23

HOW DOES COLOUR EXPERIENCE REPRESENT THE WORLD?

Adam Pautz

There is no call to treat illusory sensible qualities, and in particular colours, as actual qualities of actual entities.

David Armstrong (1984)

Many favour representationalism about colour experience. To a first approximation, this view holds that experiencing is like believing. In particular, like believing, experiencing is a matter of representing the world to be a certain way.

Once you view colour experience along these lines, you face a big question: do our colour experiences represent the world as it really is? For instance, suppose you see a tomato. Representationalists claim that having an experience with this sensory character is necessarily connected with representing a distinctive quality as pervading a round area out there in external space. Let us call it “sensible redness” to highlight the fact that the representation of this property is necessarily connected with the sensory character of the experience. Is this property, sensible redness, really co-instantiated with roundness out there in the space before you?

Since the development of the new mathematical physics of the seventeenth century, many prominent thinkers have returned a negative answer. Galileo, for instance, famously said that “tastes, odours, colours, and so on reside in consciousness”, not the external world. Following this tradition, some contemporary representationalists hold that tomatoes and other objects are just collections of particles and fields lacking sensible properties. We evolved to have experiences that habitually misrepresent objects in space as having various sensible colours, only because this helps us to discriminate them from one another.

However, other representationalists resist this radical irrealist view. They think representationalism about colour experience goes best with “realism” about sensible colours. The tomato’s surface really does exemplify sensible redness. In general, our colour experiences typically represent objects as they really are.

We will look at these different versions of representationalism. But first I will explain in more detail the basic representationalist approach they share.
I Representationalism about colour experience: the basic idea

To understand representationalism, it is best to start with non-veridical colour experience. Suppose that someone—call her “Mary”—suffers from “Charles Bonnet syndrome”, a condition which involves having vivid hallucinations. Suppose in particular that she hallucinates a tomato. Intuitively, having this phenomenal experience is essentially connected with having an experience as of a reddish and round item in a certain location. But, in this case, no physical round and reddish thing is present. So how come it is correct to say Mary has an experience as of a reddish and round item?

Before representationalism became popular, the dominant view was the *sense datum theory*. On this approach, what Mary experiences is a non-physical round and reddish item created by her brain—a “sense datum” or “mental image”. In general, sensible colours do not qualify ordinary physical objects, but such mental images of them created by the brain. However, this view faces well-known puzzles.

Enter the representationalist alternative. On representationalism, all that is going on is that Mary’s experience “represents” that there is something before her with the properties being round and reddish. This doesn’t require that there really is an item—a peculiar mental “sense datum”—that has these properties. On representationalism, Mary’s hallucination is like a belief in that it is a “representational” state that can represent properties belonging to no real (physical or mental) thing.

But, of course, Mary’s hallucinatory experience is a representational state of a kind very different from belief. Hallucination, unlike mere belief, involves a vivid impression of the real presence of a thing. To explain this, representationalists would claim that, in having her hallucination, Mary stands in a very unique representational relation to the properties being round and being reddish, even though those properties are not presently instantiated before her. I will call it the “perceptual representation relation”. This hypothesized relation is unique in that, when one bears it to some properties, one has the vivid impression that there is an object present with those properties. Indeed, on a simple form of representationalism, Mary having her tomato-like hallucination is just identical with her perceptually representing (or “perceptually predicating”) a cluster of tomato-like properties such as being reddish and being roundish (where these properties can exist even if they aren’t instantiated by any real thing). Representationalists, then, do not appeal “sensations”, or “qualia”, that are distinct from and lie behind perceptually representing the world to be a certain way (*pace* Campbell, this volume); rather, they say that experiencing is *nothing but* representing.

In general, most representationalists hold that, necessarily, if two individuals have phenomenally different visual experiences, then they perceptually represent different clusters of sensible properties. Other representationalists advocate weaker versions of representationalism that allow for some exceptions to this general principle (e.g. cases where you perceptually represent the same properties but there is a change in which ones you *attend to*).

Representationalism about colour experience has many virtues. It accommodates the undeniable fact that, necessarily, in having standard visual experiences, it seems that sensible colours are co-instantiated with certain shapes and location properties. For instance, necessarily, if one has a tomato-like experience, then one has an experience that matches the world only if a reddish, round item is present at a certain place. It also explains how both veridical and non-veridical experience can provide a subject with the capacity to think about and learn truths about sensible colours and shapes, even when they aren’t instantiated by physical objects before the subject (Russell 1912, chap. X; Johnston 2004, 130; Tye 2014, 51–2). At the same time, it avoids the “sense data” of traditional sense datum theory.
For the sake of discussion, let us assume that representationalism is right. This view implies that every sensory experience has two elements: (i) the perceptual representation relation, and (ii) a complex of sensible properties that are the relata of this relation (sensible colours, shapes, audible properties, and so on). So once we accept representationalism, we face a pair of difficult questions. For colour experience, they are as follows.

First, the sensible colours question: what is the nature of the “sensible colours”. For instance, what is the nature of the “sensible redness” which Mary perceptually represents as co-instantiated with roundness? Is it a property that tomatoes and other objects sometimes really have?

Second, there is the representation relation question. What is the nature of “the perceptual representation relation” that we bear to sensible redness and other perceptible properties? For instance, when Mary hallucinates a tomato, she bears this relation to being reddish and being round. How does Mary’s brain enable her to “reach out” and perceptually represent these properties, so that they seem present before her, even though they are not really present before her? Is this “perceptual representation relation” a spooky non-physical relation between her mind and these sensible properties? Or can it be identified with some kind of unproblematic physical relation (e.g. the kind of “tracking relation” we’ll discuss in the next section)? As Mark Johnston has observed (2011, 215–16), a central puzzle about the mind is how to “explain the relation of sensory intentionality”.

Let us now turn to different versions of representationalism about colour experience. They differ in how they answer these twin questions.

2 Response-independent representationalism

I will explain response-independent representationalism by way of a simple argument for it based on considerations of uniformity. The argument starts with an account of our experience of shape, and then generalizes that account to the experience of colour.

Above we supposed that Mary has a hallucination of a tomato. Let’s now suppose that she has an ordinary, non-hallucinatory experience of one. Let’s start with the question: how does Mary perceptually represent the shape, being round? This is the “representation relation question” concerning shape.

Most philosophers are realists about space as we perceive it: they think that the shapes we perceptually represent are real properties of objects that are detected by our visual systems (but see note 26 for dissent). So the simplest answer would seem to be that Mary perceptually represents this property in much the same way that a thermometer represents a temperature. Very roughly: she perceptually represents the property of being round by virtue of undergoing an internal state (namely, a neural pattern) that, under biologically normal conditions, is typically caused by (“tracks”) the instantiation of that shape in the environment. (Likewise for more primitive spatial features, like edges and angles.) This is a simple version of what we might call the “tracking” theory of how we perceptually represent shape.

Once we accept this “tracking” model of how Mary perceptually represents roundness, considerations of uniformity suggest using that same model to explain how she perceptually represents the reddish quality that seems to her to be co-instantiated with roundness. On such a uniform view, sensible redness, like roundness, is a real, objective property of the tomato that our visual system causally detects or tracks. What could this property be? Our visual system tracks reflectance properties: that is, dispositions to reflect certain amounts of incident light across the visible spectrum. Objects that have different colours will have different reflectances (see Figure 23.1 below). So if we say that sensible redness just is a particular reflectance property of the tomato, the payoff is that we can then say that Mary perceptually represents this property in the same way she represents the property of being round.
This view may seem immediately implausible because sensible colours seem to be non-dispositional, simple properties that you can’t define in other terms. But maybe this is just wrong. On the basis of ordinary experience, it doesn’t seem to us that water is H₂O. But that is what water is. Similarly, maybe sensible colours are in fact complex dispositional properties involving light, even if this is not visually evident.⁵

This approach also applies to illusion and hallucination. For example, return to Mary’s tomato-like hallucination. How can she stand in the perceptual representation relation to the property being reddish (and also the property being round), even though it is not instantiated before her? What is it for this sensible property to be ostensibly present to her? On the tracking view, the answer is that the property being reddish is just a reflectance property. And Mary stands in a tracking relation to this property, even though she is hallucinating and this property isn’t instantiated before her. Roughly, this consists her standing in this relation to it: she undergoes a neural pattern that is normally caused by its instantiation (when she sees the tomato, for instance).⁶

This yields response-independent representationalism about colour experience. It combines representationalism with the hypothesis that sensible colours are (like shapes) real properties objects that are totally independent of how we respond to them.⁷

Response-independent representationalism is externalist. On this view, the characters of your colour experiences are not fixed by the intrinsic features of your neural processes. Rather, they are fixed by what reflectance properties those neural processes track in the extracranial world. For instance, Tye, a leading proponent, says that “[the brain] is not where phenomenal character is to be found” (1995, 162–3). Instead, his slogan is “phenomenal character is in the world” (2009, 119). So, for instance, the phenomenal character of your experience of red is not constituted by a neural pattern, but by a reflectance property (see Figure 23.1 below).

Response-independent representationalism, in the version we are considering, is also reductive. It combines representationalism with a reductive physicalist approach to experience. It holds that sensible colours are identical with complex physical properties (namely, reflectance properties) and the perceptual representation relation is identical with a complex physical relation between subjects and those complex physical properties (namely, a ”tracking” relation). So it is attractively simple.⁸

However, there are also arguments against response-independent representationalism. Here we will look at two types of argument: the structural-mismatch argument and arguments from variation.

First, the structural-mismatch argument. To illustrate, suppose that someone—call him Maxwell—consecutively experiences a purple-looking grape, a blue-looking ball, and finally a green-looking leaf. Now suppose Maxwell says, “blue resembles purple more than green”. This is an evident truth about the resemblance structure of the sensible colours.

But it is a truth that response-independent representationalists apparently cannot accommodate. For the actual reflectance types typical of such objects are shown in Figure 23.1 below. On response-independent representationalism, the sensible colours purple, blue, and green just are these reflectance-types. So, on this view, Maxwell’s statement is true just in case the blue reflectance type (in the middle) resembles the purple reflectance type (on the left) more than the green reflectance type (on the right). But, somewhat surprisingly, this is apparently not true. In fact, if anything, the blue reflectance type resembles the green reflectance type more than the purple reflectance type. So the identification of the sensible colours with the corresponding reflectances types implies that Maxwell’s statement “blue resembles purple more than green” is false. Indeed, it seems to imply that “blue resembles green more than purple” is true! Since these implications are incorrect, the fact of structural-mismatch shows that sensible colours must be distinct from the corresponding response-independent reflectance types of objects.⁹
Why is there such a surprising mismatch between resemblances among the apparent colours of objects and the resemblances among those objects’ reflectances? The answer seems to lie in our neural responses to those reflectances. For instance, even though the blue reflectance and the green reflectance are very similar, they lead to quite different neural responses in the brain, and this is why they lead to experiences of quite different sensible colours. This strongly suggests an alternative to response-independent representationalism. In particular, it suggests an account of sensible colours, and of their resemblances, in terms of our neural responses. We will return to these points later.

Some response-independent representationalists have tried to answer the structural-mismatch argument. In particular, Byrne and Hilbert (2003) have tried to give a response-independent account of the truth of “blue resembles purple more than green” in terms of reflectances. First, they assert that there are four basic hue-magnitudes, namely, being reddish, being greenish, being yellowish, and being bluish. They identify each of these four hue-magnitudes with a disjunction of reflectances (2003, 55). Thus, being bluish is just the disjunction of reflectances possessed by bluish objects (e.g. blue objects, purple objects, etc.). Then they say that the truth of “blue resembles purple more than green” amounts to the following: the blue reflectance-type and the purple reflectance-type, but not the green reflectance-type, imply a common hue-type (in this case, being bluish), that is, belong to one of the aforementioned disjunctions of reflectances. In short, they hold that when we say “blue resembles purple more than green”, what we are saying is true just in case the relevant reflectance-types stand in the hue-difference relation, that is, the relation $x$ and $y$, but not $z$, imply belonging to a common hue-type (that is, imply having one of the aforementioned disjunctions of reflectances).

Byrne and Hilbert’s response to the structural-mismatch argument faces several problems of detail. However, here we will focus on one general worry about their discussion. Although we are focusing on sensible colours, it is important to realize that the structural-mismatch argument poses a general problem for response-independent representationalism. Therefore, to answer the argument, it is not enough to focus just on the case of sensible colours, as Byrne and Hilbert do.

Let’s look at the experience of smell. Byrne and Hilbert (2003, 59) note that, because we expect a uniform account of different sensible properties, response-independent representationalism about the experience of colours stands or falls with response-independent representationalism about the experience of other types of sensible properties. So if response-independent representationalism is right for the experience of colour, you would expect a parallel account to apply to the experience of smell. On such an account, smell qualities (citrus, minty, etc.) are response-independent chemical types, and we perceptually represent them because our olfactory systems track them.
However, the structural-mismatch argument against response-independent representationalism rears its head here as well. To illustrate, let’s take another example based on our actual resemblance judgements. Suppose that, after seeing objects with the reflectances shown in Figure 23.1, Maxwell consecutively smells the chemical types shown in Figure 23.2 below:

As a matter of fact, even though the first two chemical types are very different, they produce very similar neural responses in us. They cause in us experiences of similar, but distinguishable, citrus-like qualities (Howard et al. 2009). The third chemical type smells minty. So, Maxwell will judge "the middle smell (citrus₁) resembles the first smell (citrus₂) more than the third smell (minty)". This statement is evidently true. But, again, response-independent representationalism seems to have the absurd implication that it is false. For, on this view, the smell qualities that Maxwell experiences just are the external response-independent chemical types shown in Figure 23.2, which his olfactory system tracks. And the middle chemical type does not resemble the first chemical type more than the third—if anything, it resembles the third more than the first. Given the empirical facts, response-independent representationalism about smell just seems clearly wrong.

Response-independent representationalists like Byrne and Hilbert haven’t responded to the structural-mismatch argument in the case of smell resemblance. Nothing like their “hue-difference” account applies here. As is well-known, there simply are no privileged “basic smell categories” in the way that Byrne and Hilbert think that there are four basic hue categories.

Here is another element of the structural-mismatch problem for response-independent representationalists. To illustrate, suppose that we encounter an island where all humans lack the experience of chromatic colours—they only experience black-white-grey—and also lack the sense of smell. Still, because they do experience many other qualities (achromatic colours, taste qualities, audible qualities), they can acquire a general, topic-neutral concept of comparative resemblance among qualities, “quality x resembles quality y more than quality z”. Now, suppose that we tell one of the islanders—call him “Larry”—about the colours and smell qualities I’ve discussed above, which are unlike any qualities he’s experienced. And suppose we ask him to just guess the resemblances of these qualities that he cannot experience. Employing his general concept of comparative resemblance among qualities, he makes the following guess:

[#] Blue resembles purple more than green and the citrus₁ smell resembles the citrus₂ smell more than the minty smell.

Intuitively, Larry’s guess is (determinately) true. Analogy: given a finite number of examples of arithmetical sums, Larry can acquire a general concept of plus, which he can then use to make true (or false) claims about new cases.
Colour experience representing the world

However, response-independent representationalists cannot explain how Larry’s guess [#] could be determinately true. For, on this view, the qualities his guess is about are the reflectance-types and chemical properties represented in the figures above. So, on this view, Larry’s guess is true just in case the ordered-triple <the blue reflectance, the purple reflectance, the green reflectance> and the ordered-triple <R-limonene, citral, R-carvone> both satisfy a general, topic-neutral relation of comparative resemblance that is expressed by his use of the resemblance-predicate in [#]. But how could Larry’s history on the island (his interaction with other physical properties) determine that his use of the resemblance predicate picks out a relation that is satisfied by both of these trios of disparate physical properties, physical properties he has never interacted with? The problem is not that there are no relations that are satisfied by these trios. Since relations are abundant, there are in fact infinitely many such relations; they are extremely disjunctive, unnatural relations, since they are instantiated by trios of disparate physical properties. Rather, the problem is that there are also infinitely many relations defined over reflectances and chemical properties that are not satisfied by these trios. And since all these relations are equally unnatural and disjunctive, it’s very hard how Larry’s history on the island could have determined that his use of the resemblance predicate picks out a relation of the first sort rather than a relation of the second sort. (It is not the case that relations of the first sort are more “natural” than relations of the second sort, in the way that plus is more natural than minus, and therefore easier for Larry to refer to: see Lewis 1983, 375–6.) That is why, on response-independent representationalism, it’s hard to see how Larry’s guess [#] could be determinately true.11

In the next section, we will see that alternative forms of representationalism avoid these problems concerning structural-mismatch.

Let us now turn to a second type of argument against response-independent representationalism: arguments from variation. We will consider a few different arguments of this kind.

Consider first an argument from actual cases of variation. While we are trichromats, some birds have a tetrachromatic colour vision system that is sensitive to UV light. There is reason to think that, when a human and such a bird see a tomato, they experience different sensible colours. There is also variation among normal humans. For instance, the colour chip might look pure blue to John and greenish blue to Jane, where John and Jane both have normal colour vision. On representationalism, such variation means that different perceivers disagree concerning what sensible colours they perceptually represent objects as having. But if so, then it looks like they are stuck with saying that some perceivers are getting it right and the others are getting it wrong. But this kind of misrepresentation is implausible. All the perceivers have equal claim to being veridical perceivers of colour.

However, this argument is unpersuasive. Response-independent representationalists have another option: they might accommodate such cases by accepting colour pluralism (Tye and Bradley 2001; Byrne and Hilbert 2003, 16). To explain why the tomato looks different to us and to the birds, they might say that the tomato objectively possesses the colour red and it also objectively possesses an alien colour that we cannot imagine constituted by UV light. We track and thereby perceptually represent the red colour while the bird tracks and thereby perceptually represents the alien colour. So there is no colour misrepresentation. Byrne and Hilbert (1997, 272–3) and Kalderon (2011) extend pluralism to variation among humans. The colour chip viewed by John and Jane is both pure blue and slightly greenish blue, where these shades are identical with distinct but overlapping ranges of reflectances. John and Jane’s different visual systems track and thereby represent these different but equally real properties of the chip.12

Because actual cases do not clearly undermine response-independent representationalism, some rely on the conceivable of certain hypothetical “altered experience” cases to try to undermine the view. One prominent example is David Chalmers. To illustrate, let’s turn back to the
example about Suppose Maxwell. Maxwell looks at the ball with the middle reflectance shown in Figure 23.1. Then it looks bluish to him. Chalmers (2010, 415–16; and 400, n. 7) would say that it’s conceivable that there should be another individual—call him Twin Maxwell—belonging to another human-like species who bears the tracking relation to the very same response-independent reflectance property of the ball as Maxwell, but who phenomenally represents a totally different sensible colour—say, being greenish. If the conceivability of this “altered experience” case shows that it is really possible, then of course the phenomenal representation relation cannot be identified with a tracking relation, and sensible colours cannot be identified with reflectance properties.

In response, response-independent representationalists like Tye (2000, 109–11) have said that, when it comes to such cases, mere conceivability doesn’t establish possibility. According to this response, such altered spectrum cases are conceivable, but maybe they are not really possible. Chalmers himself concedes (2010, 152) that such conceivability arguments may fail because we are ignorant of the true nature of the physical world. So there is reason to doubt that the mere conceivability of hypothetical “altered spectrum” cases refutes the externalist approach of response-independent representationalism.

This brings us to a final type of argument against response-independent representationalism based on variation in colour vision. This argument is similar to the previous one, except that it relies on research in neuroscience, rather than mere conceivability, to argue for the possibility of a certain kind of “altered spectrum” case that is inconsistent with response-independent representationalism. This argument may succeed where the previous arguments from variation fail.

To illustrate the relevant research in neuroscience, we can use the same example that we used to illustrate the structure argument: the example where Maxwell consecutively experiences a grape, a ball, and a leaf with the reflectance-types shown in Figure 23.1. He then experiences purple, blue, and green. His colour experience of the ball resembles his colour experience of the grape more than his colour experience of the leaf. This phenomenological fact cannot be explained in terms of the reflectances of those objects tracked by his visual system: as we noted before, the reflectance of the ball does not resemble the reflectance of the grape more than the reflectance of the leaf. Now here is a question that we haven’t looked at yet: what then is the explanation? Recent neuroscience supports the hypothesis that the explanation lies in his neural responses to those reflectances. In particular, Brouwer and Heeger (2009) and Bohon et al. (2016) found that, in brain area V4, neural similarity among distributed patterns of activity predicts phenomenal similarity among colour experiences. In fact, if you order these distributed neural patterns by overall similarity, then they form a circle akin to the “colour circle”. So we can conjecture that the reason why Maxwell’s colour experience of the ball resembles his colour experience of the grape more than his colour experience of leaf is that his internal V4 neural representation of the ball is more like his V4 neural representation of the grape than his V4 neural representation of the leaf. The explanation is to be found inside, not outside.

Now we can use this research to construct a two-step argument for the possibility of a hypothetical altered spectrum case that is inconsistent with the externalist approach of response-independent representationalism.

First step. Imagine that Maxwell has a twin, Twin Maxwell, in a hypothetical human-like community that evolved differently than humans. Just like Maxwell, Twin Maxwell consecutively experiences a grape, a ball, and a leaf with the reflectances shown in Figure 23.1. Let us suppose that Twin Maxwell’s V4 neural representations of the grape and the leaf are identical with Maxwell’s. However, because of naturally evolved differences in this community’s post-receptoral wiring, let us suppose that Twin Maxwell’s V4 neural representation of the middle object only—the ball—differs from Maxwell’s. Specifically, while Maxwell’s V4 neural
representation of the ball resembles his V4 neural representation of the grape more than his V4 neural representation of the leaf, in Twin Maxwell the opposite is true: his V4 neural representation of the ball resembles his V4 neural representation of the leaf more than his V4 neural representation of the grape. As a result, while Maxwell sorts the ball with the grape, Twin Maxwell sorts it with the leaf. At the same time, we can suppose that Twin Maxwell’s V4 neural representation of the ball, although it is different from Maxwell’s V4 neural representation, tracks exactly the same reflectance-type as Maxwell’s V4 neural representation, namely a reflectance-type that includes the reflectance shown in the middle in Figure 23.1. Therefore, on viewing the ball, they bear the tracking relation to the very same reflectance-type. Analogy: two people, in who belong to different language communities, can use different words to track the same type of object.  

So far, we have only stipulated the physical facts about the case. We have left open how Twin Maxwell’s colour experiences of the objects compare to Maxwell’s. That is the crucial issue.

Second step. What is the most reasonable view on this issue? Well, when it comes to the first and third objects, namely the grape and the leaf, everything is physically the same between Maxwell and Twin Maxwell: they respond to the reflectances of these objects with the very same neural processing and the very same V4 neural states. So, presumably, the grape looks purplish to both, and the leaf looks greenish to both. But what about their experience of the ball in between? We know that, while Maxwell’s V4 neural representation of the ball resembles his V4 neural representation of the purple-looking grape more than his V4 neural representation of the green-looking leaf, in Twin Maxwell the opposite is true: Twin Maxwell’s V4 neural representation of the ball resembles his V4 neural representation of the green-looking leaf more than his V4 neural representation of the purple-looking grape. We also know that similarity among the reflectances tracked by the colour system is a very poor predictor of phenomenal similarity and that similarity in V4 neural states is the only good predictor of phenomenal similarity. So, the most reasonable view is that, on viewing the ball, while Maxwell experiences blue, Twin Maxwell experiences a greenish colour. This verdict is also supported by their different sorting dispositions.

But this most reasonable verdict about the case is inconsistent with response-independent representationalism. On response-independent representationalism, they should have the very same colour experience, because they bear the tracking relation to the very same reflectance-type. The neural difference shouldn’t make a difference to the character of their experiences, because it doesn’t make for a difference in what they track. Since this verdict is not reasonable, we must reject response-independent representationalism. Even though they track the same response-independent reflectance-type, Maxwell and Twin Maxwell experience different sensible colours, contrary to response-independent representationalism.

Let us use “internal-dependence” to name the thesis that there are possible hypothetical cases in which two individuals have different experiences (and hence, given representationalism, phenomenally represent different properties) because they have suitably different neural states, even though those neural states track exactly the same response-independent properties. In general, the present argument has two steps. First, many different lines of research, across the sense-modalities, provide overwhelming support for internal-dependence. Second, response-independent representationalism is incompatible with internal-dependence. Call this the internal-dependence argument. Notice that the internal-dependence argument against response-independent representationalism differs from the structure argument, even if we used the same example to illustrate both arguments. The structure argument was that response-independent representationalism cannot
accommodate the truth of claims like [#], given the mismatch between our judgements of similarity among sensible properties and the similarity relations among the corresponding response-independent physical properties. The internal–dependence argument is different. The internal–dependence argument is that response-independent representationalism, as an externalist theory of phenomenal character, cannot accommodate an additional fact: the empirically supported role of the brain in shaping the phenomenal character of experience.

The internal–dependence argument may succeed where other arguments from variation fail. It is invulnerable to the “pluralist” response that response-independent representationalists may use to handle actual cases of variation. For, even if the ball has a plurality of chromatic properties, we have stipulated that Maxwell and Twin Maxwell track exactly the same one when they see the ball. So response-independent representationalism inevitably delivers the mistaken verdict that they experience the same sensible colour, despite the radical neural and behavioural differences between them (Cohen 2009, 86). Further, the internal–dependence argument is invulnerable to the “mere conceivability doesn’t entail possibility” response that can be used to answer David Chalmers’ argument against response-independent representationalism. In fact, the internal–dependence argument doesn’t rely on conceivability at all to establish the possibility of individuals experiencing different sensible colours while tracking the same response-independent properties. Instead, it provides a two-step empirical argument for this possibility. Given internal–dependence, internal neural responses play a big role in determining what sensible colours we perceptually represent in external space. This brings us to response-dependent representationalism.

3 Response-dependent representationalism

I will introduce response-dependent representationalism by way of explaining how it might accommodate “internal–dependence”. Recall that, given internal dependence, Maxwell and Twin Maxwell track the same reflectance property of the ball, but nevertheless perceptually represent it as having different sensible colours, namely being bluish and being greenish, due to their differing internal neural states.

In one version, the proposal of response-dependent representationalists is that the different sensible colours that Maxwell and Twin Maxwell perceptually represent are identical with the different dispositions of the ball to cause their differing internal neural states (call them “N1” and “N2”). So, the bluish sensible colour that Maxwell experiences is identical with a response-dependent property of the ball along the lines of: normally causing neural state N1 in members of one’s kind. (Or maybe normally causing N1 in members of some kind or other.) By contrast, the greenish sensible colour that Twin Maxwell perceptually represents is identical with a different such property: normally causing neural state N2 in members of one’s kind. On this view, the visual system is narcissistic: it represents how objects affect the visual system.

So response-dependent representationalism shows more promise in accommodating internal–dependence than response-independent representationalism. Unlike response-independent representationalism, it may also be able to accommodate our “structure judgements”. On this view, judgements about the resemblances among sensible colours are not about response-dependent reflectances properties of objects like those in Figure 23.1, as on response-independent representationalism. Rather, they are judgements about response-dependent properties of objects of the form normally causing neural response N. Now, such response-dependent properties resemble insofar as the responses in terms of which they are defined resemble (Johnston 1992, 240). So if, as empirical research seems to indicate (Brouwer and Heeger 2009), the resemblances among our internal neural states in the colour vision system match our judgements of resemblances.
among sensible colours, then those judgements come out true. Response-independent representationalists can generalize the same account to our judgements of resemblances among other sensible properties, such as judgements about the resemblances among smells considered in the previous section.

However, response-dependent representationalism faces problems of its own. We will focus on two.

First, response-dependent representationalists have focused almost exclusively on what I have called the “sensible colours question”. But what is their answer to the “representation relation question” about how we represent sensible colours?

As noted in §2, it is natural to use a broadly “tracking” model to explain how we can perceptually represent spatial properties, and considerations of uniformity suggest generalizing this account of the perceptual representation of sensible colours. But response-dependent representationalists cannot accept a tracking model for the perceptual representation of sensible colours.

To see this, return to the case of Maxwell and Twin Maxwell. On response-dependent representationalism, they perceptually represent the ball as having different sensible colours, identified with relational properties of the form normally causing so-and-so neural state. But they do not in any sense track these different complex, relational properties. For instance, Maxwell’s internal neural state N1 is not caused, under biological normal conditions, by the instantiation by the ball of the relational property: normally causing that very neural state, N1, in Maxwell’s population. For the property of normally causing N1 isn’t itself casually efficacious in causing N1. Our visual systems are not causally sensitive to such properties. Furthermore, N1 arguably doesn’t have the biological function of tracking this specific ecologically insignificant relational property. Similar remarks apply to Twin Maxwell. On the contrary, Maxwell and Twin Maxwell’s different internal neural states are causal in, and have the function of tracking, the same reflectance property of the ball (Cohen 2009, 86), since the colour system is geared to recovering surface reflectances. Since response-dependent representationalists claim that they perceptually represent different response-dependent properties, it follows that they cannot identify the perceptual representation relation with a standard tracking relation.

But if Maxwell and Twin Maxwell’s visual systems do not have the function of tracking the relevant response-dependent properties of the ball, then what is the alternative account of how they perceptually represent them? Sydney Shoemaker, a prominent defender of this approach, candidly admits, “I have no fully satisfactory answer” (1994, 37). Call this the representational problem.

The problem is made more difficult by the fact that the ball has multiple response-dependent properties of this sort involving different subjects, different responses, and different conditions: for instance, it normally causes certain V4 neural responses in Maxwell under certain conditions, it normally causes certain V4 neural responses in Twin Maxwell under certain conditions, it normally causes certain retinal activity in Maxwell, and so on. What makes it the case that Maxwell perceptually represents one specific response-independent property on this list, and Twin Maxwell perceptually represents another? What makes it the case that certain specific response-dependent properties are visually represent to them?

In fact, the problem is even more difficult than this. For, in addition to perceptually representing sensible colours, we of course perceptually represent spatial properties, such as shapes, positions, and distances. These spatial properties are evidently not response-dependent properties of the form: normally causing neural response X. So, response-dependent representationalists must answer the following question: what is the single, uniform reductive theory of the perceptual representation relation that implies that in some cases we perceptually represent funny
response-dependent properties of the form normally causing neural response X in individuals with colour system S (or whatever response-dependent with which they identify the sensible colours), while in other cases we perceptually represent response-independent spatial properties of a radically different sort? It would be implausible to answer by saying that the shape-system has the function of indicating response-independent shape properties, while the colour-system has the sole function of indicating these entirely different, funny response-dependent properties (rather than, say, the function of indicating response-independent reflectance properties, or simply the function of guiding our behaviour in a useful way). Since the notion of “the function” of a system is unclear, it is just not at all obvious what could this mean and what could make it the case.19

There is a second, much simpler problem for response-dependent representationalism. Briefly, on this view, sensible colours are relations to types of neural states and types of perceivers. But this doesn’t fit the phenomenology. Sensible colours just don’t look like relations. They seem non-relational. Indeed, they seem “simple”. Call this the phenomenological problem.20

So far, we have looked at one version of response-dependent representationalism. But it is not the only version. Colin McGinn (1996) has devised a complicated novel version that is designed to avoid the phenomenological problem. As a bonus, it suggests an answer to the representational problem. Let me explain.

The version of response-dependent representationalism we have so far considered is reductive: it identifies sensible colours with response-dependent, physical properties of objects (see note 8). McGinn (1999) defends a form of representationalism. But, against reductive representationalism, he insists (1996, 541–2) that, “when we see an object as red we see it as having a simple, monadic, local property of the object’s surface”, distinct from any complex physical property. He is a non-reductive representationalist.

Even though he thinks colours are simple, non-relational properties of objects, McGinn also holds that it is just a brute “law of metaphysics” that their instantiation by objects is tied to those objects’ effects on perceivers. This is what makes his view qualify as response-dependent. In particular, he postulates the following general principle: it is metaphysically necessary that an object possesses the fact that the ball has this relational property grounds the fact that it actually instantiates the property being blue. Nevertheless, according to McGinn, being blue is a “simple”, non-relational property of the ball that is distinct from this relational property that always grounds it. So McGinn thinks that [Bicon] doesn’t amount to a reduction of being blue. Likewise, because the ball is also disposed to normally look greenish to normal twin humans (like Twin Maxwell), it also instantiates the simple property being greenish, according to [Bicon]. In general, thanks to [Bicon], objects are guaranteed to really possess exactly those simple colours that they normally appear to possess.

McGinn’s view provides a unique origin story for colour. Given [Bicon], before sentient creatures evolved, objects had no sensible colours, for the simple reason that they did not habitually look to have any colours to any individuals. Then brains evolved that have an intrinsic capacity to enable us to perceptually represent objects as having certain simple sensible colours. That is, brains evolved that in this sense “project” simple colours into objects. Thanks to [Bicon], those objects thereby acquired those (and only those) simple sensible colours, in addition to their scientific properties. In other words, [Bicon] guarantees a fortunate match between appearance and reality when it comes to sensible colours.
McGinn’s non-reductive response-dependent view avoids the two problems with the reductive version considered above. First and most obviously, it avoids the “phenomenological problem” facing the reductive version. For, on McGinn’s non-reductive version, sensible colours themselves are simple, non-relational properties of objects just as they appear to be, even if by [Bicon] they are grounded in distinct complex, relational properties of those objects.

As we saw, the reductive response-dependent theorist also faces a “representational problem”. McGinn faces a question here too: how do we manage to perceptually represent these alleged “simple” sensible colours? McGinn also cannot identify the perceptual representation between subjects and sensible colours with the externally determined “tracking relation”, for a couple of reasons. This goes against internal-dependence. And it requires that sensible colours are “out there” prior to our representation of them, which goes against [Bicon].

However, McGinn’s view suggests an alternative answer to the “representational problem”. Because he is already an anti-reductionist about sensible colours, he might provide a parallel non-reductive account of our ability to perceptually represent those sensible colours. In particular, he might say that this relation is irreducible, just as he thinks sensible colours are. That is, there is no true identity of the form: the perceptual representation relation = complex physical relation R. (Compare: it is implausible that there is a general reduction of the reference relation in physical terms, given that we can refer to such diverse things as chairs, numbers, and uninstantiated kinds.) Nevertheless, he might say that, whenever one bears this relation to some sensible colours, this is grounded in one’s total internal neural state. Indeed, maybe it is possible in principle to systematically “decode” what sensible colours an individual perceptually represents from her V4 neural patterns (Brouwer and Heeger 2009; Haynes 2009, Bohon et al. 2016). This yields a non-reductive and internalist form of representationalism (Chalmers 2010; Horgan 2014). It might be combined with non-reductive physicalism: experiential facts are grounded in physical facts, even if a general reduction is impossible.

It is worth noting another virtue of McGinn’s view: it nicely avoids the structural-mismatch problem that we brought against response-independent representationalism (§2). On his view, the sensible colours that we perceptually represent are simple properties, which are quite distinct from the types of reflectances shown in Figure 23.1. So McGinn is free to say that these simple properties generally stand in exactly the resemblance relations they seem to stand in, even if the corresponding reflectances-types do not. For instance, he might say that, when Maxwell says “this shade of blue resembles this shade of purple more than that shade of green”, he is correctly reporting that the shades stand in a unique unanalysable relation of intrinsic qualitative resemblance (see e.g. Allen 2017, 124–5). Since this relation is very “natural” (in the sense of Lewis 1983), McGinn’s view also accommodates the point (illustrated in §2 with the story about Larry on a remote island) that, once we experience a finite number of resembling qualities, we can easily determinately grasp this general relation. Then we can apply it to new qualities, and thereby make true or false claims.

McGinn’s view also avoids a general problem that Chalmers (2010, 400) raises for all realist views of colour. Chalmers suggests that realists are led to chromatic explosion. Let’s go back to the ball with the reflectance shown in Figure 23.1. It normally looks one colour to Maxwell. It could have looked another colour to “Twin Maxwell”. Indeed, Chalmers holds that, for any colour you choose, the ball could normally look that colour to a possible perceiver. So, to avoid favouritism or arbitrariness, mustn’t realists say that, right now, the ball has all of those colours? McGinn’s view—which Chalmers doesn’t consider—avoids this radical chromatic explosion. Given his [Bicon] principle, the ball possesses only those simple colours it appears to have to some actual normal perceiver under some normal conditions. In general, an object only has enough colours to make the experiences of actual perceivers generally veridical, and not more. This is a limited colour pluralism.
While McGinn’s view avoids some problems, it also faces a few new ones. Here I will mention two.

First, the case for [Bicon] is unclear. It is not a priori. This is shown by the fact that the irrealist representationalist view to be considered in the next section rejects [Bicon], but it cannot be ruled out a priori (Tye 2000, 170). The only argument for [Bicon] seems to be that it allows McGinn to agree with our pre-theoretical conviction that objects have the sensible colours they appear to have. But our pre-theoretical convictions have a bad track-record.

Second, simplicity considerations count against [Bicon]. It reports a necessary connection between two highly disparate sorts of properties, namely, simple, occurrent sensible colours and complex dispositions of objects to appear to have those sensible colours. It is not derivable from logic or from any general modal principles. It would have to be accepted as an additional, brute principle—a special brute “law of metaphysics” (Rosen 2009, 133), or a “grounding dangler”.

Third, [Bicon] faces a problem about borderline cases. As Hardin (1988) has emphasized, lighting conditions can radically influence colour appearance but there is no precise cutoff between normal and abnormal lighting conditions. Now suppose that you view a colour chip in lighting conditions C, where this is a borderline case of normal conditions. On McGinn’s view, there is a specific simple shade that you determinately perceptually represent, which is somewhat different from the apparent shade of the chip in perfect daylight. Suppose you dub it “blue,31”. Now, by [Bicon], the chip really instantiates blue,31 iff it looks blue,31 under normal conditions. Since in this case the right-hand side is indeterminate (since C is a borderline case of “normal conditions”), it follows from [Bicon] that the left-hand is indeterminate too. That is, there is this specific shade, blue,31, such that it is indeterminate whether the chip instantiates it. Hence [Bicon] requires “vagueness in the world”, which many consider to be incoherent (Lewis 1993).

4 Irrealist representationalism

Irrealism holds that ordinary physical objects don’t instantiate sensible colours. All that is out there are particles and waves and fields. It became very popular after the scientific revolution of the seventeenth century and was defended by Galileo, Newton, Descartes, and Locke. Hardin (1987) and Chalmers (2010) describe the rejection of realism in dramatic terms: they call it a “fall from Eden”.

Traditionally, irrealists have said that, although sensible colours do not qualify ordinary physical objects, they do qualify something. For instance, if you look at a tomato, the reddish quality is instantiated by a round item you experience (called an “idea”, a “sense datum”, or a “visual field region”) located in a kind of private mental arena created by your brain (Boghossian and Velleman 1989). On this view, you only ever experience such very life-like mental images, which you mistakenly believe to be physical objects. But, as I noted at the outset, such “sense data” create serious puzzles.

Representationalism has made possible a new and more defensible form of irrealism. Irrealist representationalism holds that, when you view a tomato, you perceptually represent the property being reddish as co-instantiated with the property being round. So it seems that these properties are instantiated together before you. But, in fact, the property of being reddish is not instantiated by anything. It is not instantiated by a physical round thing before you; and it is also not instantiated by a mental round thing—there is no such thing. There is no reddish thing there of any kind, even though there seems to be one. So the property being reddish is a bit like the property being a unicorn: it is an entirely uninstantiated property. When it comes to sensible colours at least, the brain is a kind of “partial virtual reality machine”, which projects onto objects some features that they don’t really have.21
In some respects, irrealist representationalism resembles McGinn’s view considered above. It shares with McGinn’s view the following two claims. First, we perceptually represent objects as having simple colour properties, which cannot be identified with reflectances or the like. Second, the brain has an intrinsic and innate capacity to enable us to perceptually represent these properties, without any contribution from the world. The difference is this: McGinn wants to accept the commonsense view that objects normally have the very simple colour properties that the brain projects onto those objects, so he posits a brute principle, [Bicon], that guarantees this result. By contrast, irrealists reject [Bicon]. They deny that objects really have the simple colour properties that the brain projects.

What is the argument for irrealist representationalism? In his influential paper, “Perception and the Fall from Eden”, David Chalmers provides a mostly a priori defence of this view. As we saw (§2), he uses the conceivability of “altered spectrum cases” to rule out the view of response-independent representationalists that sensible colours are identical with reflectance properties. More generally, he argues that sensible colours cannot be identical with any physical properties, because there is an intuitive “epistemic gap” between sensible colours and all physical properties (2010, 399, 415). So, Chalmers concludes, sensible colours must be simple, irreducible properties. Then, as we noted before (§3), Chalmers argues that, to avoid arbitrariness, realists would need to accept the conclusion of “chromatic explosion”: every object has every simple colour (p. 400)! (Allen (2017, 71–2 apparently defends this view.) Chalmers says that “this conclusion is even more counterintuitive than the [irrealist] conclusion that all color experiences are illusory” (ibid.). So, he opts for irrealism.

But this complex a priori rationale for irrealism faces problems. First of all, as we already noted, Chalmers’ conceivability argument against externalist approaches like response-independent representationalism is open to the standard criticism that “conceivability doesn’t establish possibility”. In other words, response-independent representationalists might agree that “altered spectrum cases” like the case of Maxwell and Twin Maxwell are conceivable, but then just deny that they are possible. Second, even if representationalists accept the possibility of such cases, realism doesn’t require “chromatic explosion”. As we saw, McGinn’s form of realism, which Chalmers does not consider, puts the brakes on chromatic explosion, and yet it is liberal enough to accommodate the veridicality of our actual colour experiences in normal conditions.

However, there is also a more empirical argument for irrealist representationalism. This argument has two stages. First stage: representationalists can rule out reductive theories of sensible colours on the basis of empirical arguments, without having to rely on Chalmers’ a priori considerations. In particular, sensible colours cannot be identified response-independent reflectance properties of objects. For this view is undermined by the structural-mismatch argument and the internal-dependence argument (§2). Sensible colours also cannot be identified with response-dependent properties of objects of the form: normally causing neural response N in community C. This view is undermined by the representational problem and the phenomenological problem (§3). So if representationalists want to be realists about sensible colour, then their only option is to accept a view like McGinn’s on which objects have simple sensible properties over and above their physical properties. Second stage: this type of view can now be argued against on the basis of simplicity considerations (§3). McGinn’s view requires the [Bicon] principle, which posits a massive unexplained coincidence. In particular, [Bicon] is the conjecture that, if we evolved so that our brain projects certain simple sensible colours onto objects, then viola those objects thereby acquire those very simple sensible colours. But why should nature be so obliging? There is no overwhelming a priori or empirical reason to accept this principle. Of course, it could be true. But irrealist representationalist will say that we are obliged to reject it for the same reason we are obliged to reject other needlessly complicated empirical theories.
Another consideration in favour of irrealist representationalism about colour experience over the realist forms of representationalism is this. There are especially strong empirical and \textit{a priori} reasons to think various \textit{qualities} we feel in bodily regions (pain qualities, “pins and needles”, etc.) cannot be identified with any real, mind-independent physical properties of those bodily regions. There are also especially strong empirical reasons to think that smell properties (discussed in §2) as well as audible properties (e.g. phoneme-type) cannot be identified with real physical properties of external items. These are all “projected” qualities. Why should we continue to insist that \textit{sensible colours} are an exception? Why insist that they are really possessed by external items, if these other sensible properties are not (Locke 1869, II.viii.16)?\textsuperscript{23}

Finally, there is an evolutionary argument for taking irrealist representationalism seriously. The primary function of the sensory systems is to enhance adaptive fitness—not to represent the way the world really is. There is every reason to expect that this should sometimes involve embellishment or error, depending on an individual’s unique ecology. For instance, even if fruits are not intrinsically bright or sweet, it is understandable that we evolved to experience them as bright and sweet.

Irrealism in some form or another is the dominant view among colour scientists. For instance, Cosmides and Tooby (1995, xi) maintain that “color is an invention that specialized circuitry computes and then projects onto physically colorless objects”.

While irrealism is a popular view among colour scientists, it is typically considered to be an outlandish “position of last resort” among philosophers. There is a curious disconnect between philosophy and science. Let us consider some of the philosophical objections to irrealist representationalism, together with possible replies.

\textbf{Objection.} The irrealist representationalist holds that, when you see a tomato, then \textit{redness} is in some sense \textit{ostensibly present to you}, and this is bound up with the character of your experience, even though \textit{it is not instantiated in the world or in your experience}. But this is difficult to understand.\textsuperscript{23}

\textbf{Reply.} To see that this is not an objection, consider a different example. Suppose you have the condition known as “Charles Bonnet syndrome” which involves vivid hallucinations. You hallucinate an object with a very unusual shape. In this case, everyone should admit that, necessarily, in having this experience, you in \textit{some sense} ostensibly presented with an unusual \textit{shape property}, and this is bound up with the character of your experience, even if this property is not instantiated in your environment or in your experience (setting aside exotic objects like “sense data”). This explains why everyone recognizes a sense in which your experience is \textit{non-veridical}. It also explains how your experience gives you the ability to refer to this unusual shape property (a shape property you might have not recently encountered in real life). If this kind of delusory presentation of an uninstantiated property can happen with respect to shape, why can’t it happen with respect to colour?

\textbf{Objection.} But how could such colour illusions \textit{always} happen? How could the brain enable us to perceptually represent properties of a wholly novel sort that have \textit{never} been instantiated in the world? Our standard models for explaining how we represent properties appeal to \textit{tracking} real instances of those properties in the world, but irrealists cannot accept this view. In short, irrealism requires a total mystery when it comes to the “representation relation question” (Tye 2000, 166).

\textbf{Reply.} Irrealist representationalists can make a few points in reply. \textit{First}, we already know from internal-dependence that tracking theories fail in the special case of perceptual representation. For instance, Maxwell and Twin Maxwell experience different sensible colours but track the same reflectance properties. In view of this, the best view of perceptual representation may be the internalist, non-reductive view we considered in connection with McGinn’s approach.
And that view is consistent with irrealism. On this internalist view, the brain has an intrinsic capacity to represent sensible colours and other sensible properties, whether or not they are instantiated in the world.  

Second, we already know, independently of the colour debate, that the brain is “creative”: it is not limited to representing what is there. For instance, it can represent uninstantiated kinds and abstract objects (Chalmers 2010, 417). So why not uninstantiated sensible colours?  

Third, the claim of irrealist representationalists that the brain has an intrinsic capacity to represent unreal sensible properties may look mysterious only because we do not know enough about the brain. Maybe, if we had a more systematic understanding of the “neural code”, we would understand why one neural state is the basis of representing the colour red, while another is the basis of representing the colour orange (Brouwer and Heeger 2009; Haynes 2009, Bohon et al. 2016).

Objection. Suppose you view a tomato. Then you have a neural representation of sensible redness “bound” in the brain with a neural representation of roundness. We just saw that irrealists must accept an internalist theory of how we perceptually represent sensible redness. But previously (§2) we saw that it is very natural to accept an externalist, tracking theory of how we perceptually represent shape. So irrealist representationalists require a non-uniform theory of perceptual representation. Since the perceptual representation of shape is inextricably bound up with the perceptual representation of sensible colour, such a non-uniform theory is untenable: if the perceptual representation of sensible colour is internally determined, so too must be the perceptual representation of shape.

Reply. In fact, irrealist representationalists are not committed to a non-uniform theory of perceptual representation. Chalmers (2010, 443) and Horgan (2014) have argued the perceptual representation of spatio-temporal properties is internally determined, just like the perceptual representation of sensible colours. In fact, one might go further and argue that perceived spatio-temporal relations are no more “out there” than sensible colours. For instance, it seems obvious that the spatio-temporal relations we phenomenally represent are not frame-relative, but relativity theory shows that no such relations are out there.

Objection. Irrealist representationalism requires an implausible across-the-board error theory about all of our talk of colours.

Reply. In fact, irrealist representationalism does not require this. Recall Maxwell viewing the grape, the ball, and the leaf. He judges “the blue colour resembles the purple colour more than the green colour”. Remember that response-independent representationalists have trouble accommodating the truth of this judgement (see Figure 23.1). By contrast, irrealists can hold that this is a perfectly true judgement about the relevant sensible colours, even though they are not instantiated. Compare: there can be truths about resemblances among complex shapes, even if they are not instantiated. Maxwell also judges “the grape is purple, the ball is blue, and the leaf is green”. Irrealists can say that he thereby expresses or implicates a truth about how these things habitually look (Boghossian and Velleman 1989; for a somewhat different conciliatory approach, see Chalmers 2010).

4 Conclusion

One general moral to draw from the preceding overview is that representationalism about colour experience is a very flexible doctrine. It is compatible with just about every major view on the metaphysics of sensible colours. The hope of representationalists is that their view can simultaneously accommodate two facts: (i) the phenomenological fact that we ostensibly experience sensible colours “out there” along with shapes and locations, and (ii) the empirical fact that the experience of sensible colours is bound up with neural processing in the head.
Notes

1 A note on terminology: Most representationalists, for instance Byrne and Hilbert (2003) and Tye (2000), think that ordinary colour terms in English like “the colour red”, or “redness”, or “that colour” refer to what I’m calling “sensible redness”. If they are right (and I think that they are), then we could drop the cumbersome term “sensible redness” and simply use “redness”. But other representationalists, for instance Shoemaker (1994) and Chalmers (2010, essay 12), deny that ordinary colour terms to refer to this property. So, to refer to this property, they invent special technical terms, like “phenomenal redness” (Shoemaker), or “perfect redness” (Chalmers). We all have the same property in mind; we just use different terms to refer to it. To stay neutral on this terminological disagreement, I have decided to introduce the neutral term “sensible redness” to refer to this property.

2 However, the sense datum theory of colour experience has recently been defended by Brown (2010) and Robinson (1994, 59–74), among others.


4 Armstrong (1987, 36) and Byrne and Hilbert (2003, 7) stress the desirability of a uniform account of colour and shape. The point to be developed is that it is desirable to have a uniform answer to the question of how we manage to represent colour and shape (“the representation relation question”).

5 Due to “metamerism”, the best view in the vicinity is that colours are extremely unnatural disjunctions or types of reflectances (Tye 2000; Byrne and Hilbert 2003). There is a serious problem about how we might “track” and represent such unnatural properties (Armstrong 1987, 42 and Byrne and Hilbert 2003, 792). But, for the sake of simplicity, I will ignore this issue.

6 More exactly, proponents of the tracking model identify the dyadic perceptual representation relation between individuals and sensible properties with a complex relation along the following lines: \( \exists x y (x \text{ is in an internal state that is poised for cognitive access and that has the systemic biological function of tracking property } y). \) See Dretske (1995, 19–20), Tye (2000, 62), and Tye (2014, 51–2). This is a very simple sketch. Indeed, Byrne and Hilbert (2003) claim that all existing accounts are too simple. (MacPherson (2003) also raises a difficult puzzle about “novel colours.”) However, I will use “tracking” generically to name whatever physical relation it is that we bear to reflectance properties which enables us to perceptually represent them, under the assumption that colours are reflectance properties. In this generic sense, even Byrne and Hilbert hold that perceptual representation is a matter of “tracking”.

7 David Armstrong (e.g. 1984) is the originator of response-independent representationalism. More recent proponents of some form of this view include Dretske (1995), Tye (2000), Lycan (2001), Byrne and Hilbert (2003), Hill (2009).

8 On one general formulation, “reductive physicalism” about some property or relation \( R \) (being reddish, being a city, etc.) is the view that \( R \) is identical with a complex property built up from the properties on some short list of “basic” physical properties and relations (e.g. the fundamental physical properties, together with various “topic-neutral” properties and relations such as causation and similarity). For different conceptions and defences of reduction, see Lewis (1994), Jackson (1998, especially 62, 123), Dorr (2008), and Sider (2011, chaps. 7 and 13).

9 Philosophers who have put forward the structural-mismatch argument include Hardin (1988, 66–7) and Thompson (1995, chap. 3). Scientists have long recognized the point when it comes to other sensible properties. For instance, Stevens et al. (1937) famously argued that perceived pitches cannot be identified with response-independent frequencies, on the grounds that the relations (in particular, equal intervals) among apparent pitches do not match those among frequencies.

10 Here are three main problems of detail facing Byrne and Hilbert’s hue-magnitude proposal. (1) Byrne and Hilbert give no response-independent account of truths involving degree such as: the degrees of blueness and reddishness in the ball are roughly equal (so that the ball roughly equally resembles a pure blue object and a pure red object). Intuitively, this is not a truth about numbers, any more than the height of the rectangle is equal to (or double, etc.) to its width is about numbers. But if degrees of hue-magnitudes are not numbers, then what are they, and what is it for them to be “equal”? Byrne and Hilbert (2003, 55) apparently hold that the degree of reddishness that the ball possesses is just a certain disjunction \( D1 \) of reflectances, namely the disjunction of the reflectances of all objects (all balanced blue-red objects, all balanced blue-green objects) that are bluish to this same degree. Likewise, they hold that the degree of blueness that the ball
Colour experience representing the world

possesses is just another disjunction $D_2$ of reflectances. This implies that the degrees of bluishness and reddish-
ness in the ball are roughly equal iff $D_1$ is roughly equal to $D_2$. But if $D_1$ and $D_2$ are just distinct (but over-
lapping) disjunctions of reflectances, as Byrne and Hilbert claim, then what is it for $D_1$ to be “roughly
equal to” $D_2$? (2) Byrne and Hilbert also give no response-independent theory of degrees of brightness.
One idea would be to equate brightness with physical luminance. But this idea fails. At equiluminance,
reds are brighter than greens (Corney et al. 2009). Indeed, a yellow patch might be brighter than a
white patch, even if the yellow patch has lower luminance (Conway 2013, 11). So how would Byrne
and Hilbert analyse “colour $x$ is brighter than colour $y$” in response-independent terms, in particular,
in terms of reflectances? (3) Finally, any adequate account must accommodate the truth of general claims
of colour resemblance such as “every colour on the hue-circle resembles nearby colours more than
farther away colours”. But Byrne and Hilbert’s hue-difference account of colour resemblance is unable
to accommodate the truth of such a general claim (Pautz 2011, 423–4, fn. 6). For instance, there are
shades of purple $P_1$, $P_2$, and $P_3$ on the hue-circle such that that $P_1$ resembles $P_2$ more than $P_3$; but
they certainly don’t satisfy the hue-difference relation as explained in the text. In response, Byrne and
Hilbert might suggest that, for arbitrary colours $x$, $y$ and $z$, $x$ resembles $y$ more than $z$ iff there is some
hue-magnitude $h$ such that the degree of $h$ of $x$ is more similar to the degree of $h$ of $y$ than the degree of $h$
of $z$. (Here I’m indebted to Keith Allen.) But this too is open to counterexamples: for instance, let $x =
R_60$, $y = R_63$, $B_37$, and $z = R_70$, $Y_30$. For other problems with Byrne and Hilbert’s proposal,
see Allen (2017, chap. 6).

11 The argument here is not that response-independent representationalists must show that the truth of
Larry’s guess [#] can be deduced a priori from the response-independent character of the physical prop-
erties shown in Figures 23.1 and 23.2. In fact, they certainly should reject this a priori claim for
standard reasons (see Davies 2014, 301–3). Rather, the argument is that, even if they do reject this a
priori claim, they still must at least gesture at an a posteriori account of how Larry managed to make a
true claim, rather than a false (or indeterminate) one. Compare: a type-type neural identity theorist for
pain needs to gesture at an a posteriori account of truths like “my second pain was more intense than my
first pain” in terms of neural firing rates or whatever, even if they deny that such truths are a priori
deducible from truths about firing rates and the line.

12 There is evidently no formal contradiction in the pluralist view of Byrne and Hilbert (1997) that the
colour chip has one colour that is completely bluish and a distinct colour that is not completely bluish
but rather a bit greenish as well as bluish. But it is very counterintuitive. For this reason, more recently,
Byrne and Hilbert (2003) have converted to inequalitarianism about this case: they now assert that the
chip does not have both shades. So, given response-independent representationalism, either John or
Jane must be misrepresenting the chip. But now a new problem arises: since John and Jane are both
normal perceivers, it is hard to see what could make it the case that one but not the other is guilty of
misrepresentation (Cohen 2009). Byrne and Hilbert (2007, 90) reply that this is a general problem faced
by anyone who accepts that John and Jane represent fine-grained colours. But, since Byrne and Hilbert
also accept inequalitarianism instead of pluralism, they do face an additional question about John and Jane
that others (for instance, pluralists) do not face, namely: what is the general theory of perceptual repren-
tation that entails that one of them perceptually represents a reflectance-type that the chip possesses,
while the other perceptually represents a reflectance-type that it does not possess?

13 Let me clarify this case a bit. As I noted before (see footnote 5), because of metamerism, response-
independent representationalists identify the blue colour of the ball with a general “class” or “disjunc-
tion” of reflectances. This might be called a “reflectance-type” or “reflectance-class” (Tye 2000; Byrne
and Hilbert 2003). This reflectance-class includes the reflectance of the ball shown in Figure 23.1 (call
this $R_1$) and all the other reflectances ($R_2$, $R_3$, $R_4$) of objects that normally have the same effect on the
visual system and so look the same shade of blue. So, here is a more complete description of the case
I’m imagining: When Maxwell views the ball, he has a V4 neural representation that tracks (is normally
caused by), and thereby represents, this reflectance-class (that is, it can be caused by $R_1$ or $R_2$ or $R_3$ or $R_4$).
When Twin Maxwell views the ball, his V4 neural representation, although different, tracks (is nor-
mally caused by) this very same reflectance-class ($R_1$, $R_2$, $R_3$, $R_4$).

14 For research on colour vision supporting internal-dependence, see Brouwer and Heeger (2009),
Conway (2013), Schmidt et al. (2014), Danilova and Mollon (2016), and Forder et al. (2017). For
research on audition, taste, smell, and pain supporting internal-dependence, see Coghill et al. (1999),
Howard et al. (2009), Chang et al. (2010), and Crouzet et al. (2015). For an overview, see Kriegeskorte
and Kievit (2013). Since internal-dependence is especially well-supported for non-visual modalities, it
is unlikely that colour vision is an exception.
15 For the internal-dependence argument, see Pautz (1998), (2003), and (2010). For discussion, see Cohen (2009, 82–8), Kalderon (2011, 250–6), Fish (2013, 58–9), and Allen (2017, 68–73). Kalderon (this volume) suggests that Shoemaker (2003, 269) uses a similar argument. But Shoemaker’s argument is quite different. In fact, Shoemaker (2003, 269) gives no argument at all thinking that the relevant kind of “altered spectrum” case is possible except that we can conceive of it (he says “we can imagine” such a case). So he seems to defend the conceivability argument discussed above. We saw that this argument is vulnerable to the response “conceivability doesn’t entail possibility”. By contrast, the internal-dependence argument relies on empirical research across the various modalities to argue that such a case is possible, not mere conceivability. Therefore, it is not vulnerable to the response “conceivability doesn’t entail possibility”.

16 For the same reason, the response-independent representationalist cannot in this case use the inelegant response that Maxwell or Twin Maxwell is misrepresenting the colour of the ball, contrary to Byrne and Tye (2006, 253). For this point, see Cohen (2009, 86).

17 There is another important problem for response-independent representationalism quite different from the empirical problems I have focused on. In particular, Chalmers (2010, 354–5) suggests that this view faces a problem about experiential indeterminacy. Here is an especially difficult case (not considered by Chalmers). Consider the well-known “inverted earth case”. (For details, see for instance Lycan 2001, 30–1.) But let’s add a twist. To begin with, suppose that there is an Earthian male whose neural representation \( S \) of the sky has the biological function of tracking blue, and an inverted Earthian female whose neural representation \( S \) of the sky instead has the biological function of tracking yellow. Strictly speaking, they belong to different species, which evolved on the different planets independently. However, let us suppose that, by an outstanding coincidence, they are genetically nearly identical (except for their different colour vision), so that they can interbreed and have a child. Now here is the problem: on response-independent representationalism, it will presumably be indeterminate whether the child’s neural representation of the sky, \( S \), represents yellow or blue, since it is indeterminate whether in this creature \( S \) has the historical, biological function of tracking yellow or blue (for we may suppose that there is nothing that could settle whether the child belongs to one parent’s species rather than the other’s). If so, it will indeterminate whether it has a yellowish or bluish colour experience. But it is deeply counterintuitive that colour phenomenology could be radically indeterminate in this way.

18 This might be called the “neural response version” of response-dependent representationalism. Harman (1996, 10) and Kriegel (2009, 90) are proponents. Shoemaker (1994) defends a different version, on which the relevant responses are colour experiences. But it faces a serious circularity problem that is avoided by the neural response version. For discussion of this point, see Harman (1996, 10), Kriegel (2009, 88), and Levine (forthcoming). For these reasons, I will focus on the superior “neural response” version, which avoids the circularity problem. Fregean representationalism about colour experience (Chalmers 2010, essay 11) is very similar to response-dependent representationalism. It holds that there is a neural experience iff one has an experience that represents that there is something with the unique property that normally causes bluish experiences. I will not discuss it separately here.

19 For presentations of this “representational problem” for the response-dependent representationalism, see Pautz (2010, 350–5), and Byrne and Hilbert (forthcoming). For a response, see section 3.5 of Jonathan Cohen’s contribution to this volume. Cohen’s response is that we certainly can represent in thought and language fine-grained response-dependent properties. For instance, I can say “consider a property of the form causing neural state, \( N1 \), in Maxwell’s population under precise viewing circumstances \( C1 \)”, and thereby refer to a property of this kind. Cohen then suggests, “if so, then it’s open to [response-dependent representationalists] to hold that whatever accounts for the mental representation of such properties [in thought and language] can also serve as an account of how states of the visual system represent the very same properties”. But this is mistaken. The correct account of my ability refer to such a fine-grained response-dependent property in thought is that I speak a language with a compositional semantics, and this language has symbols like “neural state”, “Maxwell’s population”, “precise viewing circumstances”. So I can form a complex predicate out of these terms in order to represent to a fine-grained response-dependent property. Contrary to Cohen, the same compositional, language-based account definitely cannot serve to explain the perceptual representation of these properties, for a very simple and decisive reason: the format of perceptual representation is not language-like but rather iconic (Tye 1995). Even setting this point aside, the suggestion is incredible. What evolutionary advantage would come from the visual system being innately equipped with symbols like “neural state”, “my population”, “precise viewing circumstances \( C1 \)”, and forming, whenever one enjoys a colour experience, a complex representation along the lines of “there is an object out there that causes
neural state, NI, in my population under precise viewing circumstances CI". There are other problems with Cohen’s response. For instance, he doesn’t address the “uniformity problem” mentioned in the text (see also Pautz 2010, 353).

20 For this problem, see Armstrong (1984, 170), Boghossian and Velleman (1989), and McGinn (1996). One response depends on the “neutrality thesis”. For discussion, see McGinn (1999).

21 Recent proponents of irrealist representationalism, or something close to it, include Mackie (1976), Maund (1995), Wright (2003), Pautz (2006), and Chalmers (2010).

22 For a host of empirical and a priori reasons for denying that pain qualities and other sensible properties aren’t instantiated in the extracranial world, see Pautz (2010) and Levine (2019). For instance, in the domain of speech perception, it often happens that a categorical change in the perceived sensible property (e.g. from /da/ to /ta/) corresponds to no categorical change in the objective stimulus. It is impossible to be a realist about audible qualities; they are projections of the brain.

23 For this objection, see Campbell (this volume), Papineau (2016, sect. 13), and Levine (2019).

24 A clarification: Even if irrealists must hold that the brain alone explains our ability to perceptually represent colours, they needn’t say that it explains the existence and character of the colours themselves. Instead, they might accept a “transcendent” view of colour properties on which they exist necessarily and their character is entirely mind-independent (Russell 1912, chap. IX). Compare: the brain doesn’t create numbers or the facts about numbers; it only grounds our capacity to represent numbers.

25 Indeed, on irrealist representationalism, the central puzzle about experience becomes: how can we perceptually represent uninstantiated sensible properties? To illustrate, consider Frank Jackson’s knowledge argument. In her black and white room, Mary learns all the truths about the instantiation of fundamental physical properties in our world. On irrealist representationalism, when she is released, she learns certain truths about non-physical sensible colours, for instance, that there exist these specific qualities, with specific natures or quiddities. But since these properties are uninstantiated, they are not themselves a problem for the physicalist claim that all instantiated properties are grounded in physical properties (Jackson 2004, 431). (Compare: physicalists can believe that there exists the non-physical property being a unicorn, and that we are related to it in thought, as long as they say that it is uninstantiated.) So if we accept irrealist representationalism, then we should say that the “sensible properties question” is not the real problem for physicalism. Rather, it is the “representation relation question”: how can the brain enable us to perceptually represent these particular, uninstantiated non-physical properties?

26 For discussion of the idea that relativity theory and quantum mechanics support irrealism about the spatio-temporal properties we perceptually represent (perceived simultaneity, perceived threedimensional shape), see Boghossian (2011, 56), Chalmers (2012, 296–7, 333), and Ney (2013, 177–81).

27 Contemporary naïve realism is not as flexible as representationalism. While there are internalist versions of representationalism, naïve realism is essentially externalist and requires a response-independent theory of the sensible properties. So it faces versions of the structural-mismatch problem and the internal-dependence problem from section 2 above. For discussion of this issue, see Kalderon (2011, 250–6), Fish (2013, 58–9), Logue (2016), and Allen (2017, 68–73 and chap. 6).

28 I thank Keith Allen, Derek Brown, Brian Cutter, and Fiona Macpherson for very helpful comments on this chapter.

References


Adam Pautz


Colour experience representing the world