The Autonomy of Colour

Justin Broackes

This essay* takes two notions of autonomy and two notions of explanation and argues that colours occur in explanations that fall under all of them. The claim that colours can be used to explain anything at all may seem to some people an outrage. But their pessimism is unjustified and the orthodox dispositional view which may seem to support it, I shall argue, itself has difficulties. In broad terms, Section 2 shows that there exist good straight scientific laws of colour, constituting what one might call a phenomenal science. Section 3 offers a larger view of what we are doing when we attribute colours to things, a view which makes it a case of holistic explanation, similar in many ways to psychological explanation. Section 2 emphasizes the model of scientific explanation, and Section 3 the holistic model found in rational explanation; but it will emerge that colour explanation in different ways fits both models, as it also does the two principal notions of autonomy that the first section identifies.

1. Are colours explanatorily idle?

Philosophers often say that colours are explanatorily idle. As McGinn has put it:

* For discussion and comments on earlier versions of this paper, I am grateful to David Bell, Quassim Cassam, William Child, Larry Hardin, Kathleen Lennon, Michael Martin, Peter Smith, Paul Snowdon, Helen Steward and Tim Williamson, as well as to other members of audiences where I have presented it. I owe a special debt to John Campbell: without his own writings on colour this would have been very different work, and without our many discussions, it would have been a good deal less enjoyable. To David Charles I am grateful for criticism and encouragement that go way beyond those of a generous editor.
The Autonomy of Colour

First, these qualities are not ascribed to things as part of the enterprise of explaining the causal interactions of objects with each other: colour and taste do not contribute to the causal powers of things. Primary qualities are precisely the qualities that figure in such explanations ... Secondly, secondary qualities do not explain our perception of them; primary qualities are what do that.¹

On the other hand, some philosophers have thought the contrary,² and they have painters³ and much of everyday speech on their side. We regularly hear claims like these:

(1) The red paint turned pink because he added white to it.
(2) The house gets hot in summer because it is painted black.
(3) The orange light of the evening sun made the façades of the buildings seem to glow.
(4) The yellow of a life-jacket caught his eye as he looked across the water into the distance.
(5) He stopped at the traffic lights because they were red.

Colours are invited to explain the appearance of things (3), human perception (4), action (5), and even the characteristics of non-sentient items, like the colour of a paint (1) and the temperature of a house (2).

³ The painter Philipp Otto Runge said in a letter to Goethe: ‘This has driven me on at least to study the characteristics of the colours, and whether it would be possible to penetrate so deeply into their powers, that it would be clearer to me what they achieve, or what can be produced by means of them, or what affects them.’ (Runge, as quoted in Goethe’s *Farbenlehre*, in the Zugabe that follows §920; my translation. (p. 266 in Vol. 10 of J.W. Goethe, *Sämtliche Werke*, K. Richter et al., eds. (München: Carl Hanser, 1989)))

The Autonomy of Colour

Why then deny that colours explain what they seem to? Of course colours are not fundamental physical properties, like mass and charge. But we should not need reminding that good explanation is not always explanation in the terms of basic physics. As Putnam has said, we can explain why a square peg will not fit into a round hole, by saying the board and the peg are rigid, and the round hole is smaller than the peg. An ‘explanation’ in quantum mechanics, or whatever other basic terms, would miss the relevant features. For ‘the same [higher-level] explanation will go in any world (whatever the microstructure) in which those higher level structural features [rigidity and size] are present. In that sense this explanation is autonomous.’

Why shouldn’t colours occur in explanations that are autonomous in a similar way? Autonomy means different things to different people, and it may be helpful to clarify the two main senses I shall be giving it. The first involves no more than is introduced in the last quotation from Putnam: the same explanation would go through in any world where the same higher-level properties were present. It is no part of an autonomy claim in this sense, therefore, that properties invoked in an autonomous explanation are in every way independent of properties at other levels. The squareness of the peg is, for example, supervenient on the basic physical properties and arrangement of the peg’s constituent parts. Colours similarly will be supervenient upon physical properties. In this usage, therefore, interdependence is not, as it is for Patricia Churchland, ‘autonomy’s opposite’. Colour-explanations will be autonomous in this sense if they are indifferent to the underlying realization of the property—if the same explanation would go through if the object’s redness, for example, were realized in some microstructurally different way. But that does not imply that the explanation is independent of other properties in every way. It means that it is independent of microstructural variations that would result in the same macroproperty.

In what is probably a different sense, explanations are autonomous if they rule themselves, in that they are responsible to, and to be judged by, criteria internal to that style of explanation—and

---

not by criteria from another domain. On this understanding, an autonomous explanation will typically (though not necessarily) be ‘epistemically independent’ of other explanations, in that knowledge that it meets the appropriate internal standards of success will be independent of knowledge concerning any other form of explanation. But it may still (and typically will) be the case that the explanation fails to be ‘ontologically independent’, in the sense that the higher-level causal relations would hold even if the underlying lower-level causal relations did not. In between these two senses of independence there is a third issue, of whether for any higher-level classification there has to be at a lower-level a classification that corresponds (a type-identity, or restricted type-identity); and a fourth issue, of whether every higher-level causal generalization must correspond to (or be reducible to) one or more lower-level causal generalizations. But to claim that a property figures in autonomous explanations in the sense explained at the start of this paragraph is not to claim independence in any sense stronger than the first of these four.

I shall be exploring the prospects that colours figure in explanations that are autonomous in these two senses. Why should one resist the idea? Localized error, so to speak, is of course unavoidable: individual explanatory claims are bound sometimes to turn out false—as I shall later suggest is actually the case with (4). But some people may still suspect a global error—a mistake in the very idea that colours can occur in ‘autonomous explanations’. The resistance must have a theoretical source, and I shall briefly consider four.

One might suspect that colour explanation, if there were such a thing, would compete for space with physics, and each would crowd the other out. The obvious reply is that explanatory schemes at different levels may peacefully co-exist if they stand in appropriate relations. It is widely believed that mental and physical schemes of explanation can peacefully coexist if mental phenomena are supervenient upon physical phenomena. Could not colour explanation in a similar way coexist with the physical sciences, if a corresponding supervenience relation held there too? The suggestion is plausible.

---

Though no definition of supervenience is uncontroversial, a first approximation would be: $f$-properties supervene on $g$-properties iff it is
The Autonomy of Colour

My present perception of blue, for example, would be explained by
the blueness of the mug in front of me, while the underlying visual
processes were explained by whatever physical features are relevant.
There will be no competition between the explanations, if the colour,
as seems plausible, supervenes on the physical features.

One might worry about the fact that the putative effects of colour
are primarily on humans, or, if on other things, on their colour, rather

metaphysically necessary that situations indiscernible in g-terms are also
indiscernible in f-terms, but not vice-versa.

The case for peaceful coexistence of mental and physical discourses has
been made in different ways by Davidson, Fodor and Dennett. There are
people who have argued for the elimination of the mental. But the most
notable proposals (like that of Patricia Churchland, Neurophilosophy, esp. Chs.
7-9) attack the mental scheme of discourse not on the a priori ground that
different levels of discourse can never coexist peacefully (—Churchland is
herself a defender of a plurality of levels of scientific discourse—v. p. 358), but
rather on the a posteriori ground that ‘folk psychology’ is not the best candidate
for the job it is intended for, and looks set to be replaced by an ideally
developed neuroscience (v. e.g. p. 396). So the belief in a plurality of legitimate
levels of discourse is I think not a contentious one. The second issue—raised
by the claim that folk psychology should be replaced by neuroscience—is not
one which I can pretend to discuss properly here. But since the issue lurks in
the background of discussions later in this paper, it may help to say now that
my own response would be to call into question Churchland’s view of what
the aim of mental discourse is. If its aim is not to ‘explain and predict’ in the
manner of the physical sciences, then failure at that job is not a reason to
abandon it. (For alternative views of the domain and methods of rational
explanation, see e.g. Davidson, ‘Mental Events’ and ‘Psychology as
Philosophy’, repr. in his Essays on Actions and Events (Oxford: Oxford
University Press, 1980); John McDowell, ‘Functionalism and Anomalous
Monism’, in E. LePore & B. McLaughlin, eds., Actions and Events. (Oxford:
Blackwell, 1985); K. Lennon, Explaining Human Action (London: Duckworth,
1990).) A third and different matter is Churchland’s claim that ‘cognitive
psychology’ looks unlikely to succeed in isolation from neural science (or if it
persuades itself that it is ‘autonomous with respect to neuroscience’ (p. 362)).
On this issue, Churchland’s view seems completely persuasive, if the domain
of cognitive psychology includes such questions as why we sleep, and why we
forget as much as we do. But that does not immediately lend support to the
second claim, unless the aims of ordinary mental discourse and ‘folk
psychology’ are the same as those of cognitive psychology.

For the comparison between the relation of mental explanation to physical
explanation and the relation of colour explanation to physical explanation, see
urged the same point in Justin Broackes, ‘The Identity of Properties’ (D.Phil.
thesis, Oxford University, 1986), p. 228. Not being able to argue all points at
once, the succeeding discussion presumes the peaceful coexistence of the first
pair, and considers the suitability of using that as a model for the relation of
the second pair.
than on, say, their size and shape. But that is hardly a reason to deny them causal efficacy: the primary effects of economic factors like an increase in the money supply are also on humans and other economic factors (rather than directly on the size and shape of physical objects); but we do not treat that as a reason to say they are causally idle.

A third worry might be that colours are parochial: there are totally colour-blind humans, and if they had been the only ones around, then they would hardly have felt they were missing something. But we can admit the parochiality of something without denying it causal efficacy. Economic factors, again, are parochial (there are societies without money, and a view of our own existence that makes money an irrelevance); but that does not make us deny the reality of economic causes.

Perhaps the most serious concern is that colours are dispositions, and dispositions neither cause nor explain. The issues are too complex to discuss properly here, but there are problems at each stage of the argument. First, one may doubt whether colours are in fact dispositions. I shall later be giving reasons to deny the orthodox view of them as dispositions to produce experiences in us, while suggesting that they are dispositions of a different kind. But we can certainly not simply presume that colours are dispositions of any kind, in the face of the substantial body of philosophers who have recently argued that they are not. Secondly, it is doubtful whether dispositions are explanatorily idle. There is a tradition of scoffing at explanations in terms of virtus dormitiva. But dispositions are not all like dormitive virtue (—what about the engineer’s properties of capacitance, inductance, resistance and elasticity?—), and even dormitive virtue has its explanatory uses. (The man fell asleep at the controls of the machine because he had drunk too much of a cough mixture with a dormitive virtue.) The issues can hardly even be aired here, but even if colours are dispositions, we cannot assume that they are

---

7 The effect of the black paint on the temperature of the house seems an exception to this. But we will shortly have reasons to doubt whether the claim is strictly true.
8 I discuss them more fully in my book *The Nature of Colour* (in preparation)
9 e.g. P.M.S. Hacker, *Appearance and Reality* (Oxford: Blackwell, 1987), Chs. 3 & 4; Campbell *op. cit.*; and Barry Stroud in his John Locke lectures in Oxford.
The Autonomy of Colour

explanatorily idle for that reason, any more than for the other reasons I have considered.

2. Colour laws and colour science

What are the prospects of finding good straight scientific explanations that employ colours? What can colours be used scientifically to explain? We might distinguish three possible uses of colours: to explain (i) the effects of bodies (and light) on humans and other animals (notably in perception), (ii) the effects of bodies (and light) on the colour properties of other bodies, and (iii) the effects of bodies (and light) on the non-colour properties of other bodies (like their temperature, motion or size).

The last category is the least promising. The most conspicuous cases where colour affects the non-colour properties of non-sentient things are cases where they do so only by affecting sentient beings, who in turn produce the effects in the non-sentient things. (The colour of the traffic-lights affects the motion of the cars, but only as it is seen by the drivers.) These cases therefore reduce to the first category. There may seem to be cases where colour is directly responsible for the non-colour properties of something: the warehouse walls, we may say, are heating up in the sun because they are painted black. But on closer inspection, the explanation seems to be invalidated by lower-level facts. The walls, we find out, are really heating up because they fail to reflect the infra-red light from the sun, rather than because they fail to reflect the visible light in the way that makes us call them black. It is not the blackness proper that explains the effect. Black things commonly absorb infra-red as well as visible light, so we naturally say ‘if it hadn’t been black, it wouldn’t have heated up like that’. But the counterfactual is strictly false: the house could well have been some other colour and still heated up like that (if, say, it were painted with a green paint that absorbed in the infra-red); and it could well have been black and not heated up like that (if the paint absorbed light in the visible range but not in the infra-red). The threat of invalidation by lower-level explanation may well be endemic to purported explanations of this kind. We have reason to believe that non-colour physical phenomena can in principle be explained in physical terms;
The Autonomy of Colour

we know (particularly from the existence of metamerism, described below) that colours are (often quite strikingly) variably realizable in physical terms; so it seems likely that for any physical effect produced by an object of one colour, there could be another object of the same colour that did not have that effect. The prospects, therefore, of finding laws by which colours could be treated as causes of the instantiation of other physical properties (other than via perception) seem remote—though the argument does not rule out the possibility in principle.

The initial prospects look better of giving scientific explanations on the basis of colour for the other two ranges of phenomena: human perceptions and the colours of things. Here a similar challenge arises, but in this case I think it can, at least often, be met. Quite aside from any general prejudice that the only decent explanation is explanation in the terms of physics, there is a worry that these phenomena are simply (as a matter of fact) better explained in terms of physical properties more basic than colour. The main reason for saying this is the existence of metamerism. Because of the limited sensitivity of the eye, two lights may have the same colour though their spectral composition is different. Two objects may look the same colour though the light coming from them is spectrally different, and their spectral reflectance profiles are different. This in itself is no reason to say that colours are unexplanatory: so far it seems a standard case of variable realizability. But many of the effects that we commonly ascribe to colours, turn out to be determined not by colours but by their realizations. We may say that the tomato looked brown because it was in green light. But it turns out that the colour of the object and the colour of the light are not sufficient to determine the object’s appearance: it is the spectral reflectance of the object and the spectral composition of the light that determine the character of light reflected from the object, and hence its appearance. As is well known, a shirt and a pair of trousers may match in the midday sun, but differ in fluorescent lighting. Clearly, therefore, the appearance of the objects in the fluorescent light cannot be determined by the colour of the light and the colour of the objects.10 The threat is not that there is lower-level

The Autonomy of Colour

explanation which goes deeper than the higher-level account (—that in itself would not invalidate the account—), but rather that the lower-level explanations show us that the purported higher-level account rests on claims that are just not true. It is simply *not true*, the challenge runs, that the reason the tomato looked that shade is that it was in green light of just that colour: for that shade of green illumination is neither necessary in the circumstances nor sufficient for the thing with that shade of red to look that shade of brown. A proper explanation will have to refer to the lower-level spectral characteristics of the object and the light. This threat, that the variable realizability of colours will invalidate purported colour explanations, I shall call ‘the challenge of metamerism’.

If the threat seems to be realized in the case just described, that does not mean that it is in all colour explanations. Maybe some colour explanations can defeat the challenge of metamerism, and some cannot. This will be so if some but not all colour explanation is ‘autonomous’ in Putnam’s sense: ‘the same explanation will go in any world (whatever the microstructure) in which those higher level ... features are present’.\(^\text{11}\) Some explanations do seem to meet this condition. The mug looks blue to John because it is blue and John is looking at it in decent lighting, and he has good colour vision. The claim cannot be undermined by considering objects whose blueness is realized in a different spectral reflectance profile: whatever its spectral profile, as long as the mug is blue, then it will look blue to John in the circumstances described. It is the blueness that nomologically correlates with the effect, not just some lower-lying property that happens to be coinstantiated with the blueness. Other candidates for autonomy status come to mind: the object is opaque because it is white;\(^\text{12}\) the paint is this particular green because it was mixed from paints of this blue and this yellow, in these particular proportions; the

knowledge of what has happened to the relative energy distribution of the light .... Two light sources having completely different energy distributions may look exactly alike to an observer and yet may produce entirely different colors if the light from them falls on the same object. It is apparent that no description of these lights in terms of *colors* can ever explain the situation, but knowledge of the energy distributions may make it entirely obvious.’ (my emphasis in the first sentence)

\(^{11}\) ‘Philosophy and our Mental Life’, quoted above.
The Autonomy of Colour

yellow book looks brown because it is in violet light. But are all these explanations in fact autonomous? Are they in fact immune to the challenge of metamerism? To see which are and which aren’t, it will help to review some of the attempts of colour theorists to come up with serious laws.

I shall put aside the theory of the aesthetic qualities of colours in various combinations. Working often to develop a discipline parallel to those of harmony and counterpoint in music, Alberti, Goethe, Munsell and Itten, to name only some of the more prominent, have tried to set out principles of the harmony of colours. Some of the attempts have bordered on the fanatical: Munsell, trained as a painter, conceived his system of colour notation, with its numerical measures of hue, chroma and value, as a prerequisite for the proper statement of the principles of colour harmony. ‘COLOR ANARCHY IS REPLACED BY SYSTEMATIC COLOR DESCRIPTION,’ he exclaimed in capital

---

13 Alberti, for example writes: ‘Grace will be found, when one colour is greatly different from the others near it... This contrast will be beautiful where the colours are clear and bright. There is a certain friendship of colours so that one joined with another gives dignity and grace. Rose near green and sky blue gives both honour and life. White not only near ash and crocus yellow but placed near almost any other gives gladness. Dark colours stand among light with dignity and the light colours turn about among the darks. Thus, as I have said, the painter will dispose his colours.’ (Leon Battista Alberti, Della Pittura, Book II, near end; pp. 84-85 in J.R. Spencer’s translation, On Painting (New Haven / London: Yale University Press, Revised ed. 1966).) (John Spencer’s reference in his notes to ‘the colour chords’ of ‘the Albertian colour system’ (p. 130 n. 83, cp. p. 105 n. 23), however, finds more systematicity in Alberti’s comments than I can find there.)

Goethe’s Farbenlehre, Part VI, esp. §§ 803 ff., sets out principles of colour harmony, and traces them to the eye’s ‘tendency to universality’ (§ 805). The same tendency as he uses to explain contrast effects (when the eye ‘spontaneously and of necessity ... produce[s]’ the complementary colour) is responsible also, he thinks, for our finding combinations of complementary colours harmonious. (§§ 805-7, cp. § 61)

Munsell carefully distinguishes the aesthetic characteristics of three typical paths one may take from a given colour: the ‘vertical’ path (taking lighter and darker values of the same hue), the ‘lateral’ path (changing the hue without changing either value or chroma) and the ‘inward’ path (towards the centre of the colour solid and beyond to the opposite hue). Describing their uses, he adds that the third is ‘full of pitfalls for the inexpert’. (A.H. Munsell, A Color Notation (The Munsell Color Company, 1905, 8th ed., 1936), 38).
The Autonomy of Colour

letters.  

The status of such principles of harmony is a matter of such complexity that I shall set it aside for another time.

In the field of straight experimental science, there is a fine body of explanation of colour phenomena. The work straddles areas which are otherwise often separated, like optics, quantum electrodynamics, chemistry, psychology and psychophysics. The best-known work—like that of Newton in optics, and of Helmholtz and Maxwell in psychophysics—is on the explanation of colour phenomena in terms of physics. The kinds of explanation and law that are our present concern, on the other hand, are those that explain colour phenomena in terms of other colour phenomena. From a wide possible range, I shall consider five types of law, as developed in the work of the nineteenth-century colour theorists, Grassmann, Chevreul and Rood.

One might take Newton’s ‘centre of gravity’ law of additive colour mixing to be the first straight scientific law where colours are among the explanantia. Newton offers a method to predict the colour of a mixture of lights from the colours of its component lights, using a diagram in which the colours of the spectrum are arranged on the

---

14 Munsell op. cit., 24.

15 One point, however, may be worth raising. There is a tendency to think that aesthetic responses to colour are direct and unaffected by reasoning, training and cultural influence. But this (like the ‘tingle-immersion’ view of aesthetic response in general) will not survive scrutiny. Aesthetic judgement, with respect to colour as anything else, is always open to revision, at least in its details, as a result of aesthetic experience of other situations, reflection and thought. None the less, it is remarkable that there is such a discipline as harmony and counterpoint at all—and it seems therefore that, at least with respect to what might be called a strictly delimited aesthetic language, it is possible to make rough-and-ready aesthetic judgements on the basis of rules: ‘this is not (for the language of Haydn and Mozart) discordant’, ‘the chord progressions of this chorale harmonization are more or less in the style of Bach’. Something similar in the case of colour seems promising: you cannot tell in advance that this particular colour combination will never, in any context, look good. (Maybe the four colours look terrible together on the walls of a room, and awful used in a particular textile design for a dress. But when employed in a particular way on a book jacket, they suddenly look right.) It is a matter of continual artistic discovery that things that people once assumed would never look right or sound right, can suddenly begin to do so, employed in a new way—a way which may itself change the artistic or musical language. But we can none the less come up with limited rules of thumb: for example that any employment of this combination of colours, if seen pretty much with present eyes, will look a bit unpleasant.

circumference of a circle, and white (the colour of the mixture of all of these kinds of light) lies at the centre. If you consider the points on the colour circle that represent the component spectral lights in the mixture, and assign to each of them a weight proportional to the intensity of light of that kind, then the centre of gravity of the resultant figure will represent the colour of the mixture of lights, as illustrated in Figure 1.

Is this an autonomous law of colour, specifying the colour of a mixture in terms of the colours of the components? Newton did not, I think, intend it quite like this: though he talks of predicting the colour of the mixture from ‘each Colour in the given Mixture’, he is using ‘colour’, I think, really for refrangibility, or as we would now say, wavelength. He is giving a rule to allow us, from the relative amounts of light at each wavelength, to predict the colour of the resulting mixture. But Newton’s rule can be transformed into an autonomous law of colour-mixture, given one of the empirical laws of colour which the German mathematician Grassmann proposed in the mid-19th century. As a matter of empirical fact, the results of mixing, say, a green light and a blue light of a particular hue and saturation will be perceptually indistinguishable, whatever the spectral composition of the two lights—however the green and blue are ‘realized’. To generalize: ‘Two colours, both of which have the same hue and the same proportion of intermixed white, also give identical mixed colours, no matter of what homogeneous colours they may be composed.’

So we can take colours rather than wavelengths as the explanantia in Newton’s colour mixture law. In the laws on the results of mixing coloured lights, therefore, the ‘challenge of metamerism’ can be met: in so far as Grassmann’s Third Law is true (and that is within wide limits) any variation in the spectral composition of the lights

---


Some qualifications should be noted. Maxwell’s triangle and the subsequent empirical research leading to the 1931 CIE x, y chromaticity diagram showed that in place of Newton’s circle, the locus of spectral points in this kind of mixing diagram needs rather to be in the shape of a plectrum or tongue. Secondly, each point on a modern chromaticity diagram represents a certain hue and saturation, but brightness is not taken into account, as it needs to be in a full explanation of additive mixing. (For further details, see e.g. R.W.G. Hunt, Measuring Colour (Chichester: Ellis Horwood, 1987), 58-60.)
which does not change their colour will not change the colour of the mixture. Explanation of the colour of the mixture in terms of the colour of the lights combined is, in Putnam’s terms, autonomous.

Fig. 1: Newton’s colour wheel: to predict the colour of mixtures of light

The circumference DEFGABCD represents ‘the whole Series of Colours from one end of the Sun’s colour’d Image to the other’. Let \( p, q, r, w, t, v, x \) be the ‘Centers of Gravity of the Arches’ DE, EF, FG, GA, AB, BC, and CD respectively; ‘and about those Centers of Gravity let Circles proportional to the Number of Rays of each Colour in the given Mixture be describ’d’. ‘Find the common Center of Gravity of all those Circles, \( p, q, r, s, t, v, x \). Let that Center be \( Z \); and from the Center of the Circle ADF, through \( Z \) to the Circumference, drawing the Right Line \( OY \), the Place of the Point \( Y \) in the Circumference shall shew the Colour arising from the Composition of all the Colours in the given Mixture.’ The ratio of \( OZ \) to the radius of the circle gives the relative saturation of the colour. (From Newton, *Opticks*, 154-5.)

In the mid-nineteenth century, Chevreul, a chemist working at the Gobelins tapestry factory, tried to formulate other scientific laws of colour. Out of the many which he offered in his *Principles of Harmony and Contrast of Colours*,\textsuperscript{18} I shall consider three important types: laws on the mixing of coloured pigments; laws for what appearance is produced when light of one colour falls on objects of a second colour; and laws of colour contrast, either simultaneous or successive. Two of these turn out, I think, not to meet the initial challenge of metamerism.

The mixture of coloured pigments is today called subtractive mixing, because each pigment may be thought of as subtracting a certain amount from the incident light in the process of (only partially) reflecting it. Crudely considering only red, green and blue components of light, we may say that a yellow object reflects only the red and the green (and filters out the blue), a blue object reflects only the green and the blue (and filters out the yellow). If we paint therefore a layer of blue paint on top of a layer of yellow on some white paper, then the only light that is reflected is the light that can pass so to speak through both filters, namely, the green light. The effect of mixing two pigments subtractively is the same as that of superimposing two filters, so we seem to have an explanation of why, when we mix yellow and blue paint, we get green.¹⁹

The trouble is, however, that this kind of explanation is not well equipped to resist the challenge of metamerism. It turns out that the colour of a pigment mixture is not determined solely by the colours and quantities of the components, but also by their spectral reflectance characteristics. It was a simplification to think of the filtering action of a blue pigment as simply removing the yellow. Two pigments of different spectral characteristics may look the same colour in a certain form of daylight. But now imagine we also have a coloured celluloid filter, and look at the metameric samples through it. It may well happen that, through the filter, the samples now look different from each other.²⁰ And the situation is essentially the same if, instead of having a celluloid filter in front of the eyes, we paint a layer of colour on top of the sample, as in subtractive mixing of

---

¹⁹ In colour printing, if dots of different colours are printed on top of one another, the mixing can be considered subtractive; if dots too small to be individually seen are printed next to one another, then the mixing can be considered additive. Some forms of colour mixing are neither precisely subtractive nor precisely additive.

²⁰ Mathematically, to get the character of light reaching the eye we would (at each wavelength $\lambda$) multiply the power of the illuminant at $\lambda$ by the reflectance of the sample at $\lambda$ and then by the transmittance of the filter at $\lambda$. Now (as far as the character of light reaching the eye is concerned) it does not make any difference whether the filter comes between the illuminant and the object, or between the object and the eye. (The two cases may be perceptually different, if the context results in different adaptation.) Since we already know that two metameric samples may become distinguishable in colour when the illumination is varied, it is not surprising then that the same happens when the samples are seen through a filter.
The Autonomy of Colour

pigments. The result is, then, that we cannot pretend that it is colour alone that determines the outcome of subtractive mixing: there may be differences between the result of mixing one yellow paint with blue and the result of mixing another yellow paint with it, even if the two paints look the same colour in daylight. In this area, it seems, rough-and-ready generalizations are all we can hope for.21

Chevreul extensively studied the effects of seeing an object of one colour in light of another colour. Pages of his *Principles of Harmony and Contrast* are filled with experimental generalizations like these:

Yellow rays falling upon a Black stuff, make it appear of a Yellow-Olive.

Yellow rays falling upon a White stuff, make it appear of a light Yellow. ...

Yellow rays falling upon a Blue stuff, make it appear Yellow-Green, if it is light, and of a Green-Slate if it is deep.

(Chevreul, *Principles*, 92)

Violet rays falling upon a Yellow stuff, make it appear Brown with an excessively pale tint of Red.

(Ibid., 94)

Unfortunately, these laboriously-collected generalizations are simply not reliable. Recognizing some of their shortcomings, theorists later in the nineteenth century (like Ogden Rood) tried to amend them. But there is a problem in principle: the existence of metamers makes it impossible to predict accurately the apparent colour of an object of one colour seen in light of another. There will be metameric yellow objects, for example, that in daylight are indistinguishable in colour, but in the same violet light no longer match each other. It therefore

Fred W. Billmeyer & Max Saltzman, *Principles of Color Technology* (New York: John Wiley, 2nd ed. 1981), 130, gives a remarkable CIE diagram (after Johnston 1973) showing the results of mixing various paints with increasing amounts of titanium dioxide white. The colour changes often show up as lines that are anything but straight, and in some cases the result of adding a small quantity of white is to move the colour not in the direction of white but at 90° to it.

In certain circumstances, there will also be chemical reactions in colorant mixing, and then clearly no mixing rule that attended only to the colours of the colorants and not to their chemistry could hope to account for the results.
The Autonomy of Colour

cannot be simply the colour of the object and the colour of the light that determine the appearance. I shall return to the question whether this conclusively scotches claims of lawlike autonomous status for all colour generalizations of this sort, but the initial challenge of metamerism has not been met.

The fourth kind of law which I shall consider concerns colour contrast. The appearance of a region of colour is affected both by the colours of objects that were in the field of view a short time before, and by the colours of other objects which are in the field of view at the same time. After-images are dramatic examples of the first phenomenon. Leonardo noted cases of the second: ‘white garments make the flesh tints dark, and yellow garments make them seem coloured, while red garments show them pale.’\(^{22}\) The phenomena were studied widely in the eighteenth century,\(^ {23}\) but it was Goethe and Chevreul who made them famous. Goethe explained that in both cases, it is ‘the colours diametrically opposed to each other which reciprocally evoke each other in the eye’. (*Farbenlehre*, § 50, my emphasis) Each colour evokes its complementary—hence the names *successive contrast* and *simultaneous contrast*.\(^ {24}\) An object seen as red will leave a green after-image, one seen as turquoise will leave a


\(^{23}\) Chevreul (*Principles*, Pt. I, Sect. II, Ch. II) acknowledges that Buffon (1743) noticed examples of both types, and also mentions Scherffer (1754), CErpinus (1785), Darwin (1785) and Count Rumford (1802).

\(^{24}\) Chevreul gives the impression that he invented the terms. (In *Principles*, 374, he talks of earlier writers who lacked ‘the fundamental distinction which I had made between two sorts of contrast under the names of simultaneous contrast and successive contrast of colours’.) But Goethe had grasped the distinction quite clearly (e.g. in the *Farbenlehre*, § 56).

I have not seen any reference in Chevreul to Goethe’s work, and I do not know what the relation between them was. But there is a fundamental agreement between them, notably on the universality of these contrast effects (Chevreul: ‘every colour seen simultaneously with another, appears with the modification of an accidental colour’ (*Principles*, 376, my emphasis); Goethe § 51) and in the view that an understanding of contrast is the foundation for understanding the laws of colour harmony (Chevreul, 376-77; Goethe §§ 60-61, §§ 805-7.). This is why both of them attack earlier writers’ descriptions of contrast effects as ‘accidental colours’ or ‘adventitious colours’ (Goethe §§ 1-2, Chevreul 376), whereas in fact ‘they are the foundation of the whole doctrine’ (Goethe §1).
The Autonomy of Colour

yellow one. The laws here are precisely ones of colour: the colour of the after-image is the complementary of the colour of the original appearance, and is not affected by the spectral realization of that colour. The robustness of these laws is easy to understand given a knowledge of the causal processes involved: given that the cones in the retina respond in indistinguishable ways to two metameric red objects, and that our after-images and simultaneous contrast effects causally depend on our retinal responses, we would expect metameric red objects to be indistinguishable in their contrast-effects. The challenge of metamerism in this case is therefore fully met.

Now for the final type of law. Ogden Rood worked on all forms of colour science at Columbia in the 1870’s, and his Modern Chromatics is still well worth reading. He offers one type of causal colour generalization that we have not considered before: describing how hue changes with saturation. One of several tables details ‘the effects of mixing white with coloured light’:

<table>
<thead>
<tr>
<th>Name of colour</th>
<th>Effect of adding White</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vermilion</td>
<td>More purplish</td>
</tr>
<tr>
<td>Orange</td>
<td>More red</td>
</tr>
<tr>
<td>Chrome-yellow</td>
<td>More orange-yellow</td>
</tr>
<tr>
<td>Pure yellow</td>
<td>More orange-yellow</td>
</tr>
</tbody>
</table>

25 One description of the phenomenon in the Farbenlehre is almost as remarkable as the event it describes: ‘I had entered an inn towards evening, and, as a well-favoured girl, with a brilliantly fair complexion, black hair, and a scarlet bodice, come into the room, I looked attentively at her as she stood before me at some distance in half shadow. As she presently afterwards turned away, I saw on the white wall, which was now before me, a black face surrounded with a bright light, while the dress of the perfectly distinct figure appeared of a beautiful sea-green.’ (Goethe Farbenlehre, §52, p. 22).

26 It is worth noting however, as Goethe and Chevreul did not, that after-image complementaries are not always the same as mixture complementaries: the colour of the after-image produced by a coloured light may not be the same as the colour of a second light that when mixed with the first yields white. Perhaps unfortunately for Goethe, the eye’s ‘tendency to universality’ (Goethe § 805) is not entirely precise. See M.H. Wilson & R.W. Brocklebank, ‘Complementary hues of after-images’, Journal of the Optical Society of America, 45 (1955), 293-99.

The Autonomy of Colour

Greenish-yellow  Paler (unchanged)
Green           More blue-green
Emerald-green  More blue-green
Cyan-blue       More bluish.
Cobalt-blue    A little more violet
Ultramarine (artificial) More violet
Violet         Unchanged
Purple         Less red, more violet.

Rood, Modern Chromatics, 197

The general formula covering these effects is this: ‘when we mix white with coloured light, the effect produced is the same as though we at the same time mixed with our white light a small quantity of violet light.’ (op. cit., 197) And this generalization again seems able to meet the challenge of metamerism. The process of mixing white with a light is one of additive mixing, which we have seen already is unaffected by change of spectral composition. So if one orange of one spectral type turns redder on the addition of white, then a metameric orange of another spectral type will do so too.

Of the five types of law considered, three immediately meet the challenge of metamerism: the laws of additive mixing, of colour contrast, and of change of hue as a colour is desaturated. Only the first is readily expressed in mathematical terms (when developed, for example, in connection with a CIE $x$, $y$ diagram). But they are all of them robust in the sense that substitution of metamers will not invalidate them. We can accept without worry therefore the autonomy in Putnam’s sense of a wide variety of explanations that invoke colour.

Two of the types of law considered fail the test for robustness. Laws of subtractive mixing and of the appearance of objects seen in light of another colour are at risk of being invalidated by substitution of metamers. But the challenge may be being exaggerated. The existence of metamers certainly means that we will not be able to find absolutely precise and indefinitely refinable generalizations on these issues that are also true. But that does not mean that there cannot in this area be broad generalizations which are quite literally true and lawlike in the sense that they are supported by their instances and
support counterfactuals. (There is a difference between a narrow claim that is only partly true, and a broad claim that is precisely true.) Though the appearance of a yellow object in violet light may not be determined \textit{exactly} by the colours of the object and the light, it may still be determined \textit{roughly} by them, for example, to lie somewhere in the brown region. Similarly, even if we cannot predict the exact appearance of a mix of yellow and blue paint, we may none the less be able to predict that it will be some form of green. And if this is so, then explanations that use precisely these deliberately broad terms will still be autonomous in Putnam’s sense.\footnote{Davidson makes a telling admission when discussing psycho-physical generalizations. He admits that, notwithstanding his denial of psycho-physical laws, it would be embarrassing to deny that there are any ‘inductively established correlations between physical and psychological events.’ (The Material Mind’, in his Essays on Actions and Events, at p. 250.) ‘The burned child avoids the flame’ is his example; one even less susceptible of counter-example might be: ‘A conscious person with no otherwise detectable abnormality, holding a hand in a flame, will begin to feel pain’. Davidson’s comment is that such generalizations ‘are lawlike in that instances make it reasonable to expect other instances to follow suit without being lawlike \textit{in the sense of being indefinitely refinable}.’ (‘Mental Events’, 224, my emphasis.) If what Davidson says here is right, then the broad generalizations about colour contemplated at this point in the main text will be ‘lawlike’ only in the weaker of two senses of the term. But Davidson’s stronger sense of ‘lawlike’ may itself be anomalous. Why should being lawlike in any sense be a matter of being ‘indefinitely refinable’ or ‘sharpen\[able\] without limit’? Davidson’s original explanation of the notion was that ‘Lawlike statements are general statements that support counterfactual and subjunctive claims, and are supported by their instances.’ (‘Mental Events’, 217) But it is not clear that this notion of law has any internal connection with that of indefinite refinability. If this is right, then Davidson’s denial of psycho-physical laws is rather weaker than at first appears (strictly all it denies is \textit{indefinitely refinable} psycho-physical laws, \textit{sharpenable without limit}); and his argument for the identity theory would require not just the Nomological Character of Causality as naturally interpreted, but the stronger principle that ‘events related as cause and effect fall under strict deterministic laws, \textit{indefinitely refinable and sharpenable without limit}’. (‘Mental Events’, 208, amended by the addition of the words in italics.) A further difficulty with Davidson’s view comes to the surface here: if on his view there are two senses of ‘lawlike’, and it is in only one of these that he wishes to deny the existence of lawlike psycho-physical connections, then it is odd that the argument for that denial actually seems to make no play with the difference between those two senses. To rephrase the point: if the criterial and evidential differences between the nature of mental and physical discourse do not rule out the existence of lawlike psycho-physical connections in the weaker sense, it is not clear that they can rule out the existence of lawlike psycho-physical connections in the stronger sense—given that Davidson’s argument for the latter claim shows no obvious sensitivity to the difference between these two senses.}
The Autonomy of Colour

How far exactly this attempt can succeed in rehabilitating candidate colour laws is an empirical matter. If we employ sufficiently broad colour classifications, then it will be fairly easy to find true generalizations about them. If we employ colour classifications that are too narrow, then it will be easy to find metameric counter-examples. How broad the classifications must be in order to be ‘sufficiently broad’, is an empirical question to which I do not know the answer. But the apparent availability of broad generalizations suggests that the challenge of metamerism is not an objection in principle to the idea of lawlike colour generalizations, even in those cases where colour does not determine all the colour properties at issue in the outcome. Even if the generalizations are broad and intrinsically incapable of indefinite refinement, they may still be lawlike, unobvious and non-trivial.

In this section we have considered the prospects of good scientific laws about colours. The challenge to be met was not that lower-level explanation would be available as well (for that is innocuous), but that the proposed colour laws would actually be invalidated by metamerism: colour effects could only be explained in terms of information at the level of wavelengths, rather than colours. We considered five types of law as examples. Three of them conclusively meet the challenge: laws of additive mixture, of colour contrast, and of change of hue with saturation. Two of them on the other hand do not: laws of subtractive mixture, and of how objects of one colour appear in light of another. But even in these cases, there are still prospects at least of broad generalizations that are lawlike, and of correspondingly broad causal explanations. The case for the autonomy of at least some scientific colour explanations seems complete.

3. Colour interpretation

We raised earlier the possibility that colours might have their home in a holistic interpretative scheme like that of rational explanation. The scheme might even be an extension of the scheme of rational explanation. This would certainly seem right if the traditional dispositional thesis were correct. If colours were dispositions to
produce experiences in us, and if experiences were ascribed to people in the course of interpreting them as (more or less) rational creatures, then colours would figure in an extension of rational explanation. One striking consequence would be that a second form of autonomy, mentioned in section 1, would then seem to attach to ascriptions of colour. On one familiar picture, the ascription of psychological states is autonomous in the sense that it is responsible only to assessment by standards internal to that form of explanation. In one form of this view, set out by Davidson and developed by McDowell, psychological explanation is explanation of a special sort, namely ‘rational explanation’, which has different aims from those of broadly ‘physical’ explanation. The ascription of psychological states has ‘its proper source of evidence’ only ‘in terms of the vocabulary of the propositional attitudes’. If colours were simply dispositions to produce psychological occurrences, then we might expect the former to inherit the same sort of autonomy as attaches to the latter—though the line of inheritance might not be direct. Though of course colours would not themselves be psychological states, they would in a sense be ‘offshoots of the psychological’.

This is a powerful picture, but things are not that simple. Though the main figures are already recognizable, much will need to be repainted. The traditional dispositional thesis fails to make sense of some of our colour attributions, and this shows, so I shall argue, that we have a deeper conception of the nature of colours—namely, as (in the case of surfaces) ways in which objects change the light. This might threaten to remove any essential connection of colour with the psychological, but that depends on how exactly we characterize the ‘ways’ in which objects change the light. The ascription of colours does indeed prove to be part of the implementation of a scheme of explanation of everyday experience. That, by itself, might be said

---

29 ‘To recognize the ideal status of the constitutive concept of rationality is to appreciate that the concepts of the propositional attitudes have their proper home in explanations of a special sort: explanations in which things are made intelligible by being revealed to be, or to approximate to being, as they rationally ought to be.’ By contrast, what might roughly be called ‘physical’ explanation is characterized as ‘a style of explanation in which one makes things intelligible by representing their coming into being as a particular instance of how things generally tend to happen.’ McDowell, ‘Functionalism and Anomalous Monism’, op. cit., 389, my emphasis.
30 Davidson, ‘Mental Events’, 222, 216.
equally of other characteristics of external objects, like their size and shape, and does not guarantee that the qualities ascribed are themselves interestingly ‘psychological’. Nonetheless, it will turn out to be not just a prejudice that there is something both ‘relative’ and ‘subjective’ about colours which puts them in a different position: they really are connected with the psychological in a way that size and shape are not. The view of psychological ascription as the employment of a scheme of rational interpretation will turn out to need qualification; but the claim to autonomy in the special sense of this section will survive.

3.1 Away from the dispositional thesis

I shall set aside criticisms of the dispositional thesis on grounds of circularity or triviality, since these are not, I think, conclusive against versions that make no pretence at a reduction of colour. There is a more telling criticism, that touches even the truth of the coextensiveness claim that something is red, yellow or whatever iff it would look red, yellow or whatever to normal observers under normal circumstances. The problem is this. There are, we may imagine, killer yellow objects that kill anyone who looks at them. Far from having a disposition to produce experiences of yellow in normal observers, they have a disposition to end all experience in them whatever. A defender of the dispositional thesis might insist: ‘Such an object would still look yellow if only it could be seen by normal observers under normal circumstances.’ But this does not need to be true: there is a difference between the nearby possible worlds in which an object is visible, and the nearby possible worlds in which it is visible and also has its actual present colour. Imagine a situation where there are a lot of killer yellow objects around, but we have learnt to deal with them: they tend to be small, they emit a

---

31 There are good reasons for adding ‘actually’ and perhaps also ‘now’ operators to this. But I shall not go into these here, and they are independent of my present concerns.

32 I have heard the killer yellow example ascribed to Saul Kripke and Michael Smith. I do not know how either of them has used the example however, and I can only apologize if I have distorted it or omitted the best bits.
distinctive bleeping sound, and (taking care not to look at them) we can easily cover them with thick black paint which the death-rays cannot penetrate. In such a situation the sentence ‘If only the object could be seen ..., then it would look yellow’ will be false, simply because, if only the object could be seen, it would be covered in black paint.

The natural response is to plug the gap: instead of saying ‘If only it could be seen ...’, the dispositionalist will say ‘If only it could be seen without changing its colour...’. This is progress, but not enough. There will be situations where an object can be seen and its colour is unchanged, but the colour itself cannot be seen. It may be a very short and faint flash of light; it may put our cones out of order but not the rods; it may be visible but only through a dark filter (and otherwise the death-rays get us). It looks as though we have to ensure not only that the object is visible and its colour unchanged, but that the colour itself is visible. We seem to have arrived at this:

\[(D) \quad x \text{ is yellow iff, if only } x \text{'s colour could be seen by a normal observer } y, \text{ then } x \text{ would look yellow to } y.\]

This is true but trivial. It is tantamount to:

\[(D2) \quad x \text{ is yellow iff, if only a normal observer } y \text{ could see what colour } x \text{ is, then } y \text{ would see } x \text{ to be yellow.}\]

And the same of course could be said of other properties too:

\[(D3) \quad x \text{ is a dog iff, if only a normal observer } y \text{ could see what kind of animal } x \text{ is, then } y \text{ would see } x \text{ to be a dog.}\]

and even:

\[(D4) \quad x \text{ is a piece of platinum iff, if only a normal observer } y \text{ could see what kind of substance } x \text{ is, then } y \text{ would see } x \text{ to be a piece of platinum.}\]

Of course there is a difference between these claims. The ‘if only’-clauses become progressively harder for humans to satisfy: we can
often tell at sight whether something is yellow, perhaps less often whether it is a dog, and very seldom whether it is platinum. But by the time the dispositional thesis has been reduced to (D1), the interest lies not in what it says, but in any surrounding commentary that can be given to explain what part it plays in our thought about colours, and what else plays a part in addition.

A deeper problem should also be evident. The puzzle cases are ones in which a normal observer cannot in fact see the colour (or chemical composition) of the object, but (struggling to maintain a version of the dispositional thesis) we insist on talking about how he would see the object if only he could see its colour (or chemical composition). But obviously if in any particular case we take a view on such counterfactual matters, this can only be because we have already—quite independently of these dispositional theses—taken a view of what is necessary for an object to count as yellow or whatever. (In the parallel case, if we say, ‘if only he were able to see what substance it is, he would see it as platinum’, then this can only be because we know already, for reasons other than its appearance, that the sample really is platinum.) If, therefore, we make sense of these cases of killer yellow, it can only be because we have—independently of the dispositional thesis (D1)—a view of what it takes for something to be yellow. We must have, therefore, a conception of the nature of colours, just as we have a conception of the nature of platinum and of other chemical substances.

What is this conception? What is the common factor between a visibly yellow object and one that is yellow but not visibly so? We cannot say simply: the primary qualities of the objects and their parts. Something that had exactly the same primary qualities (of all constituent parts) as a killer yellow object would kill people just as surely as the killer yellow object does. (Interestingly, even an ordinary yellow object in the dark cannot be said to have exactly the same primary qualities as a similar object in the light: a yellow book

---

33 I bracket here questions of what Putnam has called the Division of Linguistic Labor. It may be that individual users do not need such a conception, if others in the community, to whom they are prepared to defer, have one. So I talk here of what ‘we’ need. Our conception may itself be ‘obscure and relative’—like a definite description of which we only later identify the satisfier.
The Autonomy of Colour

cover, for example, will in the light be absorbing photons and emitting photons, which it will not be doing in pitch darkness. If in the dark it were literally in the same primary quality state as it is normally in when illuminated, then it would be glowing! If it has to be some physical property, then the common factor to all yellow objects will at best be some subset of primary qualities, or some relatively complex high-level primary quality. This of course is the physicalist proposal of Armstrong, much criticized in the recent writings of Hardin and Westphal. It turns out that it is much harder to find a physical property common and peculiar to yellow things than one might imagine. There plainly is no one structural property responsible for the yellowness of all yellow objects. Neither is there any simple physical characteristic common to the light they give off, even when normally illuminated. (They will not, for example, merely give off light from the ‘yellow’ part of the spectrum: something that only reflected light of wavelengths from 565 to 575 nanometres would look so dark it would be black.) But on the other hand the suggestion that there might well be no physical property at all common and peculiar to yellow things is an odd one: for more than a century, psychophysics has investigated the physical characteristics of things that look yellow without obviously wasting its time. The idea of building a machine to tell the colour of any object from a ‘purely physical’ specification of it is hardly in the same class as the idea of building one to identify sentimental poems or baroque façades.

There are both phenomenal and non-phenomenal elements in our conception of colours; the difficulty is to see their interconnection. At one extreme there is the temptation to recognize only the phenomenal element: to say, for example, as the dispositional thesis does, that being yellow is solely a matter of how a thing looks to people. But we know from killer yellow that this is not true. At the other extreme there is the temptation to recognize only the non-phenomenal element: to say, for example, that being yellow is solely a physical

The Autonomy of Colour

property.\textsuperscript{35} But apart from the difficulty of saying how ordinary observers could (on this conception) have any confidence in their own ability to tell the colour of something at sight (—why should unaided perception be a reliable guide to a ‘physical property’ of yellowness any more than it is to alkalinity, or to being made of platinum?—), there is the difficulty of identifying the ‘physical property’ in question, and in offering an account of why we should wish to dignify with the title ‘yellow’ the items that have that physical property. A promising compromise is to make colour a non-phenomenal property identified phenomenally: ‘yellowness is picked out and rigidly designated as that external physical property of the object which we sense by means of the visual impression of yellowness.’\textsuperscript{36} This instantly accommodates killer yellow: killer yellow objects have the same underlying non-phenomenal property as those which produce impressions of yellow, though they do not themselves produce that impression because of the death-rays they give off as well.

But we should not let our minds be narrowed by the thought that the only range of properties available for our fundamental conception of colour is the range of physical properties. There are, at least at first sight, many properties that are not physical properties: mental, economic, functional, aesthetic and moral properties, to cite only a few of the more striking (and perhaps overlapping) categories. Of course philosophers will take different attitudes to the apparent diversity: some will accept it wholeheartedly; others will deny it outright, insisting that there are no genuine classifications other than those of physics. In the middle will be people who accept certain

\textsuperscript{35} Abstracting from a number of contentious issues, I shall offer as a first elucidation of the notion of a physical property: ‘a property that can be introduced in the language of physics.’

\textsuperscript{36} Saul Kripke, Naming and Necessity (Oxford: Blackwell, rev. ed. 1980), 128 fn. 66. I take it here that Kripke means by ‘physical property’ a property introduced by the predicates of physics. He also says, however, that yellowness is ‘a manifest physical property of an object’ (128 fn. 66; cp. 140 fn.71), and it would be possible to take ‘manifest physical properties’ to be something other than properties of physics: perhaps physical in a broad sense (part of the natural world) but also ‘manifest’ in the sense, \textit{fully open to view, fully grasped by anyone who understands the term}, so that colour would be a phenomenal property rather than one of physical science. This would make the view closer to the ‘simple view’ of John Campbell, the non-reductive realisms of Hacker, Stroud and Putnam, and the view which I defend myself.
ranges and try to reduce the others. To assure one’s title to any of these views is not the task of a paragraph or two. The pressure towards a pluralism of types of property will be great for anyone impressed with Strawson’s and Grice’s observation\(^\text{37}\) that where there is agreement on the use of expressions with respect to an open class, there must necessarily be some kind of distinction present—unless one also believes that every distinction made with any predicate can be identified with a physical distinction, quite unrestrictedly\(^\text{38}\) and without massaging the extensions of the terms. The arguments that this kind of physicalist property-identity is unlikely to be available are known well enough from the works of Davidson and Fodor not to need repeating here. Accepting a pluralism of properties, the principal challenge is, then, I think to show how this pluralism is compatible with a belief in the fundamentality of physics. Again, I think the essential moves have been made by Davidson and Fodor, showing that the irreducibility of, for example, mental to physical classifications is no obstacle to a broader physicalism that treats the former as supervenient upon the latter.

None of these views is uncontroversial. But given the _prima facie_ availability of a view of mental properties of this general type, the possibility is open of regarding colours in a similar way: as properties that figure in an autonomous explanatory space (or subspace), irreducible to physical science, but supervenient upon it. Colours would figure in a distinctive form of colour explanation. The concerns


\(^{38}\) Note that even those who (like Lewis) advocate restricted type-identity must (if Grice and Strawson are right) allow that there is some property common, say, to Martians in pain and humans in pain, even if it is not what these theorists call ‘pain’. This may be the property of being in a physical state which plays the role of pain for that kind of organism, and we can call it a second-order physical property if we like. But it should be noted that it is not definable in the language of physics alone, unless the specification of the higher-level _functional role_ is itself definable in the language of physics—which there is reason to doubt.

Some would take refuge at this point in a metaphysically charged use of the word ‘property’, designed not to apply to just any classification that people are capable of coherently effecting. But that would still leave the problem of making sense of a plurality of classifications, to replace that of making sense of a plurality of properties.
The Autonomy of Colour

of colour attribution would be different from those that govern the attribution of other ranges of property, though there would no doubt be points of contact.

3.2 Colour interpretation

To say this is not to be relieved of the burden of doing something more to characterize the nature of the properties that colours, thus conceived, are, or to characterize the scheme of explanation which equips us thus to conceive them. I shall take the latter question first, and the former in the section that follows.

On the present suggestion, colours are in some ways parallel to mental properties, figuring in their own explanatory discourse. It may help, therefore, to employ some of the same techniques to elucidate colour discourse as have elsewhere been used to characterise mental discourse. In Davidson, and in a different way, in functionalists like Lewis and Shoemaker, one can find the suggestion that psychological discourse is the employment of an explanatory theory governed by certain constitutive principles. To characterise that discourse, we need to articulate some of the constitutive principles, and make clear how the overall explanatory scheme is applied to empirical cases.39

In Davidson’s procedure of holistically interpreting either aliens or fellows, we make use of \textit{a priori} principles which, for example, link belief, desire and action, or (in another famous example) belief, meaning and holding true. In Lewis and Shoemaker, there is a parallel idea that each psychological state can be characterized by the role it plays in a functional organization implicit in the platitudes and other \textit{a priori} claims of common-sense psychology. What will be the

39 There are of course notable differences in these philosophers’ attitudes to the constitutive principles. Lewis and Shoemaker, for example, believe that the significant interrelations between psychological states can be captured in \textit{topic-neutral} causal terms (which yet are sufficient to form an individuating description of each such state), whereas Wittgenstein would not: the constitutive principles might themselves embody a particular ‘point of view’, the grasp of which was not available to just anyone.

The idea of treating colour discourse in parallel to mental discourse can be found in Wittgenstein: ‘The colour concepts are to be treated like the concepts of sensations.’ (\textit{Remarks on Colour}, III.72)
parallel principles for colour explanation? Some of them will indicate relations between the various colour concepts; others will indicate their relations to the phenomena which they are used to explain.

Perhaps the fullest articulation of such principles is provided in Wittgenstein. In the Remarks on Colour, he considered a variety of propositions which seem to characterize a kind of 'logic' of colour. 'Pure yellow is lighter than pure, saturated red, or blue.' (III.4) 'A shine, a "high-light" cannot be black.' (III.22) 'Yellow is more akin to red than to blue.' (III.50) '[We don't] speak of a "pure" brown.' (III.60) 'Grey is between two extremes (black and white), and can take on the hue of any other colour.' (III.83) 'Black seems to make a colour cloudy, but darkness doesn’t.' (III.156) The source of many of the most interesting remarks is the letter from the painter Runge, which Goethe reproduced in his Farbenlehre—which in turn Wittgenstein had before him as he wrote some of the Remarks.40 'If we were to think of a bluish orange, a reddish green or a yellowish violet, we would have the same feeling as in the case of a southwesterly northwind.' 'Both white and black are opaque or solid.' 'White water which is pure is as inconceivable as clear milk.' (Runge’s letter, sections 2 and 11; cp. Wittgenstein, III.94.) 'Black makes all colours dirty, and if it also makes them darker, then they equally lose their purity and clarity.' (Runge’s letter, § 5) 'The opaque colours lie between white and black; they cannot be either as light as white or as dark as black.' (Runge’s letter, § 12, my translation) A particularly important claim in Wittgenstein links the notions of surface colour and film colour: 'Something white behind a coloured transparent medium appears in the colour of the medium, something black appears black.'41

Wittgenstein’s Remarks contain a fascinating discussion of the ways in which our mastery of such principles depends upon our

40 It is a great pity that the letter is omitted from Eastlake’s translation, and even from some German collected editions of Goethe’s works. The letter follows section 920 of the Didaktischer Teil (Didactic Part) of Goethe’s text (Munich edition, pp. 264-271).

41 Wittgenstein Remarks on Colour, III.173; cp. Goethe Farbenlehre, § 582. Wittgenstein may have unconsciously remembered Goethe’s comment, or he may have independently rediscovered it. In general Wittgenstein seems to have paid more careful attention to Runge’s letter than to the rest of Goethe’s text.
natural capacities and our innate endowment—as also, we might add, does our mastery of parallel principles connecting psychological states. A fuller account of colour classification than my own would have to investigate this issue. But for present purposes, it will be sufficient to draw out other types of parallel between colour explanation and psychological explanation.

The ascription of psychological states to people is part of the holistic explanation of behaviour; similarly, the ascription of colours to things is part of the holistic explanation of our perception. To describe just one small element in the picture: rather as beliefs and desires (in a certain context) produce action, the colours of objects, the lighting conditions, and the presence of observers together (in a certain context) produce perceptions of colour. The explanation is holistic: one and the same perception may have been produced by say, a blue object in white light, or by a white object in blue light, just as one piece of behaviour may be caused by alternative combinations of beliefs and desires; and only the accumulation of evidence can allow us to choose between such alternatives. Most importantly, the mastery of the holistic scheme depends on mastery of a range of a priori principles which together constitute a kind of theory. It is of course a posteriori that any particular system can be interpreted as having beliefs and desires, but it is a priori that if a system has beliefs and desires, then they relate in various particular ways to each other, and to inputs and outputs. (In the hackneyed example, if someone desires that \( p \) and believes that her \( \phi \)-ing will bring it about that \( p \), then other things being equal, she will tend to \( \phi \).) In similar fashion, though it is a posteriori that any particular region of the universe is coloured, it is a priori that colours relate in various ways to each other, and to things that they cause and are caused by. (For example: ‘Nothing can be red and green all over.’ ‘If there is something blue at place \( p \), and a person is present with \( p \) in front of her, conscious and with a normally functioning perceptual system, then (subject to various provisos) she will have an impression as of a blue object in front of her.’42)

---

42 For parallels between the explanation of action and the explanation of perception (though he is not talking of colour in particular) see C.A.B. Peacocke, Holistic Explanation (Oxford: Oxford University Press, 1979), especially Ch. 1.
Most interesting, is the way the parallel lends support to both the subjectivity and (in a different sense) the objectivity of colour. We are used to the idea that the states we ascribe to people when we employ the psychological scheme are ‘subjective’ in the sense not just that they are states of subjects, but also that the states are themselves perspective-dependent: only a theorist who has a particular point of view (embodied in his grasp of the a priori theory) will so much as comprehend the states thus ascribed. (Martians might make nothing of our talk of jealousy, through lacking the constitution necessary to grasp the theory, or to feel jealous.)

And yet the ascription of such states is perfectly objective in the sense that those ascriptions can perfectly well be really (and not merely ‘apparently’ or ‘for practical purposes’) true. (It may a perfectly objective fact that person $a$ is jealous.) In a similar way, the employment of colour terms is subjective in the sense that it embodies a particular point of view, inaccessible to certain perfect rational people. At the same time, the colour ascriptions of people who have that ‘point of view’ are objective, in the sense of being assessable as genuinely true or not.

Of course there is more than merely a parallel between the psychological scheme and the colour scheme. There are direct connections between them, because of the fact that colours give rise to experiences of colour, and (to put the matter carefully) variations in physiology determine simultaneously variations in the colours of which we are capable of being aware, and variations in the kinds of perceptual experiences in which we are aware of them.

How exactly does this picture of colour explanation relate to the issue of autonomy? In the second usage which I have distinguished, the term ‘autonomy’ is used for the status of a discipline or explanatory scheme which is self-governing, in the sense that it sets its own standards of correctness and is open to critical assessment.

---

43 A further form of subjectivity lies in the fact that, in the case of ordinary psychological descriptions, the scheme employed by the theorist is also a scheme employed by the object of the theory: persons are in a special sense self-interpreting systems. This form of subjectivity clearly does not have a direct parallel in the case of colour.

only from within the discipline or explanatory scheme. In this sense, everyday classification of metals recognizes its own non-autonomy: we recognize that, on whether something is gold, common-sense judgements make claims that can only be fully justified (and may in fact by overturned) by going outside everyday methods of judging, and turning to the metallurgists. Whether an explanatory scheme is autonomous or not is highly sensitive to the delimitation of the explanatory scheme: taken together, everyday classification of metals and metallurgy might count as autonomous, though the first alone did not. The notion is not easy to apply: psychological explanation might be presented as autonomous, in the sense that it sets its own standards; but yet it may turn out that those standards themselves make essential reference to physical and externalist considerations. So autonomy in this sense does not mean ‘independence of the physical’. A claim of autonomy in itself carries no anti-scientific bias either: the physical sciences themselves could be described as autonomous. But what autonomy does mean is that if the internal standards of a field leave no place for criticism on scientific grounds of judgements in that field, then there is no place simpliciter for criticism of such judgements on those grounds. The claim of autonomy in itself is usually a trivial one, a demand to recognize an explanatory scheme for what it is: to recognize the relevance of certain kinds of consideration to judgements made in it, and the intrinsic irrelevance of other kinds of consideration. Whether the claim is news to anyone will depend on what kinds of consideration exactly are mentioned when the claim is developed in detail.

The role of colours as properties used in interpretation suggests that they occur in explanations that are autonomous in this second sense. How exactly? The most important point is that nothing outside the methods of colour discourse can force us to abandon a colour judgement: if on looking at an object in a variety of circumstances of whose character we are aware and checking with other people, we conclude that it is red, then there is no place provided for correction of this judgement on other grounds. So far, this is a pretty high-grade form of self-government. What we have also provided for, however, is the possibility that on occasion it may be impossible to tell by looking what colour something is (for example, because it would kill you first), and then it will often only be information about the
physical (for example, the object’s spectral reflectance curve) that can conclusively underwrite a colour attribution. Is this a departure from ‘autonomy’? It is a little hard to know how to answer. Not if the internal rules on colour ascription themselves provide for this. It is only possible to apply colours to objects that cannot be seen if you already have a view on what colours are. I would suggest that it is an integral part of colour ascription now, that anything which has the property that underlies our ordinary perceptions of red will itself be red. It is of the nature of a discovery, though a fairly obvious one, that the property in question, in the case of surfaces, is a way of changing the light (as I shall argue in the next subsection); but which spectral reflectance curves correspond to this way of changing the light, will be a piece of recherché information. On this view, it is integral to colour explanation that there will be cases where the colour of something can only be determined by means other than simple perception: so colour explanation is autonomous precisely by itself making provision for reference to the areas outside superficial perception. By now this is a fairly modest claim of autonomy; but less turns on the employment of the label than on the understanding of the reasons for which it is employed.

Two further issues are worth a few quick comments. To describe colour explanation and psychological explanation as forms of ‘interpretation’ should not be taken as suggesting (as it does sometimes in Davidson) an instrumentalist or anti-realist view of the items ascribed in such explanation. Unless the occurrence of a term in a theory used to explain or interpret phenomena automatically forces an anti-realist understanding of that term, then the use of the word ‘interpretation’ here will not do so. Secondly, I have sometimes used the term ‘rational explanation’ where ‘psychological explanation’ would in some respects have been less misleading. I used the former term to allude to a view of the nature of psychological explanation found in Davidson and McDowell, which I wished to use as a model for colour explanation. But the term may have shortcomings even in its original context. The domain of psychological explanation seems to be wider than that of rational explanation, if the latter is defined by the ‘constitutive ideal of rationality’. We may on occasion explain what people are doing by saying they are playing football, doodling or dancing. The explanations are perspective-dependent (and may be
impenetrable to the Martians), but it is not clear that the concept of rationality is central either to the activities or to our recognition of them. One might well say that we can describe people as engaging in such activities only if we treat them also as subjects of belief and desire, that is, as subjects of explanation that really is governed by the constitutive ideal of rationality. But even if this is true, it concedes that rationality is only one dimension of assessment of the distinctive lives of sentient beings. Even if the ‘constitutive ideal of rationality’ is the clearest source of the combination of subjectivity, autonomy and realism that we find in psychological explanation, it is not to be assumed that it is the only one.

If this section has described some of the nature of colour discourse, it leaves us with the unfinished business of describing the nature of the items that this discourse actually equips us to talk about—that is, the nature of colours themselves. A recent suggestion is that of Jonathan Westphal: the colours of surfaces are ways in which they change the light. The suggestion is immensely valuable, but it needs care to develop it in a satisfactory way.

3.3 The colour of a surface as a way in which it changes the light.

There is a difference between the light reflected from a surface and the light incident on it. At or around any one wavelength the surface will reflect a certain proportion of the light, transmit a certain proportion, and absorb the remainder. And it is a high-level empirical fact about objects in our environment that, for any one region of a surface, the proportions of light reflected, transmitted and absorbed at each wavelength, are normally constant from one time to another, regardless of how large or small is the absolute quantity of light falling on the object. It is therefore normally possible to describe how a surface changes the light by giving a ‘spectral reflectance

---

45 A physical colour is defined by Westphal as ‘the alteration of a complete spectrum’. A green object, for example, is ‘an object which will absorb or darken almost all of any incident red light and reflect or not darken higher proportions of light of the other colours.’ (Colour, 84)

46 Exceptions include objects that change colour with temperature, those that fade in the light, and things like light-sensitive sunglasses. Fluorescent, iridescent and glossy surfaces also need special treatment.
The Autonomy of Colour

curve’ for it, showing what proportion it reflects of light at each wavelength. (See Figure 2) The spectral reflectance curve for a surface determines its colour, on at least one natural understanding of that term, according to which the colour of, for example, a British post-box is red—even if it is being mistaken for brown because the light is bad, or for green by a person who is colour-blind or a recent victim of spectrum-inversion, or it is invisible because there is no light on it at all.

Fig. 2: Spectral reflectance curves of some typical natural objects


Though spectral reflectance curves determine colours, they are not identical with them. There are two problems. First, the space of spectral reflectance curves differs from colour space: spectral reflectances vary in an infinite number of dimensions (corresponding to the proportion of light reflected at each of an infinite number of wavelengths of visible light), whereas on a natural interpretation, colours vary in just three dimensions of hue, saturation and brightness or lightness. The loss of dimensions corresponds to the fact that there are only three kinds of cone in the retina, and objects with physically different spectral reflectances may be indistinguishable in colour. Secondly, the phrase ‘the colour of the object’ is indeterminate: the object may be at one and the same time red,
The Autonomy of Colour

vermilion, a highly saturated vermilion, and also R10Y 3080 (to pick a figure out of the air for the object’s coordinates in the Natural Colour System). The colour of an object is like its position, which may be at one and the same time: in the house, in the bedroom, in the top drawer of the bedside table, and at such-and-such a coordinate position. Plainly at best there is going to be a many-to-one relation between spectral reflectance and even a colour coordinate position like R10Y 3080; and it will be even more many-to-one between spectral reflectance and a colour classification like red.

This does not mean that the colours of surfaces are not ways in which they change the light; it means only that ways of changing the light can be individuated in ways other than those of physics. Westphal’s suggestion in his book Colour has two distinctive features; my own will emerge in discussion of them. First, Westphal seems to think we can characterize the colour of an object by taking its spectral reflectance curve and ‘phenomenalizing’ it—that is, reading it so that ‘the x-axis is illuminant colour, not wavelength’ (p. 32). (A surface colour itself might then be characterized by a certain type of thus phenomenalized spectral reflectance curve.) Secondly, following a suggestion of Wilson and Brocklebank, Westphal says that the significant property of objects ‘is not the colour ... of the light they reflect, but rather the colour of the light they don’t reflect’ (p. 80). So a green object will be one which ‘refuses to reflect a significant proportion of red light’ (p. 81); a yellow object will be one that refuses to reflect much blue light (p. 80). When defining the colour of lights, Westphal mentions how lights of different colours will ‘darken’ objects—that is, I think, make them appear dark (compared with other objects illuminated with the same light). Thus a red light is one that is ‘disposed to darken green objects’ (p. 85), and a blue one one that will darken yellow objects. Putting the two types of definition together, we have a medium-sized if not small circle: objects are green if they refuse to reflect red light, and light is red if it darkens green objects.

Neither of the two distinctive features of Westphal’s suggestion will quite do. First, if we simply read the x-axis of a spectral reflectance graph as illuminant colour, then it is most unlikely that for any one x-value a single y-value can be recorded as the proportion of light of that colour which the object is disposed to reflect. For, given
the facts of metamerism, it is quite possible for an object to reflect, say, 85 per cent of incident light of 570 nm, but only 75 per cent of a mixture of light of 550 and 600 nm which looks (and is) exactly the same colour. Secondly, specifying kinds of light which an object fails to reflect is not enough to determine its colour, unless we add that those are the only kinds of light which it fails to reflect—and that of course would then be equivalent to telling us that it does reflect significant amounts of other kinds. So I cannot see any advantage in talking of the kinds of light that objects fail to reflect rather than the kinds that they succeed in reflecting.

But this should not discourage us from working for a conception of colours as ways in which objects change the light, which is phenomenal but not quite in the way that Westphal suggests. Abandoning the two-step procedure that defines surface-colour in terms of the (complementary) colour of light which such surfaces refuse to reflect, we might talk instead of the light that the surfaces succeed in reflecting. Holding to the aim of characterizing colour phenomenally, we might say: the red surfaces are those that when illuminated with normal white light tend to reflect light that is, phenomenally, red (whatever the spectral composition of that light)—where for accuracy that must be interpreted: the surfaces tend in normal white light to reflect light that results in normal people when

---

47 This is significant for example for Westphal’s project of explaining why nothing can be red and green all over. The incompatibility is only a straight logical incompatibility as Westphal wishes, if it is the incompatibility of an object’s reflecting a low proportion of green light but not of red light, and also reflecting a low proportion of red light but not of green light. Without the italicized clauses, it would be open to someone to think an object could be both red and green, by being black.

48 Westphal’s motive, I think, is that the results of subtractive colour mixing are more directly predicted from the former kind of information than from the latter (though as we have seen in section 2, they will not be exactly predictable from colour information at all without recourse to spectral data). This of course does not show how ordinary people (as opposed perhaps to dyers) think of colours; and if Westphal is not making a claim at the level of thought (or Fregean sense), then there seems no reason for preferring his kind of characterization to an equivalent one in terms of colours reflected.

As a matter of fact, grass actually reflects what one might well think a significant proportion of red light: approaching 80% at the extreme red end of the spectrum, as Figure 2 illustrates. It none the less looks green partly because of the relatively low sensitivity of the eye at that point, and also because the effect of increasing by just a small amount the reflectance of the complementary colour red, will only be to lighten or desaturate the green.

37
The Autonomy of Colour

normally affected by it having a perception as of a red material object. But by then it seems no less illuminating to say, quite non-reductively, that red things are ones that change the light in a certain way, and that way is the way in which red things change the light. Does this make a pointless detour? No: for the route brings to light the fact that surface colours are ways of changing the light. What red surfaces have in common is what they do to the light: a factor that remains constant however the character of the incident and reflected light vary individually. It is like the elasticity of a spring, a constant factor characterizing the way the length varies with the force to which the spring is subjected.

Are there other ways to characterize this way of changing the light? We can say a lot within the everyday language of colour, employing either a priori or a posteriori connections: red things look darker than yellow ones; red is a deep colour; the most saturated reds seem more saturated than the most saturated yellows; red shades into orange in one direction and into purple in the other; a lightened red becomes a pink; red is a ‘unique’ hue: there are reds that seem to contain no hint of any other colour; nothing can be red and green all over. We can produce samples: these things are red, we may explain, and those are not, and there are borderline cases like these—though the samples will only be of use to people who can perceive pretty much as non-colour-blind humans do.

Can red be characterized in non-colour terms? We can clearly go a long way. We can characterize it as the colour corresponding to spectral light of approximately wavelengths 650 to 750 nanometres. But that of course uses the predicate ‘colour’. We can do better: psychophysics and colorimetry have, I think, put us in a position where we can tell of any newly presented object what colour it is,

---

49 I talk of ‘normal’ white light, to count out, for example, light that is composed of two very narrow complementary bands around, say, 480 and 580 nanometres. Such light is phenomenally detectable for the reason that coloured objects illuminated by it look quite different from normal. A defence of the employment of the notion of normality must wait for another occasion.

50 I should perhaps make clear that the claim is that (all) surface colours are ways of changing the light, not that (all) ways of changing the light are surface colours. There will be some objects that change the light in ways that preclude our ascribing any simple colour to them—e.g. mirrors, highly metameric objects, oil films—except perhaps relative to a particular angle or position of view, or of illumination.
simply from its spectral reflectance curve (together with standard data culled from human subjects).\textsuperscript{51} It should therefore be possible to find a ‘physical’ property coextensive with ‘red’. The questions remain whether this physical property will be necessarily coextensive with the colour, and whether it will be identical with it—and I cannot pretend to answer these here. But it must be obvious that the view of colours as introduced by a distinctively subjective scheme of explanation, and as phenomenally-characterized ways of changing the light (as the present section has suggested), does nothing to rule out the idea that they might be characterizable \textit{a posteriori} in other ways as well.

The attractions of taking colours as ways of changing the light are tremendous. I shall mention several.

First, consider a red car on a bright day. It clearly looks red. But you will also be able to see in it the reflections of other things around, from the road and the other cars to the sky above. The surface in one sense looks a perfectly uniform colour, but almost every point on it is, in another sense, presenting a different appearance. What is it that it constantly looks to be, when it in this sense constantly looks red? Why does it ‘look the same’? Because there are in the visual array cues\textsuperscript{52} that enable us to recognize it as a presentation of an object across whose surface there is a constant \textit{relation} between incident and reflected light: there is a constant \textit{way in which the surface changes the incident light}.

Secondly, this conception explains how it is that in order to tell what colour an object is, we may try it in a number of different lighting environments. It is not that we are trying to get it into one single ‘standard’ lighting condition, at which point it will, so to speak, shine in its true colours. Rather, we are looking, in the way it handles a variety of different illuminations (all of which are more or less ‘normal’), for its constant capacity to modify the light.

Thirdly, this makes sense also of what might otherwise be thought a strange phenomenon: aspect-shift in colour perception. I

---

\textsuperscript{51} I describe the method to be employed in \textit{The Nature of Colour}.

\textsuperscript{52} Here and elsewhere I employ some of the language and the doctrine of Gibson’s ecological optics. See e.g. J.J. Gibson, \textit{The Ecological Approach to Visual Perception} (1979; reissued, Hillsdale, NJ: Lawrence Erlbaum, 1986), Ch. 5 and pp. 97-99.
have had the experience of looking at a book cover from just one angle, uncertain whether it is dark blue or black. At that point I can see it alternately as dark blue, and then as black, shifting at will between appearances. The effect goes as soon as I turn the book at a new angle to the light: it is suddenly clear that the object is dark blue. To count a transient perception as a perception of a dark blue object is to be prepared to count some different perceptions as perceptions of the same enduring object with the same colour; if such different perceptions do not in fact materialize as expected, then the object after all is not dark blue. To be dark blue is not crudely to have a disposition to present a single appearance in a single kind of lighting; it is to present a variety of appearances in a variety of kinds of lighting, according to a constant pattern. And if it is puzzling how a dynamic property can be make itself manifest in a static perception (‘how can a disposition to present a variety of appearances be visible in a single appearance?’), then we already have, in familiar discussions of aspect-shift, the theoretical apparatus for a solution. It is because there is ‘the echo of a thought in sight’: our perception of an object as having a certain colour is ‘soaked with or animated by, or infused with’—the metaphors are Strawson’s—‘the thought of other past or possible perceptions of the same object’. If it is a shock to find even colour appearances treated as soaked with thought—rather than being the brutely given qualia of today’s descendants of sense-data—then that is a shock worth undergoing.

A remarkable related phenomenon occurs with various forms of partial colour-blindness. The impression is often given that people classed as ‘red–green colour-blind’ (namely, the roughly 8 per cent of

53 cp. Strawson: ‘there would be no question of counting any transient perception as a perception of an enduring and distinct object unless we were prepared or ready to count some different perceptions as perceptions of one and the same enduring and distinct object.... To see [a newly-presented object] as a dog, silent and stationary, is to see it as a possible mover and barker, even though you give yourself no actual images of it as moving and barking...’. (‘Imagination and Perception’, in P.F. Strawson, Freedom and Resentment and other essays (London: Methuen, 1974), pp. 52-53)

54 The phrase is Wittgenstein’s (Philosophical Investigations, II.xi, p. 212). Strawson makes much of it in ‘Imagination and Perception’.
56 There are some interesting consequences of applying this outlook to the inverted spectrum puzzles, but I cannot go into these here.
The Autonomy of Colour

men and 0.5 per cent of women who have impaired colour discrimination of reds and greens, usually in the form of a condition known as deuteranomaly) have one and the same type of perception when looking at red, green and grey objects. Nothing could be farther from the truth. I myself have red–green deficiencies of vision, according to the Ishihara colour tests. I confuse certain reds and greens in certain circumstances. But I do not have a single kind of perception from red, green and grey things in general. I have no difficulty in seeing the red of a post box, or the green of the grass, and my identification of their colour is not due to knowing already what kind of thing I am looking at. (I am equally good on large blobs of paint.) So I plainly do not have just one concept, applicable equally to red, green and grey things. My problem is that occasionally I take something to be red (or brown) which turns out later to be green. What is interesting is that, when told of my mistake (or recognizing it myself, for example, after trying the object in slightly different lighting), I can usually come to see the object as having its true colour. This involves what I earlier called an aspect-shift: the object actually comes to look different, even when the physical sensory stimulation is the same. Even in the lighting situation where I originally took the object to be brown, I do not ‘get one and the same impression’ as before if I later take the object to be dark green.

This makes perfectly good sense if colours are ways of changing the light. The person with red–green deficiencies is simply less good at telling from one viewing what is the object’s way of changing the light; but by getting a variety of views of it, he may none the less recognize that property. There is no reason to say he lacks proper colour concepts; he is simply less good in applying them. His

---

57 ‘To about 5 per cent of all men, green and red are both indistinguishable from grey’ (Hazel Rossotti, *Colour* (Harmondsworth: Penguin, 1983), 123). Such people have ‘an inability to discriminate red, green and grey’ (Hacker, *Appearance and Reality*, 151).

58 These are the cards devised by Shinobu Ishihara, Professor of Ophthalmology at the University of Tokyo, in 1917. Each card carries a circle made up, like a pointilliste painting, of small blobs of different colour. Normal trichromats will see one numeral in the pattern of blobs; the colour-blind, according to their pattern of non-discrimination, will see a different numeral, or sometimes none at all. More sophisticated tests are available now, though I have not seen them myself, and the Ishihara test is still rated as efficient in detecting red–green defects.

My own deficiency is, I think, deuteranomaly.
The Autonomy of Colour

experience is, so to speak, ambiguous where other people’s is not—in rather the same way as the experience of a person who sees with only one eye will be ambiguous, where that of a person who sees with two eyes is not. Most of the time those with monocular vision have no difficulty in recognizing the three-dimensional shape of objects around them—resolving cases of ambiguity by seeing the object from different angles. We certainly have no temptation to say that they lack a concept of shape in three dimensions.\textsuperscript{59} And something parallel can be said of people with anomalies of colour vision.

This conception of colour also makes good sense of how colour perception could have evolutionary significance: the ways in which a surface changes the light will be a constant factor the tracking of which can easily be of benefit to an organism that is tracking that surface.\textsuperscript{60}

The attractions of regarding colour as a way of changing the light are clear, though (particularly in the case of evolution, ecology and colour-blindness) they raise questions that demand attention at much greater length. None the less, this cannot be a full conception of colour, simply because colours apply not only to objects that change the light, but also to objects that emit light. ‘Yellow’ is not simply

\textsuperscript{59} There is another source of difficulty for the view that those with red–green anomalies lack normal colour concepts. They are said to be capable of recognizing variation in saturation and in brightness (obviously enough in, for example, the case of yellow and blue objects); they also (if they are not deuteranopes) see a full variation of spectral hues. But in that case, putting together the concepts of saturation, brightness and hue, they ought to have a full conceptual grasp of the variation of colour within the red and green regions too—whatever their difficulties may be in recognizing it.

There are more serious deficiencies of colour perception which result in certain spectral lights being seen as achromatic—which one would naturally suppose was bound to result in misapprehension of the hue circle. But there is still, even with such people, a real question (and I mean just that), whether they may none the less have learned our own colour concepts properly, by appreciating the abstract structure of colour space (which other people in the community can tell them about) and recognizing as much of their colour perception as is sound as the (more than usually fallible) presentation of certain colours in restricted regions of that space—which can serve as points of reference allowing them to ‘place’ other colours which they actually cannot see.

\textsuperscript{60} By contrast it is hard on the orthodox dispositional view to see any reason why colour vision should have had adaptive value. Why on earth should our ancestors have evolved so as to be good at tracking the disposition of objects to cause a sensation of red in normal humans under normal circumstances? Human beings didn’t even exist then!
ambiguous, as applied to surfaces, lights, illumination, films, after-images and so on. But in at least some of these cases, the colour plainly cannot be a ‘way of changing the light’, so if yellow is a single feature, it cannot strictly ever be literally a way of changing the light.

The beginnings of a solution of the problem are not hard to find. There are a number of links between the yellowness of light and the yellowness of a surface. Yellow light falling on a white object will make it look yellow or yellowish, depending on the degree of adaptation possible to the light. Yellow light shining through a white translucent glass (like the globe of an old station waiting-room) will look similar to white light shining through a yellow globe. The exact connection of the two categories (surface-colour and light-colour) however is complex, as Wittgenstein has taught us. Why is there no brown light? Why are there no grey lights?

What exactly are the parallels between the two areas of colour application? My own answer, which I think is also Wittgenstein’s, involves both language and innate endowment, including physiology. To put the answer schematically: it is part of our language that some terms and not others apply to both lights and surfaces; someone who failed to grasp this would count as using different concepts from us. Other species might act differently from us in this, but then they would be using colour concepts that were at best analogues of our own. But what we are here characterizing as similarities embedded in the language cannot be separated from our innate predispositions: it may, for example, be necessary to have a certain kind of neurophysiology in order to be capable of learning a language of this type. As always with rule-following, the ability to ‘go on’ in a certain way, given a certain training, is something that depends upon a certain natural endowment (and incidentally also upon environmental conditions, like a certain constancy in the objects around us). Human beings to a large extent share with other humans these natural endowments, but there are conspicuous cases where, for example in the case of colour-blindness, a person is unable to learn a concept that others try to teach him.

61 And remember: ‘If we taught a child the colour concepts by pointing to coloured flames, or coloured transparent bodies, the peculiarity of white, grey and black would show up more clearly.’ (Wittgenstein, Remarks on Colour, III.240)
The Autonomy of Colour

The suggestion is that colours of surfaces are ways in which they change the light. The colours of lights are intimately connected with them: at one level, we may say, the same colours apply in parallel cases ‘because such cases look similar’; but if pushed to say in what respect the cases look similar, we could only say ‘in colour’, and that was the very similarity we were (perhaps misguided) trying to explain. If on the other hand we say only ‘the cases strike us similarly’ then that is clearly correct: the very fact that we react by applying the same term shows that. But it is not a deep explanation of why the colour applies as widely as it does. Changing levels, we may indeed look at neurophysiological characteristics that are similar between human perception of the two cases. But in the most general terms, we may say that the similarity resides as much in our reactions to the things as in the things.

* * *

Section 2 defended autonomy in Putnam’s sense: independence from mode of realization. The present section has defended autonomy in a second sense. A third sense might come to mind: explanatory independence from the objects of other explanatory schemes. I should make clear that I am not claiming autonomy in this third sense, either for colours or for psychological states. Psychological states do not form an explanatorily closed system, and the defence of psychological explanation as a form of holistic explanation autonomous in the other two senses does not need to pretend otherwise. There are a mass of psychological phenomena that are not explicable by the methods of rational explanation. ‘Why do human beings forget as much as they do?’ ‘Why is a person who has just lost his job more susceptible to illness than one who hasn’t?’ ‘Why do people become schizophrenic?’ ‘Why do people with Alzheimer’s disease lose their mental faculties?’ Maybe part of the answer to these questions will involve rational explanation invoking psychological states like imagining and desiring. But in most cases, what explanation is available will cross into other areas, like (at different levels, more than one of which may be relevant to a single puzzle) physiology, evolutionary biology, scientific psychology, and others.

In similar fashion, colour phenomena, though sometimes explained by other colour phenomena, will not always be. A mass of questions come to mind the answers to which force us to turn to other
The Autonomy of Colour

disciplines. ‘Why do colours fade?’ ‘Why does red seem to advance in front of green?’ ‘Why is the sky blue?’ ‘Why is grass green?’ ‘Why do colour contrast effects occur?’ ‘Why do blue objects come to seem relatively brighter vis-à-vis red objects as the light goes in the evening?’ (The Purkinje phenomenon.) ‘Why does hue change with brightness?’ (The Bezold-Brücke effect.) Only physics, physiology, chemistry, and various forms of scientific psychology can tell us. But the admission that colour phenomena form anything but an explanatorily closed set is perfectly compatible with claims of autonomy in the other two senses.