

## NAMING THE COLOURS<sup>1</sup>

David Lewis

### I. Manifesto

An adequate theory of colour must be both materialistic<sup>2</sup> and commensensical. The former demand is non-negotiable. The latter can be compromised to some degree. We need not be 'ever so inclusive' in advancing all our offhand folk-theoretical opinions as conditions of adequacy on a theory.<sup>3</sup> Imperfect occupants of the folk-theoretical role of colour will be imperfect deservers of the name, but may nevertheless deserve it quite well enough.

But compromise has its limits. It won't do to say that colours do not exist; or that we are unable to detect them; or that they never are properties of material things; or that they go away when things are unilluminated or unobserved; or that they change with every change in the illumination, or with every change in an observer's visual capacities; or that the same surface of the same thing has different colours for different observers. Compromise on these points, and it becomes doubtful whether the so-called 'colours' posited in your theory are rightly so-called. Yet it is a Moorean fact that there are colours rightly so-called. Deny it, and the most credible explanation of your denial is that you are in the grip of some philosophical (or scientific) error.

### II. Folk Psychophysics

In other words, it is a Moorean fact that the folk psychophysics of colour is close to true.

Like other folk theories, the folk psychophysics of colour is a generally shared body of tacit belief. It concerns not only colours themselves, but also the inner states of colour experience that the colours tend to cause in us, and further inner states and behaviour that these colour experiences cause in turn. It is not to be supposed that we go around with careful formulations of folk psychophysics in mind, though perhaps with enough patience it might be possible to elicit such formulations Meno-fashion.

Folk psychophysics should, for instance, afford an explanation of why we can sort

<sup>1</sup> Thanks are due to Ian Gold, Frank Jackson, J.J.C. Smart, Michael Smith, and others; and to the Boyce Gibson Memorial Library.

<sup>2</sup> I say 'materialistic' where some would rather say 'physicalistic': an adequate theory must be consistent with the truth and completeness of some theory in much the style of present-day physics. ('Completeness' is to be explained in terms of supervenience.) Some fear that 'materialism' conveys a commitment that this ultimate physics must be a physics of matter alone: no fields, no radiation, no causally active spacetime. Not so! Let us proclaim our solidarity with forebears who, like us, wanted their philosophy to agree with ultimate physics. Let us not chide and disown them for their less advanced ideas about what ultimate physics might say.

<sup>3</sup> See Mark Johnston, 'How to Speak of the Colors', *Philosophical Studies* 68 (1992) pp. 221-263, esp. pp. 221-222.

dyed bits of wool, and then shuffle them together, and then sort them again into just the same heaps as before.<sup>4</sup> This folk-psychophysical explanation will involve many causal chains, each running from the colour of a bit of wool to a colour experience in the sorter, and thence to a desire to put that bit of wool on a certain heap, and thence to the desired behaviour.

The folk psychophysics of colour is doubtless a fragment of some larger folk theory. Elsewhere in that larger theory we might find the explanation of why our wool-sorter wanted to succeed at his task. We need not worry about demarcation, so long as we agree to include all that will be needed for what follows.

The folk psychophysics of colour is common knowledge among us. In the same tacit way in which we believe the theory itself, we likewise believe that others around us all believe it too; and that they in turn ascribe belief in it to those around them in the same way we do; and so on *ad infinitum*, or at least as far as is humanly possible.<sup>5</sup>

A theory implicitly defines its theoretical terms. If, without benefit of any prior definition of 'entropy', thermodynamics says that entropy does this, that, and the other, we may factor that into two parts. There is an existential claim – a 'Ramsey sentence' – to the effect that there exists some quantity which does this, that, and the other (or near enough). And there is a semantic stipulation: let that which does this, that, and the other (or near enough), if such there be, bear the name 'entropy'.<sup>6</sup> Here is another way to say it: the theory associates with the term 'entropy' a certain theoretical role. It claims that this role is occupied. And it implicitly defines 'entropy' as a name for the occupant of the role.

What thermodynamics can do, the folk psychophysics of colour can likewise do. The situation is messier, of course, because the theory is tacit. Lacking an authoritative codification, we have no sharp way to tell what is non-negotiably part of the theory, what is provisionally part of it, and what is no part of it at all. Nor are these matters of secret fact. When folk psychophysics defines its theoretical terms, we run a risk of semantic indecision. Yet within limits, that indecision is harmless. We can hope – indeed we can be confident, as a matter of Moorean fact! – that the safe limits are not exceeded.

Central to the folk psychophysics of colour are a multitude of principles connecting colours and corresponding colour experiences. They go roughly as follows.

When a red thing is before someone's eyes, it typically causes in him an experience of redness

or better

When a red thing illuminated by normal light is a suitable distance before the eyes of someone with normal visual capacities in normal surroundings, it typically causes in him an experience of redness.

<sup>4</sup> See J.J.C. Smart, 'Colours', *Philosophy* 36 (1961) pp. 128-142.

<sup>5</sup> See my *Convention: A Philosophical Study* (Cambridge, MA: Harvard University Press, 1969) pp. 52-60; Stephen Schiffer, *Meaning* (Oxford: Oxford University Press, 1972) pp. 30-42; Margaret Gilbert, *On Social Facts* (London: Routledge, 1989) pp. 186-197 and elsewhere.

<sup>6</sup> See my 'How to Define Theoretical Terms', *The Journal of Philosophy* 67 (1970) pp. 427-446.

'Normal light' can be explained in terms of the range of illumination that most people – actually, nowadays, and hereabouts – mostly encounter. Likewise, *mutatis mutandis*, for normal capacities and normal surroundings. Some examples of the abnormal are: sodium vapour light; eyes adapted to sodium vapour light; a room with purple walls, floor, and ceiling, filled with purple furniture, with purple curtains over the windows. I do not suppose it is incumbent on folk psychophysics to do much by way of listing abnormal cases – if it tried, it would soon outrun common knowledge. A statistical conception, rigidified to actuality and nowadays and hereabouts, should suffice. It is otherwise with distance: I dare say that folk psychophysics has a lot to tell us about which distances are 'suitable', but for me that part of the theory remains tacit.

Further, it should really have been

When a red thing is at such-and-such direction before someone's eyes, it typically causes in him an experience of redness-in-such-and-such-direction

or

When a red thing fills such-and-such solid angle . . .

None of these refinements will prove relevant to the principal topic of this paper, so for the most part I shall ignore them.

Drawing on these connecting principles, and also on the part of folk psychophysics that classifies colours as properties (for the most part) of the surfaces of opaque things, and colour experiences as inner states of people (and perhaps other animals), we have folk-psychophysical roles for the colours and for the colour experiences. When we take the theoretical terms to name the occupants of the theoretical roles, we arrive at 'definitions' such as these.

*D1* *Red* is the surface property of things which typically causes experience of red in people who have such things before their eyes.

*D2* *Experience of red* is the inner state of people which is the typical effect of having red things before the eyes.

If, as a matter of contingent fact, the surface property that causes experience of red is a certain reflectance property – that is, a property that supervenes on the thing's reflectance spectrum – then *D1* may serve as a premise for a chromophysical identification: red is, as a matter of contingent fact, that reflectance property. And if, as a matter of contingent fact, the inner state which is the typical effect of red things before the eyes is a certain pattern of neuron firings in the visual cortex, then *D2* may serve as a premise for a psychophysical identification: experience of red is, as a matter of contingent fact, that pattern of neural firings.

*D1* and *D2* are all very well as truths; as widely shared tacit beliefs, and even as items of our common knowledge; and as premises for chromophysical and psychophysical identifications. But as a pair of definitions they are almost totally useless, by reason of

circularity. Anyone who needed definitions both of 'red' and of 'experience of red' would be little wiser if we gave him both *D1* and *D2*. That circularity, its remedy, and the new problem that arises out of that remedy, shall be our principal business in the rest of this paper.

### III. Objections and Replies

But first I stop to address some other likely worries, apart from circularity, about what I have already said.

*Objection.* Our colour experiences do not depend solely on the colours of things before the eyes. They depend also on the illumination, the visual capacities of the observer, and the surroundings. And it is not good enough just to make an exception for 'abnormal' cases, understood simply as infrequent cases (infrequent actually, nowadays, and hereabouts). There is no range of 'normal' cases such that we only seldom find ourselves outside that range, and also such that within that range the effects of differing illumination, etc., are entirely absent.<sup>7</sup>

*Reply.* No; but there is a range such that we seldom find ourselves outside it, and such that within it the effects of differing illumination, etc., are so small that they are hard to notice. Go from noonday sun to a cloudy afternoon to sunset; go from any of these to incandescent lighting to fluorescent lighting (but not to sodium or mercury vapour lighting); go from the room painted yellow to the room painted purple (but not to a room where almost everything in sight is purple); go from one to another so suddenly that your eyes are maladapted at first. Take with you a book with a green cover. It will continue throughout to cause experience of green, and will not start to cause experience of blue instead. Indeed, if it is the right book, it will continue throughout to cause experience of some fairly specific shade of green: Brunswick green, not malachite green or apple green.

(To be sure, it does not continue throughout to cause *precisely* the same sort of colour experience. And if our topic were not the naming of the colours that we name in daily life, but instead the naming – or rather, numbering – of the myriad precise shades that must be distinguished by matchers of paint, we would have to be much more fussy.)

In short, the barely noticeable differences in colour experience due to frequently encountered differences in illumination are not a problem. The much more striking differences in experience that are caused by much less frequently encountered differences in illumination are not a problem either: we need only say, for instance, that sodium vapour light causes colour illusions. It is only if we conflate these two non-problems, and we try to discuss the entire range of dependence of experience on illumination in a uniform way, that we get a problem. It would be daft to deny that sodium vapour light causes illusions. It would be still more daft to say that *all* light causes colour illusions, with the sole exception of cloudy Scottish daylight (or some other precisely and arbitrarily chosen standard).

<sup>7</sup> See Keith Campbell, 'Colours' in *Contemporary Philosophy in Australia* (eds.) R. Brown and C.D. Rollins (London: Allen and Unwin, 1969); but for a somewhat different view, see also Campbell, 'The Implications of Land's Theory of Colour Vision' in *Logic, Methodology, and Philosophy of Science* (ed.) L.J. Cohen (Amsterdam: North-Holland, 1982).

Another way to create an illusory problem is to ignore our daily experience, follow where theory leads, and then back an oversimple theory. Imagine that the eye is simply a spectrometer that measures the intensities at various wavelengths of the light coming from the thing seen, and produces colour experiences that depend just on those intensities. Then the experience produced by something with an unchanging reflectance spectrum would depend a lot on illumination – even for illumination within the frequently encountered range. The colours of things could not seem as constant as they manifestly do.

So should we assume severe and ubiquitous inconstancy when we philosophize about colour? Of course not! Rather, we should stop assuming that the eye is simply a spectrometer. We should think of it, perhaps, as a *calibrated* spectrometer.<sup>8</sup> If the spectrometer's measurements are corrected to compensate for differences in illumination, they will indeed measure reflectance properties, and experience that depends on the corrected measurements will indeed exhibit constancy of colour. And if the spectrometer shares information with similar spectrometers aimed in different directions (or if it spends some of its own time scanning around the scene), and if it has definite preconceptions about the sort of distribution of reflectance values to be expected, then it will be in a position to calculate the proper correction factors. Of course the method has its limitations. No amount of calibration will permit measurement of reflectance values in a band of wavelengths that is missing from the illumination; when the preconceptions used to calculate correction factors are mistaken, the reflectance measurements will likewise be mistaken; and recalibration after a change of illumination will not be instantaneous. Nevertheless, this fallible method may be good enough to yield approximate constancy throughout a wide range of frequently encountered cases. And that suffices.

*Objection.* A brown expanse on the television is composed of diversely coloured pixels – none of them brown. The brown expanse and a certain blue pixel are before the eye in the same direction, at the same distance, at the same time. Should they not cause a colour experience of blue and brown in the very same direction? Yet they do not.

*Reply.* The pixel and the expanse are not both at a 'suitable distance', even though they are at the same distance, because the suitability of a viewing distance depends on the size of the thing viewed.

*Follow-up objection.* Then consider a pixel-sized part of the brown expanse. Now the brown thing and the blue thing are the same in size as well as distance, so if one is at a suitable viewing distance then so is the other.

*Reply.* A pixel-sized part of a brown expanse is not necessarily brown. In this case it is blue. In fact, it is the blue pixel already considered. If the pixel is at a suitable distance, no *brown* part of the brown expanse is at a suitable distance and also in the same direction as the pixel. Colours are imperfectly 'dissective' properties: they are not always shared between a thing and its proper parts.<sup>9</sup>

<sup>8</sup> Here I follow the lead of David Hilbert, *Color and Color Perception* (Chicago: Center for the Study of Language and Information, 1987). One theory that treats the eye as a calibrated spectrometer capable of measuring reflectance properties is Land's well-known 'retinex' theory. But it is by no means the only such theory on offer.

<sup>9</sup> See Hilbert, *Color and Color Perception*, pp. 29-42.

*Objection.* Some ostensible facts about the colours – for instance, that there cannot be a reddish green, or that there cannot be a shade of yellow that is closer to various shades of blue than it is to any other shade of yellow – are best explained in terms of the way our colour vision works, rather than in terms of relations between physical properties of surfaces. Then if colours are physical properties of surfaces, how can these facts of exclusion and proximity be facts about colours?

*Reply.* Our account provides a correspondence between colours and colour experiences. (Various rival accounts also provide such correspondences.) For now, that correspondence takes the form of a set of definitional circles, but soon we shall give it a more satisfactory form. Whatever form it takes, the correspondence yields relations among colours in the image of relations among colour experiences (or vice versa). So no matter where the relations of exclusion and proximity may originate, in the end we have them twice over: as relations among colour experiences and also as relations among the corresponding colours.

(We might have had an offhand opinion that these relations originated as relations among surface properties. If so, we were wrong. But I am not sure we had any such opinion at all; and if we did, we have no business elevating it into a Moorean fact of folk psychophysics.)

If it is absolutely impossible to have an experience of reddish green (or if it is nomically impossible, or if it is unimaginable, or if it is just very difficult to arrange); and if reddish green is by definition the surface property that typically causes experience of reddish green; then the desired exclusion among colours follows. Then it is absolutely impossible (or nomically impossible, or unimaginable, or difficult) for there to be a reddish green surface.<sup>10</sup> *Quod erat demonstrandum.* Likewise for the yellow in the midst of blue.

*Objection.* Not only surfaces of opaque things are coloured, but also transparent volumes, light sources, the sky, . . . Colours are not always surface properties. *A fortiori* they are not always reflectance properties.

*Reply.* I have no opinion about whether the blue of the sky and the blue of an opaque picture of the sky are one shared non-disjunctive property, or whether they are two different, but saliently related, properties.

But even in the first case there is a colour property – *surface blue*, or rather a certain shade thereof – that the picture has and that the sky itself lacks. (It is the conjunction of the shared property with the property of being a surface of an opaque thing.) And in the second case also, we have a surface blue that the picture has and the sky lacks. It is these surface colours that are my present concern, and that well might be identical to reflectance properties. My restriction of topic leaves unfinished business. But I doubt that it leads to error within my restricted topic.

*Objection.* If colours are reflectance properties, they supervene on reflectance spectra.

<sup>10</sup> It is not absolutely or nomically impossible to arrange for an experience of reddish green. But it is difficult, at least by the method presently known: that of H. Crane and T.P. Plantida, 'On Seeing Reddish Green and Yellowish Blue', *Science* 221 (1983) pp. 1078-1080. That method does not involve any special surface property before the eyes. I do not know whether those who have the experience acquire an ability to imagine it afterward when they are no longer having it.

A reflectance spectrum is a disposition, or a bundle of dispositions, to reflect various proportions of light at various wavelengths. A disposition requires a 'causal basis'. For instance, something is disposed to break when struck if it has some property – some intrinsic property, at least in the clearest cases – which, together with striking, would cause breaking. In the case of reflectance properties, the causal basis is a matter of surface microstructure. It turns out that a given reflectance property can have any of many microstructural bases, different ones in different actual cases. Now we face a nasty four-way choice.<sup>11</sup>

Shall we say (1) that the colour is not the reflectance property after all, but rather it is the microstructural causal basis thereof? Or shall we say (2) that the colour is the reflectance property, but that a reflectance property – and any other disposition likewise – is identical to its causal basis? Either way, we end up concluding that what we thought was one single colour, indeed one single shade of one colour, is really many different microstructural properties, different ones in different cases. That is seriously uncommon-sensical. Or shall we rather say that a colour is a reflectance property, a reflectance property is a disposition, a disposition is not identical to any of its causal bases, but rather a disposition is the property of having some suitable causal basis or other? Then what shall we say about how the selective reflection of light – or more generally, the manifestation of any disposition – is caused? We cannot deny that the causal basis does the causing. That is how we defined 'causal basis', after all. Shall we say (3) that only the causal basis does the causing, and the disposition itself is impotent? In the present case, that would mean denying that surface colours cause selective reflection of light; and denying also that they cause more distant effects, such as colour experiences or colour-discriminating behaviour, via causal chains that begin with selective reflection of light. And it is safe to say that colours do not cause colour experiences some other way! But to deny that colours cause colour experiences is seriously uncommon-sensical.

Or shall we say that a disposition and its basis, for instance a colour and a microstructure, *both* cause the manifestation? What kind of multiple causation is this? It seems wrong to say that the basis causes the manifestation by first causing the disposition, or vice versa, or that the basis and the disposition jointly cause the manifestation. The remaining alternative is to say (4) that we must have a case of overdetermination: the disposition and its basis redundantly cause the manifestation. For instance, the colour and the microstructure redundantly cause the selective reflection of light, and thereby they redundantly cause the colour experience. This alternative, too, is seriously uncommon-sensical.

*Reply.* Alternatives (1), (2), and (3) should be discarded. Alternative (4) is close to right, but needs correction. The key to sorting out this mess is that we should not have been talking about properties as causes in the first place. That is loose talk. (Most of the time it is harmlessly loose talk, and I do not propose to give it up just because it got us into trouble this once.) Really, the causing is done by particular *havings* of properties – particular 'events', as we call them. (But that name is not meant to imply that all events are short-lived or involve change. If something in a frozen world is changeless throughout some short or long interval, or indeed throughout all time, its changelessness during

<sup>11</sup> See Elizabeth W. Prior, Robert Pargetter, and Frank C. Jackson, 'Three Theses About Dispositions', *American Philosophical Quarterly* 19 (1982) pp. 251-257.

the interval is still called an 'event'.) The very same event that is essentially a having of some causal basis of a certain disposition is also accidentally a having of the disposition itself.<sup>12</sup> So an effect of this event is caused by a having of the basis, and caused also by a having of the disposition. But since these havings are one and the same event, there is no redundant causation. So a colour experience may be caused by a colour, and also by the microstructural causal basis of the reflectance values that comprise that colour, without having a case of causal overdetermination.

*Objection.* If we do not want to say that what we thought was one single colour is really many different properties, different ones in different cases, then we also have a problem about *metamers*: different reflectance spectra that cause the same colour experience.<sup>13</sup> What to say about them?

*Reply.* At worst, we have a problem about *some* metamers. I note, first, that even if we think that the colours will turn out to be reflectance properties, we have no reason to think that all reflectance properties – all properties that supervene on reflectance spectra – will turn out to be colours. Metamers that differ in their reflectance spectra, but do not differ in those reflectance properties that participate in causing colour experience, are no problem. We have the same colour, exactly as in the case where there are different microstructural causal bases of the same reflectance spectrum.

Suppose, for example, that there is a certain division of the visible spectrum into long-wave, middle-wave, and short-wave intervals such that, if we take the triple of integrals of reflectance over the three intervals, the colours will turn out to supervene not only on reflectance spectra but on the triples of integrals.<sup>14</sup> When the spectra differ but the triples do not differ, we have the same colour. Only if different triples caused the same colour experience would we have a problem; only then would we face a choice between denying the identity of colours with triples of reflectance integrals and admitting, uncommonly, that some colours cannot be distinguished by sight. (Not without varying the illumination, anyway. Metamers under one illumination may look different under another, especially if we resort to illuminations not frequently encountered.)

I note, second, that it would not be so very bad to acknowledge some cases of indistinguishable colours, so long as such cases were uncommon. David Hilbert reports an encouraging consequence of one well-developed model of colour constancy (that is, of how the spectrometer gets calibrated in accordance with preconceptions):

Although there are possible differences of colour that are undetectable in normal circumstances, the actual occurrence of such differences appears to be relatively rare.<sup>15</sup>

<sup>12</sup> See my 'Events' in *Philosophical Papers, Vol. II* (Oxford: Oxford University Press, 1986) pp. 247-254.

<sup>13</sup> See David Hilbert, *Color and Color Perception*, pp. 81-100; C.L. Hardin, *Color for Philosophers: Unweaving the Rainbow* (Indianapolis: Hackett, 1988) pp. 28, 46, and elsewhere.

<sup>14</sup> Perhaps the three intervals should be 'fuzzy' intervals. These are taken as functions – bell curves, or similar – measuring the degree to which a given point is deemed to fall within the interval. To integrate a quantity over a fuzzy interval, we multiply the quantity by the function just mentioned, and we integrate that product *simpliciter*.

<sup>15</sup> Hilbert, *Color and Color Perception*, pp. 130-131.

*Objection.* When the purple people eater comes out of hiding to stalk its prey, it uses its powers of telehypnosis to disrupt the colour vision of victim and bystanders. In this way it manages to be mistaken for a benign brown beast. The deception always works. The purple people eater, though it is purple from tail to teeth, never causes experience of purple. Is this a counter-example?

*Reply.* No. The question what the eater itself typically causes is the wrong question. We were supposed instead to ask what the surface properties of the eater typically cause. Does it have a surface property such that this property – or rather, particular havings thereof – will typically cause experience of purple? It does. The same property that occurs on the purple eater occurs, or at least could occur, elsewhere: before the eyes of a normal perceiver with undisrupted colour vision. In that case it typically causes, or would cause, experience of purple. Accordingly, that property is the colour purple, and the purple people eater is rightly so called.

*Objection.* The yellow killer is a more problematic predator. It too disrupts the colour vision of anyone who sets eyes on it; and it disrupts all other brain processes as well, thereby causing instant death. But in this case, it is the killer's colour that does the damage. Its special shade of yellow, 'killer yellow', is fatal regardless of what the coloured thing may be. This colour does not typically cause colour experience. It never does, and never could so long as we retain our vulnerability to it.<sup>16</sup>

*Reply.* The case is consistent if, but only if, we can subsume killer yellow as a special case under some broader property that does typically cause experience of yellow; and that is not an artificial gerrymander, but rather is unified by appropriate resemblances between its various subcases. If so, we can repeat the strategy that worked in the case of the purple people eater: we distinguish a broader property which deserves the colour name from a narrower property which, if considered just on its own, does not.

To put the point another way: killer yellow deserves to be so called, if indeed it does, in virtue of its resemblance to other shades of yellow that do cause experience of yellow. For instance, it might resemble other shades of yellow in respect of the values of its triple of long-wave, middle-wave, and short-wave integrals of reflectance, if those turn out to be the reflectance properties on which colour experiences normally depend.

In that case, the relation that unites the killer property with non-lethal shades of yellow, and thereby justifies us in classifying the killer property itself as a shade of yellow, is colour resemblance in a doubly derivative sense. First, we have resemblance among colour experiences. That resemblance does not apply directly, since there are no experiences of killer yellow. Second, we have the image among colours of this first relation among experiences, under the causal correspondence between surface properties and the experiences they typically cause: the derivative relation that holds between colours just when the colour experiences they typically cause stand in the first relation. This second resemblance also does not apply directly, since killer yellow does not cause any colour experience at all. Finally, it may be that this second relation typically correlates with a third relation: resemblance in respect of the triple of reflectance integrals. If that turns out to be so (and whether it is so is an empirical question, even supposing that the

<sup>16</sup> The example of killer yellow is due to Saul Kripke, in lectures many years ago. I am obliged to note that what I say here may not correspond to the whole of what Kripke said in those lectures.

colours do turn out to be triples of reflectance integrals) then in a still more derivative sense we may regard this third relation also as a relation of colour resemblance. So, finally, we have a relation that can unite killer yellow with other yellows.

#### IV. From Circularity to Multiplicity

I return at last to our principal difficulty: the definitional circle between the name of a colour and the name of the corresponding colour experience. Red is the surface property apt for causing experience of red, which is the inner state that red things before the eyes are apt for causing. Likewise for green, for magenta, . . .

A manoeuvre due mainly to Carnap often works to cure such circularities, but in the present case it proves disappointing.<sup>17</sup> Here is how it works, when it works. Suppose thermodynamics has *two* theoretical terms, 'entropy' and 'temperature'; and without benefit of prior definition of either, it says that entropy does this, that, and the other; temperature does so and so; temperature and entropy are related thus. Then any ordered pair of  $X$  and  $Y$  is called a *realization* of thermodynamics iff  $X$  does this, that, and the other;  $Y$  does so and so;  $X$  and  $Y$  are related thus. (Equivalently:  $X$  and  $Y$  make thermodynamics true, if taken as referents of the respective terms 'entropy' and 'temperature'.) Now we define our terms without circularity: *entropy* is the first component of the unique realization of thermodynamics, if such there be; *temperature* is the second component of the unique realization.

What to do if there is no realization, not even a near-enough imperfect realization? Here different versions of the method disagree. Carnap's version falls silent about what referents, if any, the theoretical terms have; my version and Bedard's version say that the terms fail to refer. What to do if there are two (or more) different realizations, say the pair of  $X_1$  and  $Y_1$  and also the pair of  $X_2$  and  $Y_2$ ? Again there is disagreement. Carnap's version says that either the terms refer respectively to  $X_1$  and  $Y_1$  or else they refer respectively to  $X_2$  and  $Y_2$ , but falls silent about which. My version says again that the terms fail to refer. Bedard's version says that the terms become indeterminate in reference, so that there is no fact of the matter about whether they refer respectively to  $X_1$  and  $Y_1$  or instead to  $X_2$  and  $Y_2$ ; and as usual we resort to van Fraassen's method of supervaluations and say that what is true (or false) on both resolutions of the indeterminacy alike is true (or false) *simpliciter*.<sup>18</sup> Nowadays I am inclined to split the difference between my version and Bedard's: when the two realizations are sufficiently different we get reference failure; when they are sufficiently alike we get indeterminacy of reference; in between it is indeterminate which one we get.

Now suppose our term-introducing theory is a fragment of the folk psychophysics of colour. It has just one pair of a colour name and a colour-experience name: 'red' and 'experience of red'. It says nothing to distinguish this pair of a colour and a correspond-

<sup>17</sup> Rudolf Carnap, 'Replies and Systematic Expositions' in *The Philosophy of Rudolf Carnap*, (ed.) P.A. Schilpp (La Salle: Open Court, 1963), pp. 963-966; my 'How to Define Theoretical Terms', *op. cit.*; and Katherine Bedard, 'Partial Denotations of Theoretical Terms', *Noûs* 27 (1993) pp. 499-511.

<sup>18</sup> See Bas van Fraassen, 'Singular Terms, Truth-Value Gaps, and Free Logic', *Journal of Philosophical Logic* 63 (1966) pp. 481-495.

ing colour experience from all other corresponding pairs. So we have a very severe problem of multiple realization: *every* corresponding pair of a colour and a colour experience is a realization of our theory. But it will not do to fall silent about the reference of 'red' and 'experience of red'; or to say that these terms fail to refer; or to say that these terms are radically indeterminate in reference. Our method for curing circularities, in all its versions, has been defeated by the multiplicity of realizations.

But at least we are now in a position to do some useful redefining. A *corresponding pair* is any realization of the theory just considered. A *colour* is any first component of a corresponding pair. A *colour experience* is any second component.

If we take a bigger fragment of folk psychophysics that mentions  $n$  different colours and their corresponding experiences, we make matters worse. Now we have as many realizations as there are ordered  $n$ -tuples of corresponding pairs!

Instead of our definitional circle, we have no definitions at all. We are worse off than before.

### V. Parochial Solutions

The remedy is that the folk psychophysics of colour, as we have envisaged it so far, is too abstract and general. It has a lot to say about the causal relations between colours and their corresponding colour experiences. But it needs another chapter to tell us what distinguishes each colour from all the others (or each colour experience from all the others). Then the problem of multiplicity will be solved. Once we add the missing chapter, the folk psychophysics of colour will have a unique realization. (Or near enough. Maybe there will still be multiple realizations, but they will differ so little from one another that if we follow Bedard and say that the theoretical terms are indeterminate in reference, their indeterminacy will be far from radical. Some mild indeterminacy about the exact boundaries of the colours is credible enough. In fact, it is perfect determinacy that would be hard to believe.)

When we add the chapter that distinguishes the colours by name, how shall we do it? Most simply, we could give examples. 'Red is the colour of pillar boxes' has been a philosophers' favourite, in England at least. Others might prefer to mention the people's flag. I myself would rather say that red is the colour of the diagonal stripe on an Essendon Football Club jumper.

Likewise for more specific shades. Brunswick green, for instance, was the colour of locomotives of the Great Western Railway. (Of faded ones and freshly painted ones; of those painted before and after certain changes in the recipe for the paint; wherefore it cannot be a perfectly specific shade.)

Sometimes the examples are particular and soon forgotten. I ask you 'What's magenta?' You make a mark with a crayon and say 'This is.' I trust you to have been cooperative and truthful in answering my question. So now I can say that magenta is the colour of the mark you made when I asked you what magenta was.

A week later I can no longer say that: I have forgotten the lesson. No harm done, if I retain the mental abilities which the lesson imparted to me. Any time I like, I can put myself into a state in which I can truly say 'Magenta is the colour that typically causes the colour experience I am right now imagining'. Or I can use myself (and others can use

me) as a living instrument: magenta is the colour such that I am disposed to say 'magenta' if you point to it and ask 'What colour is that?'; and if I am disposed to be cooperative and truthful, and if the magenta thing is at a suitable distance, and if the light and my visual capacities are normal.

(In calling myself a 'living instrument', I do not at all mean to suggest that my responses are mere mindless reflexes, rather than the reasoned judgements they might seem to be. No – bizarre cases aside, surely they are the latter. But for present purposes it does not matter which they are, so long as they are reliable indicators that magenta is present.)

The general chapter of the folk psychophysics of colour told us about the relations of resemblance among colour experiences, and about the derivative relations of resemblance among the colours themselves. So once we have distinguished some colours by the methods already mentioned, we can distinguish further colours by interpolation or triangulation: orange is reddish yellow; lemon yellow is ever-so-slightly greenish, and not at all reddish, yellow and so on.

In these ways, I can add a chapter to the folk psychophysics of colour that will distinguish all the colours to which I have been properly introduced. And so can you, and so can almost anyone. (Even a blind man, except that he cannot use the methods that depend on his own abilities to imagine colour and to serve as a living instrument.) Once the chapter on distinguishing the colours has been added, we no longer have a severe problem of multiple realization. Now it is safe to say that our colour names and our colour-experience names refer to the approximately unique occupants of theoretical roles. Such moderate multiplicity as remains will result at worst in semantic indeterminacy regarding the exact boundary between one colour (or one colour experience) and another. All is well.

Or is it? The trouble is that I (or you, or almost anyone) drew upon ever so much parochial knowledge: information that is not common knowledge throughout the linguistic community. In Ballybunnion where the pillar boxes are green, or in New Haven where they are blue and not in the shape of pillars, people may or may not know about the red pillar boxes in far-off lands. And even if they do, does everyone else in the English-speaking world know that they do? Does everyone know that everyone else knows that everyone else knows that they do? Does . . . Surely not!

Likewise it is common knowledge among certain people – call them *footy people* – that Essendon wears a red diagonal stripe. But this is probably not known at all to certain other people, call them *rugby people*. So it is not common knowledge throughout the linguistic community. It takes all kinds to make the English-speaking world.

Surely very few other people know of the lesson whereby I was introduced to magenta, especially if I myself forgot it in a week. Nor do people who have never heard of me know anything about my ability to imagine magenta or to serve as an instrument for indicating its presence. Here too, the chapter on distinguishing the colours was built not upon common knowledge but upon parochial knowledge.

Maybe it is common knowledge that orange is reddish yellow. But interpolations and triangulations are useless by themselves. They only serve to distinguish more colours after we have distinguished some already. How to get started?

Our parochial solution is not good enough because language is a conventional, ratio-

nal system of semantic coordination.<sup>19</sup> I say something I take to be true under semantic interpretation  $I_1$ ; you trust me to be truthful under  $I_1$  and also to be well-informed; and in this way you come to share my information – something that both of us wanted to happen. Coordination of truthfulness under  $I_2$  with trust under  $I_2$  would have worked just as well, even if  $I_1$  and  $I_2$  assigned opposite truth conditions to the same sentences. But miscoordination – of truthfulness under  $I_1$  with trust under  $I_2$  or vice versa – would have deprived us of the benefits of communication.

Any conventional system of coordination, semantic or otherwise, is rationalised by a potentially endless system of mutual expectations. Why do you drive on the left? Because you expect me to. Why do you expect me to? Because you expect me to expect you to. Why . . . In short: we both drive on the left because it is common knowledge between us that we both will. (Likewise *mutatis mutandis* if there are more drivers sharing the roads than two.) Our coordination is rationally sustained. It is not hard-wired and it is not mere luck. Even if we drive on the left by habit, as surely we do, the habit is sustained by reason. If our common knowledge were eroded by doubt, or undermined by counter-evidence, we would swiftly change our ways.

So here is the current state of the problem: if we distinguish and name the colours by recourse to parochial information, how can we have the semantic common knowledge that is required if our talk about colour is to be part of a conventional, rational system of coordination?

## VI. A Solution that Relies Too Much on Luck

A few examples of a few colours are common knowledge throughout the linguistic community. Blood is red, well-watered leaves are green, the sky is blue, and flames are yellow. (Or rather, since we agreed to confine ourselves to surface colours, pictures of the sky are blue and pictures of flames are yellow. But let us waive the point.) These few examples will do to begin the job of distinguishing the colours, and maybe we can do the rest by interpolation and triangulation. Then the chapter of folk psychophysics that distinguishes the colours will not rely on parochial knowledge. It will be common knowledge throughout the linguistic community. So when we name the colours and colour experiences as occupants of theoretical roles, that naming will be a matter of semantic common knowledge. Our problem will be solved.

This solution is all very well for today, but consider the future. Leaves, well-watered or otherwise, will be long gone and forgotten. We will resort to brute force methods, powered by fusion, to regain oxygen from carbon dioxide. Flames will be forbidden. The polluted sky will occasionally be blue, when the smog clears, but just as often it will be red with nitric oxide, yellow with sulphur dust, or green with noxious scum never yet seen in our own time. People will have to spend their lives encased in protective armour, and so their blood never will be seen except by police and medical robots that have neither language nor colour vision. All records of better times will long since have been destroyed by the Hedonic Legion because of their tendency to sadden those who read

<sup>19</sup> See my *Convention*.

them. Yet there will still be colour – lots of it. Graffiti will cover everything.

Could there still be semantic common knowledge of colour language? I think so. That means that our presently proposed solution, though it may well succeed under fortunate conditions, cannot be the only possible solution.

#### VII. A Solution that is Unavailable to Materialists

Some philosophers think that each sort of colour experience has a simple, ineffable, unique essence that is instantly revealed to anyone who has that experience. When I was shown the crayon mark and told that it was magenta (and I believed what I was told, and it was true) straightway I knew all there is to know about experience of magenta. I knew that it was the experience with the simple, ineffable, unique essence *E*. And that is all there is to it.<sup>20</sup> (Or perhaps it is the colour magenta itself that has the simple, ineffable, unique essence that is instantly revealed to each beholder, or anyway to each beholder with normal visual capacities in normal light.)

If this doctrine of revelation were true, presumably it would be obviously true. Even those philosophers who denied it would know it in their hearts, once they had seen a few colours and experienced the workings of revelation for themselves. Thanks to its obviousness, the general doctrine of revelation could readily become common knowledge throughout the linguistic community. Then if it were also common knowledge that everyone in the community becomes acquainted with magenta early in life (and if the community were properly dismissive of sceptical doubts about inverted spectra, etc.) it would be common knowledge throughout the community that magenta is the colour that typically causes experience with essence *E*. (Or perhaps, according to the other version of the story, that magenta is the colour that itself has the simple, ineffable, unique essence *E*.) The same would go, of course, for all the other colours that we become acquainted with and have names for. So here we would have the requisite common knowledge of the semantics of the language of colour. It would be ineffable common knowledge, but what is the harm in that? The doctrine of revelation would solve our problem.

The only remaining difficulty is that the doctrine is false. At any rate, it is false for colour experiences (and colours themselves).<sup>21</sup> At any rate, it is false by materialist lights – and we have pledged ourselves non-negotiably to materialism. The essence of a colour experience is not at all simple, not at all ineffable, not at all easily known. Probably it is a matter of neural firing patterns; but if not that, something equally esoteric. Likewise for the essences of the colours themselves.

The doctrine of revelation is tailor-made to solve our problem. But we materialists must dismiss this ‘solution’ as a useless piece of wishful thinking.<sup>22</sup>

<sup>20</sup> See Galen Strawson, ‘“Red” and Red’, *Synthese* 78 (1989) pp. 193-232; Mark Johnston, ‘How to Speak of the Colors’, op. cit., pp. 223-224 and elsewhere; David Lewis, ‘Should a Materialist Believe in Qualia?’, *Australasian Journal of Philosophy* 73 (1995) pp. 140-144.

<sup>21</sup> Maybe revelation is true in some other cases – as it might be for the part-whole relation.

<sup>22</sup> You should have noticed that my continuing difficulties result from a continuing effort to associate descriptive senses with the names of the colours and the colour experiences. Is that effort misguided? Did not Kripke and his allies refute the description theory of reference, at least for names of people and places? Then why should we expect descriptivism to work any better for names of colours and colour experiences? For the supposed refutation of descriptivism, see *inter*

So much the worse for materialism? Not if we have an alternative solution. And we do.

### VIII. How Much Common Knowledge Do We Really Need?

Consider again the case of the footy people and the rugby people: two subcommunities of the English-speaking linguistic community. For simplicity, I suppose that every English-speaker is either a footy person or a rugby person. Or both? I do not need to say whether or not the two subcommunities overlap. But they do not overlap much. Plenty of footy people know nothing of rugby, and vice versa.

The two subcommunities mingle. Often, you cannot tell whether you are talking to a footy person or a rugby person, so long as you are not talking about sport.

It is common knowledge among the footy people that the diagonal stripe on the Essendon jumper is red, that the widest horizontal stripe on the Footscray jumper is red, that Swans are mostly red, . . . None of these things are known to rugby people, except for those of them, if any, who are footy people as well.

Likewise, some examples of red things are common knowledge among the rugby people. But a footy person like me – a footy person who is not also a rugby person – has no idea what these examples are.

The footy people have their version of the added chapter of folk psychophysics that distinguishes the colours. And when they define theoretical terms as names for the occupants of theoretical roles, they arrive at their version of a definition of 'red'. This definition is part of the semantic common knowledge of the footy people, but it is not common knowledge among the English-speaking community as a whole.

Likewise the rugby people have their version of the chapter and the definition. But the two definitions are parochial. In fact, no definition of 'red' is part of the semantic common knowledge of the entire community.

(Not even 'the colour of blood'? After all, the present example is *not* set in the nightmare future previously imagined. Still, these rugby people are peculiar! Look at their

<sup>22</sup> *Continued. . .*

*alia* Saul Kripke, 'Naming and Necessity' and Keith Donnellan, 'Proper Names and Identifying Descriptions', both in *Semantics of Natural Language* (eds.) Donald Davidson and Gilbert Harman (Dordrecht: Reidel, 1972); and Michael Devitt, *Designation* (New York: Columbia University Press, 1981) pp. 3-25.

I disagree. What was well and truly refuted was a version of descriptivism in which the descriptive senses were supposed to be a matter of famous deeds and other distinctive peculiarities. A better version survives the attack: *causal descriptivism*. The descriptive sense associated with a name might for instance be 'the place I have heard of under the name "Taromeo"' or maybe 'the causal source of this token: Taromeo', and for an account of the relation being invoked here, just consult the writings of the causal theorists of reference. A brief mention of causal descriptivism, with credit to Robert Nozick, is to be found in 'Naming and Necessity' itself (p. 349, fn. 38); see also my 'Putnam's Paradox', *Australasian Journal of Philosophy* 62 (1984) pp. 226-227; and Fred Kroon, 'Causal Descriptivism', *Australasian Journal of Philosophy* 65 (1987) pp. 1-17.

Causal descriptivism explains, as the causal theory itself does not, (1) how the way reference is fixed by means of causal chains is part of our semantic common knowledge, and (2) how it depends on our contingent semantic conventions. As it does: distinctive-peculiarity descriptivism *could* have been true instead, inconvenient though that would have been, but our actual conventions of language are such as to render it false.

funny ideas about the proper shape for goalposts. Might they not also have funny ideas about blood? And even if you know they do not, do you know that everyone knows? Do you know that everyone knows that everyone knows? Do you know . . .)

When footy people and rugby people mingle, they talk. Sometimes they talk about the colours of things. It never seems to happen that there is a misunderstanding. Why not? You might expect that sometimes a footy person, wrongly thinking that his conversational partner was a footy person too, would call something red because it is red in the footy sense. Then the rugby person he was talking to, wrongly thinking that the speaker was another rugby person, would believe that the thing was red in the rugby sense – and he would get a surprise when it turned out not to be.

If such conversational mishaps occurred, we would notice them. (Or at least we would hear from others who noticed them.) And if conversationalists took precautions to prevent mishaps, we would notice that. And we would hope that those who belong to both subcommunities – if there are any of those – would warn the rest of us about the risk of misunderstandings. And yet we never notice mishaps or precautions or warnings. Why? The best explanation is that there is in fact no risk of misunderstanding, because exactly the same things that are red in the footy sense are red also in the rugby sense.

Exactly the same things; or near enough. Disagreement on a few things that both subcommunities alike would treat as borderline cases would be harmless and hard to notice.

In this way, even though neither definition was common knowledge throughout the entire linguistic community, there might still be *existential* common knowledge: common knowledge (1) that some definition of 'red' was common knowledge among the footy people, and (2) that some definition of 'red' was common knowledge among the rugby people, and (3) that these two definitions agreed (or near enough) about which things were red.

If there were only that much common knowledge throughout the entire community, that would suffice. Existential common knowledge, even without fully specific semantic common knowledge, would sustain rational semantic coordination, even between footy people and rugby people conversing about what is red.

If the footy and rugby definitions of 'red' agree in extension, that will avoid the simplest sort of mishap in communication. But if the two definitions differ in intension, is there not still some risk of a more subtle sort of mishap? Suppose two people are talking; and unbeknownst to them, one is a footy person and one is a rugby person. So long as they talk about which things *are* red, they have no problem. But what happens if they talk instead about which things *would* be red under various counterfactual circumstances? Suppose the footy definition of red is simply: 'the colour of the Essendon diagonal stripe'. Now if the Essendon Football Club were taken over by the Environmentalist Party, who are no respecters of tradition and never pass up a chance for self-advertisement, what would then be red: blood or leaves? If leaves would then be red in the footy sense yet blood would remain red in the rugby sense, our conversationalists would be in trouble if they were to try to discuss this hypothetical case. Their mismatched definitions would cause a failure of communication after all.

Yet this sort of mishap also is never observed. Why not? Because the footy definition of 'red' is not simply 'the colour of the Essendon stripe'. Rather it is: 'the colour of the Essendon stripe actually, nowadays, and hereabouts'. The latter definition is 'rigidified'.

Unlike the simple one, it can be relied upon not to shift its reference when we talk about unactualized hypothetical cases, or about actualities at remote times and places.<sup>23</sup> The rugby definition, whatever it may be, is likewise rigidified. And it is common knowledge throughout the entire linguistic community that both definitions are rigidified. In this way mishaps of communication are averted even when conversation ranges beyond the actual or beyond the here and now.

(I do not say that the simple phrase 'the colour of the Essendon stripe' is definitely *unrigidified*. Rather, I suppose it to be ambiguous with respect to rigidification. Hence the custom of rigidification has two separate advantages. One is that, as we have already seen, it can prevent failures of communication between conversational partners who use different definitions. The other is that it prevents ambiguity in the speech of a single person. The cost of a custom of rigidification is that it makes unrigidified things harder to say. To gain the advantage, and pay the cost, it is not of course required that we always signal rigidification with some special form of words.)

Do the footy people's rigidified 'red' and the rugby people's rigidified 'red' have the same intension or do they not? A straight answer would be unwise. The two have something beyond their extension in common, and something not in common. But whether 'intension' is the word for what they have in common or the word for what they have not in common is an unsettled matter.

It is a simplification to imagine that the footy people define 'red' as 'the colour of the Essendon stripe' (duly rigidified). Many examples of red are common knowledge among them. Any selected one of these, or any selection of several of these, would do for a definition. It would serve no good purpose for them to make an official choice. If they have not done so, that is only to be expected.

Or perhaps there is a default choice: the biggest class of examples that does not overstep the bound of their common knowledge. If all the examples that are common knowledge among them are mentioned in their version of the chapter of folk psychophysics that distinguishes the colours, and if red is defined to be the occupant of the resulting folk-theoretical role, that is just what will happen.

But there might be a few rotten apples in the barrel: items of common knowledge that are not knowledge at all but rather error. I said that one example of red was the widest horizontal stripe on the Footscray jumper. But that begged a question about what is a stripe and what is a stripe-shaped part of the background. Arguably, the widest stripe of all is not the red stripe, but rather the very wide blue stripe right at the bottom.

As already noted, we should be prepared to say that red is a *near-enough* occupant of a role. But which one? Once we retreat from demanding perfection, there might in principle be competing near-enough occupants. So even given a default choice, there is still some risk of semantic indecision in naming the colours. Doubtless the risk is slight, doubtless the indecision risked is also slight. Yet we can still ask what to do about it; answer: nothing. Not unless there turns out to be a real problem. The advantage of being prepared is not worth the bother of solving countless problems in advance when most of

<sup>23</sup> See Hans Kamp, 'Formal Properties of "Now"', *Theoria* 40 (1971) pp. 76-109; my 'Anselm and Actuality', *Noûs* 4 (1970) pp. 175-188; Saul Kripke, 'Naming and Necessity', pp. 273-278 and elsewhere; David Kaplan, 'Dthat' and Robert Stalnaker 'Assertion', both in *Syntax and Semantics 9: Pragmatics* (ed.) P. Cole (New York: Academic Press, 1978).

them will never arise.

The footy people and the rugby people were two linguistic subcommunities, both large. Of course the subcommunities could be smaller and more numerous, so long as the necessary existential common knowledge was available. In the extreme case, I do not see why we could not have very many subcommunities of one person each, each one with his own private collection of coloured box-beetles. At that limit, the specific common knowledge within the subcommunity would vanish. Existential common knowledge would have to do the whole job of sustaining a rational system of linguistic coordination. I see no reason why this extreme case would be impossible in principle. But certainly it would be a far cry from the real world.

*Princeton University*

Received April 1996

Revised April 1997