4

The tricks of colour

JOHN MOLLON

Figure 4.1 shows an image of the Newtonian spectrum that was prepared by one of the most vehement anti-Newtonians of the eighteenth century. It was published in 1752[1] and almost certainly it is the first image of the spectrum to have been printed in colours. It serves well to illustrate how a graphic artist can create an image of high quality without necessarily grasping the theory that underlies his craft.

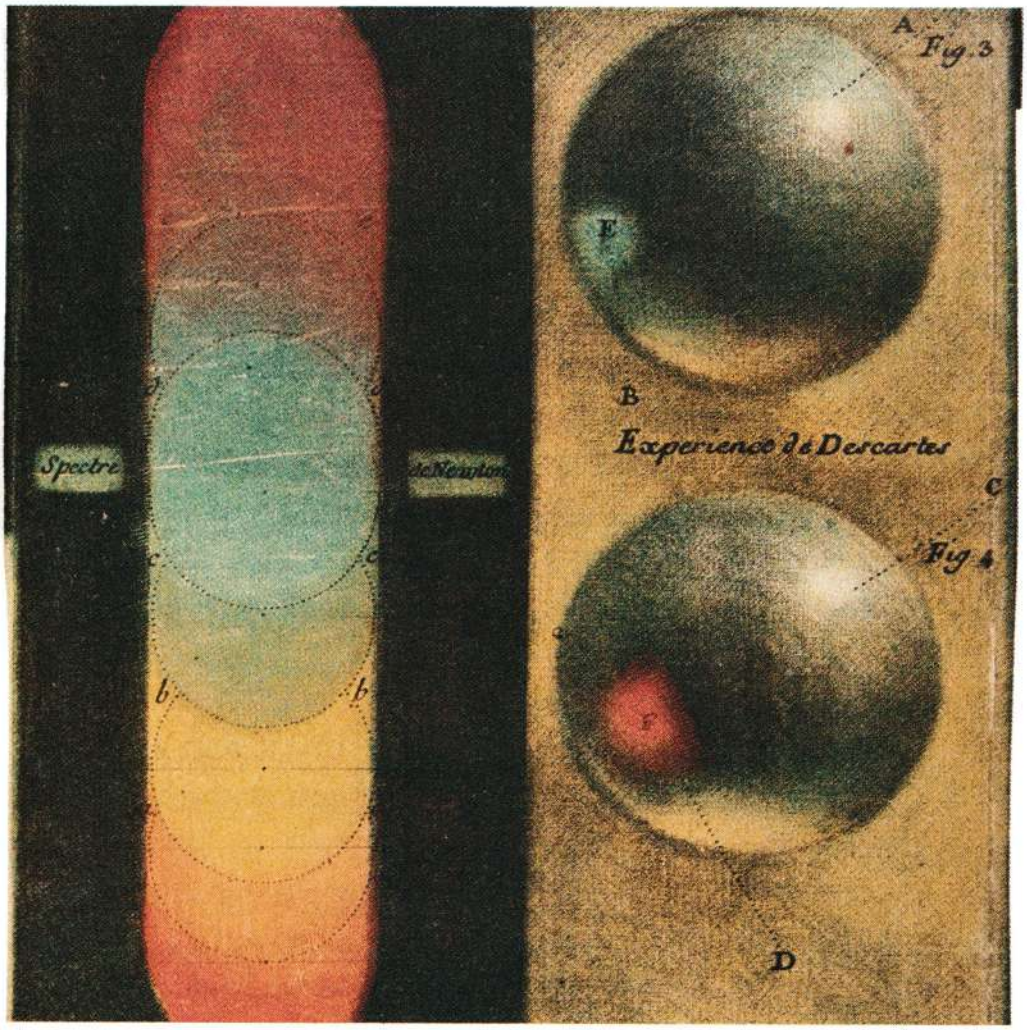


Fig. 4.1. A colour print published by Jacques Gautier D'Agoty in his popular science magazine '*Observations sur l'histoire naturelle*' in 1757 (reproduced by permission of the Syndics of Cambridge University Library).

The engraver and printer was one Gautier D'Agoty. Portalis and Béraldi[2] write of him:

> Jacques Gauthier [sic] D'Agoty, né à Marseille vers 1717, fut un de ces esprits à la fois intelligents, superficiels et inquiets, qui embrassant toutes choses à la fois, n'étreignent rien. Demi-savant et demi-artiste, il s'occupa de physique, d'anatomie, de botanique, de peinture et de gravure avec un égal insuccès.

This account gives perhaps too little credit to his artistic talent. Some of his anatomical plates – for example his extraordinary still life of the human viscera arranged on a table[3] – exhibit true flair and imagination in their composition. Many of his images lie between the surreal and the prurient; they disturb the modern eye, and we may suspect they still more disturbed his eighteenth-century contemporaries, who were not yet exposed to widespread colour printing.

But Gautier combined his successful colour printing with quite mistaken theorising about colour. He argued that Newton must be wrong in supposing that light could vary continuously in the physical property that corresponds to hue – the property we now know to be wavelength. Instead Gautier claimed that all colours were derived from three (blue, yellow and red), which were themselves derived from the interaction of white and black[4]. He found practical and daily confirmation of his theory in his mezzotint colour printing, achieved with blue, yellow and red plates, plus a black plate.

Although historical accounts of colour theory often suggest otherwise, the idea that all colours can be produced from three was already a commonplace by the middle of the eighteenth century[5]. In 1708, for example, we find a clear statement by the anonymous author of a treatise on painting in miniature[6]:

> Il n'y a proprement dit que trois Couleurs Primitives, lesquelles ne peuvent pas être composées par d'autres couleurs, mais dont toutes les autres peuvent être composées. Ces trois Couleurs sont le Jaune, le Rouge, et le Bleu. . .

As a second example, Fig. 4.2 shows part of a more elaborate statement published in 1740 by Père Castel[7], the Jesuit priest and *philosophe* who invented the 'clavecin oculaire', the first colour organ, and who was a foreign member of the Royal Society of London. It is easy to find many similar statements in the eighteenth-century, and even seventeenth-century, literature. Colour printing with three plates, first introduced by J.C. Le Blon early in the eighteenth century[8], provided empirical evidence that there were only three primitive colours. Le Blon even understood the difference between subtractive and additive mixing of colours, using the latter in his ill-fated scheme for weaving tapestries.

L'OPTIQUE
DES
COULEURS.

Fondée ſur les ſimples Obſervations, & tournée ſur-tout à la pratique de la Peinture, de la Teinture & des autres Arts Coloriſtes.

Par le R. P. CASTEL, *Jeſuite.*

A PARIS,
Chez BRIASSON, rue Saint-Jacques, à la Science.

M. DCC. XL.
Avec Approbation & Privilège du Roy.

122 L'OPTIQUE

VIII^es. OBSERVATIONS.

Sur la maniere de compoſer toutes les Couleurs avec les trois primitives.

Démonſtration de tous les dégrés poſſibles, harmoniques & pittoreſques du Coloris.

PUiſqu'il n'y a que trois couleurs ſimples, dont le mélange ſeul doit produire toutes les couleurs de la nature & de l'art, il s'agit de voir quelles couleurs peuvent reſulter des mélanges divers, qu'on peut faire de ces trois couleurs.

Tous les mélanges poſſibles ſe reduiſent d'abord en général à quatre combinaiſons. Car on ne peut mêler, 1°. que le bleu avec le jaune: 2°. Le jaune avec le rouge. 3°. Le rouge avec le bleu. 4°. Les trois enſemble.

Fig. 4.2. Title page, and description of 'the means of producing all colours from the three primitives', from *L'Optique des Couleurs* of Père Castel.

The miscategorisation of trichromacy

Of those eighteenth-century scientists who did experimental work on colour mixing, or who had practical experience of colour printing, almost all were led to conclude that there were only three kinds of light and that Newton was wrong in holding that the physical variable underlying colour was a continuous one. They were, of course, committing a category error, assigning the results of colour mixing to the wrong domain of knowledge. The fact that you can, by additive or subtractive mixing, produce all colours from three, the fact of trichromacy, is certainly true (with certain technical qualifications), but it has its basis not in the physical nature of light but in the properties of our eye. It is a fact of human physiology rather than a fact of physics. It is historically instructive that the fact of trichromacy was known empirically, but miscategorised, by artists, engravers and scientists for almost two centuries before its true nature was understood – from about the second decade of the seventeenth century until 1801. I believe that this miscategorisation of a fact

about human vision held back the understanding of physical optics more than has been appreciated by historians of science.

A general obstacle to the understanding of colour was the reluctance of the eighteenth-century mind to allow that there might be physical variations in the world that were not apparent to our senses. But the specific concept that was lacking was that of the narrowly tuned sensory transducer, the modern idea of a specialised receptor cell that converts external stimuli into the electrical signals that are the common coinage of the nervous system. The standard eighteenth-century view was that the vibrations of the air or of the ether were directly transmitted along the sensory nerves to the sensorium or sensory, the place where the soul has intercourse with the brain[9].

It was Thomas Young, celebrated for his contributions to wave optics and to the deciphering of the Rosetta Stone, who realised that the contradiction in the eighteenth-century literature could be resolved if trichromacy were taken to be a property of man, rather than a property of the physical world. Trichromacy is a limitation of our own colour vision, and one that arises, Young suggested, because there are just three classes of receptor cell in our eye[10]. Physically different lights will look the same to us provided that they stimulate the three receptor cells in the same ratios.

The receptor cells are the so-called cones, which lie in the retina of the eye and form the light-sensitive surface on which the optical image is focussed[11]. The curves of Fig. 4.3 show, for each type of cone, how the absorption of light varies with wavelength. The data are obtained by a technique called microspectrophotometry, in which a narrow beam of light is passed through individual cells from fresh human retinae (the tissue is obtained from eyes that have to be removed on account of cancer)[12]. Notice that the peak sensitivities of the three types of cone lie in the violet, the green and the yellow-green. I shall refer to the three types rather clumsily as short-wave, middle-wave and long-wave. Whatever its mnemonic convenience, the use of colour names for the different cone types has been one of the most pernicious obstacles to the proper understanding of colour vision.

It is important to emphasise that the cones themselves do not distinguish colours; the individual cones are colour blind. The electrical signal that comes out of the cone depends only on the total number of photons of light absorbed. At wavelengths at which a cone is not very sensitive, one needs more light to get the same signal from the cone than at a wavelength to which it is very sensitive; but there is nothing in the signal itself to distinguish the two wavelengths. What do vary with wavelength are the *ratios* of the signals from different cones. So, if the visual system is to know about colour, it must have the neural machinery to obtain these ratios. I shall give some description of

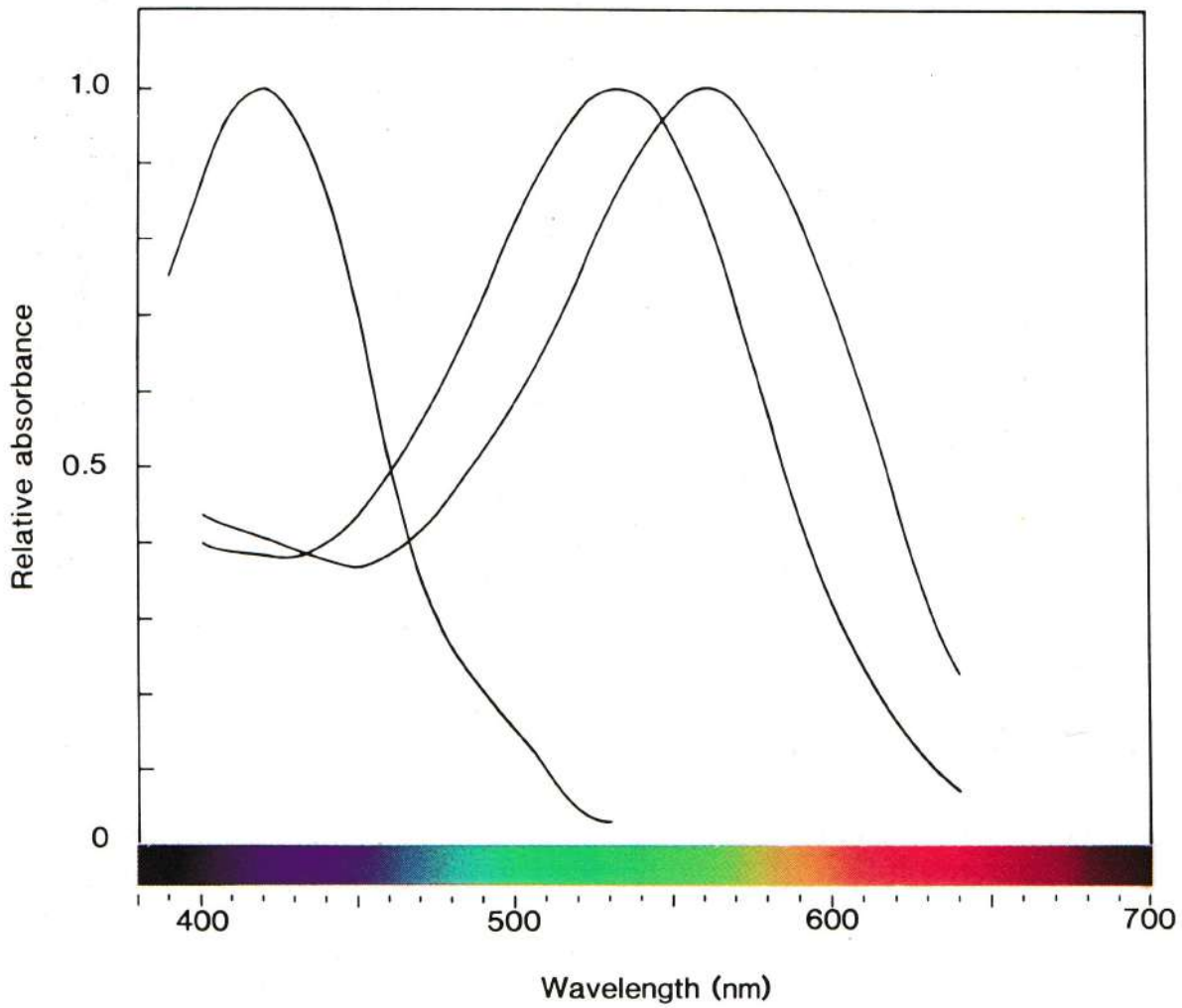


Fig. 4.3. The absorbance curves of the cone cells of the human retina. Each of the solid lines corresponds to one type of cone and shows how the absorption of light varies with wavelength (see note 12).

this neural machinery below, since colour vision offers a clear and very typical illustration of how nerves can analyse the storm of information that batters on our senses.

But before I come to the neural machinery, I should like to draw attention to several asymmetries among the three types of cone receptor, which have traditionally been thought of as equal members of a trichromatic scheme.

Five asymmetries of cone vision

1. The absorbance curves (Fig. 4.3) for the middle- and long-wave cones lie very close together in the spectrum, separated in wavelength by only 30 nm, whereas there is an interval of 100 nm between the wavelength at which the short-wave cones are most sensitive and the wavelength at which the middle-wave cones are most sensitive.

2. The short-wave cones are much rarer than the long- and middle-wave receptors, accounting for less than 5% of all cones in the human retina[12, 13].
3. Hereditary deficiencies of the short-wave cones are rare, but those of the middle- and long-wave cones are very common[14]. About 8% of men in our population exhibit some form of hereditary colour deficiency or colour anomaly: this minority in our midst live in different perceptual worlds, seeing colours as the same that are different for the normal person, and (in most cases) seeing colours as different that match for most of us. Almost invariably this hereditary change in colour vision arises from a change in either the long-wave cones or the middle-wave cones.
4. The rare genetic deficiencies of the short-wave cones, when they do occur, have the same incidence in men and women, whereas those of the middle- and long-wave cones are sex-linked, being much the more common in men than in women and showing the same pattern of inheritance as does haemophilia.
5. When our vision depends on the short-wave cones alone, we become very poor at resolving small intervals in either space or time. To demonstrate the poor resolution in *space*, one can make up a grating pattern from alternating bars of a blue and a green chosen to be of equal brightness and nearly identical in their effects on the middle- and long-wave cones. One's resolution of the grating then depends on the short-wave cones and one will have to come very close to the grating to distinguish the alternate bars – much closer than if the grating is made out of equally bright red and green bars[15]. Artists and graphic designers are aware of a distinction between 'soft' colours that melt into one another (a blue-green on a grey field) and 'hard' colours that do not (a red on a green field): the 'melting' in fact occurs when the boundary is one that is visible only to the short-wave cones[16]. It is also easy to show that our resolution in *time* is poor when we depend only on the short-wave cones. A blue and a red bar, both flickering at same rate, are presented on a bright yellow field. The yellow field serves to reduce the sensitivity of the middle- and long-wave cones, so that the blue bar is visible only by means of our short-wave cones. The red bar is primarily detected by the long-wave cones. If we slowly increase the rate of flicker of the two bars, we find that there is a range of flicker rates over which the blue bar appears steady and unflickering, while the flicker of the red bar is still clearly visible[17].

I have listed five asymmetries in our trichromatic system of cone vision. I should like to propose that these asymmetries – as well as other features of our vision – can be understood in terms of an evolutionary scheme. Our colour

vision depends on two rather different sybsystems, one recently overlaid on the other. This view has long been suggested by the very asymmetries I've mentioned, but is now particularly supported by some molecular genetics published recently by Jeremy Nathans and his colleagues[18]. Nathans has managed to isolate the genes that specify the photosensitive molecules that are found in the cone cells (Fig. 4.4) of normal and of colour-anomalous people. And for the case of one normal observer, he has been able to work out the exact sequences of amino acids that make up the three different photopigment molecules.

All the photopigments are known to be rather similar protein molecules (Fig. 4.4). In each case the molecule seven times crosses the membrane of the receptor cell in which it is embedded. It forms a kind of palisade surrounding the derivative of vitamin A that gives the molecule its light-absorbing properties. What Nathans has found is, first, that 96% of the amino-acid

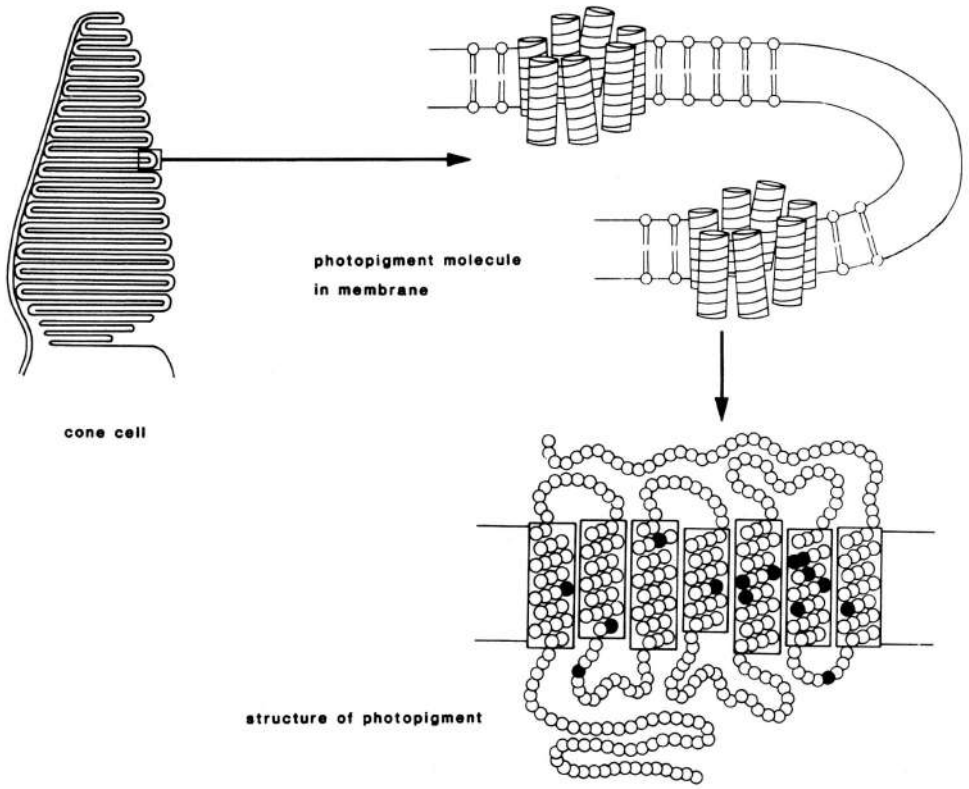


Fig. 4.4. At the top left is shown the outer segment of an individual cone cell from the retina. Its multiply-infolded membrane contains molecules of photopigment (top right). The latter have a characteristic form, crossing the membrane seven times and clustering into a palisade around the derivative of vitamin A that gives the molecule its light-absorbing properties. At the bottom right is represented (after J. Nathans) the sequence of amino acids in the protein part of the photopigment, with the filled circles indicating those amino acids that differ between the long- and middle-wave pigments of the human retina.

sequence is similar for the middle- and long-wave pigments (Fig. 4.4 bottom right), whereas the short-wave pigment shows only a 43% identity with the middle- and long-wave pigments. The second salient finding was that the genes for the middle- and long-wave pigments lie very close together. (They are both, in fact, on the X-chromosome, as has long been inferred from the pattern of inheritance of colour blindness: a man, since he inherits only one X-chromosome, will always exhibit colour blindness if his mother passes on to him an aberrant gene, whereas a woman, having two X-chromosomes, must normally inherit the aberrant gene from both parents before she will be overtly colour blind.)

The extreme similarity and the juxtaposition of the middle- and long-wave genes strongly support the long-held suspicion that they evolved very recently by duplication from a single ancestral gene. It is this evidence that leads me to argue that our colour vision is really two subsystems, a very recent system overlaid on an ancient system[19].

The subsystems of colour vision

1. Widespread among mammals is an ancient, 'dichromatic', form of colour vision that depends on a comparison of the rates at which photons of light are absorbed in (i) a short-wave class of cones and (ii) a second class of cones with peak sensitivity that varies between species but always lies in the green to yellow region of the spectrum[20]. The results of this comparison are carried within the visual system by a special neural channel.

The concept of a 'neural channel' deserves some elaboration. The cornea and lens of the eye form an optical image on the array of receptor cells in the eye, but that image is not then transmitted passively to the brain. Rather, the array of receptor cells is examined in parallel by further stages of retinal nerve cells, which identify different attributes of the images, such as colour, motion and edges. So, what is sent to the brain is already a much-analysed version of the information in the optical image formed on the retina. The subsystems, which carry information about different attributes of the image, are commonly referred to as 'channels'.

The ancient subsystem of colour vision serves well to illustrate the concept of a channel, and the way in which sensory systems extract information. Fig. 4.5a represents, to the left, the array of receptor cells in a local region of the retina, and, to the right, a later, higher-order cell that draws input from this local region. The higher-order cell draws excitatory input from one type of cone, but is inhibited by another type. Typically, such cells are excited by short wavelengths and inhibited by long wavelengths. Thus this cell becomes sensitive to the ratio of absorptions in the different cone classes, rather than to

the absolute level of light. There is no intensity of long-wave (red or yellow) light that will make the cell respond, since the inhibitory signal will always outweigh the excitatory.

The trick the visual system uses to extract information about colour is really the one main trick that sensory systems have for analysing the information in the physical image. In general terms, the trick is this: a higher-order cell is connected to two distinct subsets of cells at the preceding level and draws signals of opposite sign from the two subsets. It is the same trick, for example, that is used to extract information about edges in the image (see below).

When the first electrical records were obtained from higher-order cells of the kind shown in Fig. 4.5a, it was generally assumed that the antagonism of the inputs was there to allow 'opponent coding': it was thought that the cell signalled one colour by an increase in its response and the complementary colour by a decrease in its response. This is probably a mistaken way of looking at it. The inhibitory input from the second class of cones lends a *specificity* to the response of the higher-order cell[21]. A cone cell responds in the same way to all wavelengths – provided the light is intense enough – but the higher-order cell will give a positive response to only part of the spectrum.

Cells of the type shown in Fig. 4.5a draw their excitatory and inhibitory inputs from co-extensive, or nearly co-extensive, areas, and so they are insensitive to spatial detail, to variations in the illumination of adjacent areas[22]. Thus this ancient subsystem of colour vision carries almost pure chromatic information. It allowed our ancestors to distinguish browns and yellows from greys and whites, and blues from all of these. In other words, it allowed them – as it allows us – to estimate the direction in which the reflectance of a surface varies across the visible spectrum, and the steepness of that variation. A sign of the antiquity of such colour vision systems is the antagonism between long- and short-wave stimuli that can be recorded electrophysiologically in the pineal gland of the frog[23].

2. In the Old World monkeys, in the apes, and in man is found a second colour vision system that depends on a comparison of the rates at which photons of light are absorbed in the middle-wave cones and the long-wave cones. This system has evolved very recently through the duplication of an ancestral gene, and there does not seem to have evolved a special morphological system to carry the second type of chromatic information. Rather, this type of colour vision is parasitic upon an existing neural channel that subserves the analysis of spatial details in the image. In all mammals one finds higher-order cells in the retina that draw inputs of opposite sign from distinct, concentric sub-regions of the array of receptor cells (Fig. 4.5b). A higher-

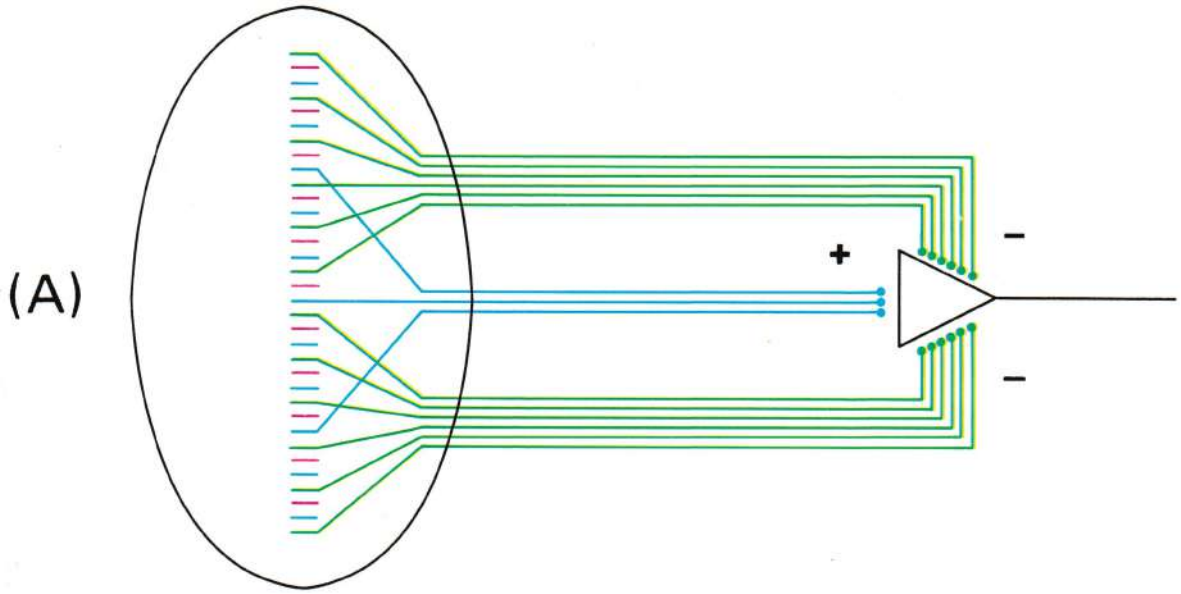
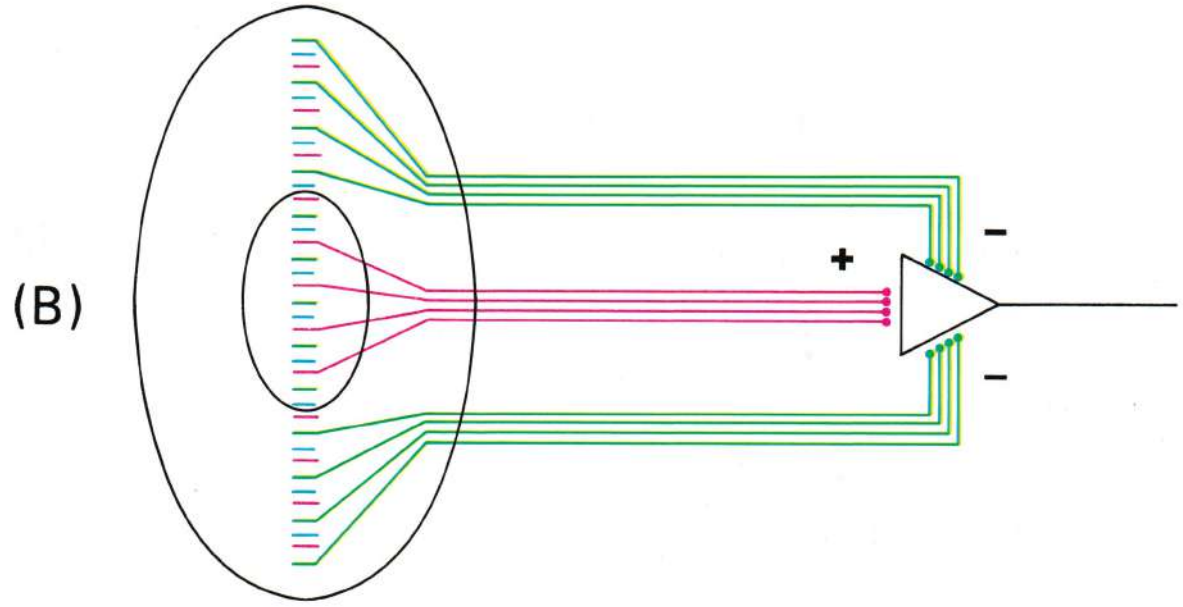


Fig. 4.5. (A) The ancient colour system of the mammalian retina. To the left is represented the array of cone cells in a local region of the retina. To the right is represented a higher-order cell (a 'retinal ganglion cell') that draws excitatory inputs from short-wave cones and inhibitory inputs from middle-wave cones. The area circumscribed by the solid line to the left is called the 'receptive field' of the ganglion cell: it is defined as the small region of receptor cells that have an influence on the response of the ganglion cell.



(B) The second colour system, found in the primate retina. The retinal ganglion cell on the right draws excitatory input from long-wave cones in the centre of the receptive field and inhibitory input from middle-wave cones in a concentric region. In other ganglion cells of this class it may be the middle-wave cones that are excitatory and the long-wave ones that are inhibitory; and the excitatory input may be drawn from the surround rather than the centre of the receptive field. (After Wiesel and Hubel.)

order cell of this kind typically gives strong response when light falls exclusively on the sub-region from which it draws excitatory input, but the cell gives little or no response when the illumination is spatially homogeneous and also falls on the other sub-region, from which the cell draws inhibitory input. Thus the cell is responsive to local contrast in the retinal image, to the spots, edges and lines that carry the information in our visual world. In the retinae of Old World monkeys, and presumably in our own retinae, such higher-order cells usually draw their antagonistic inputs not only from spatially distinct regions but from different classes of cone: the cell is either excited by middle-wave cones and inhibited by long-wave cones, or *vice versa* (Fig. 4.5b). But, significantly, there is little evidence for an input from the short-wave cones to this kind of cell[20].

Now that we have considered the evolutionary history of our colour vision, we can guess why our discrimination is so odd when it depends only on the short-wave cones. Signals from these sparsely distributed cones reach us only over a neural channel that was not designed to carry precise spatial information. If we abut a green and a blue patch, the two being equally bright and chosen to produce identical absorptions in the long- and middle-wave cones, then our visual system cannot know the exact position of the edge between the two patches. And subjectively the colours melt into one another. For our ancestral colour vision system (Fig. 4.5a) was not designed to detect spatial contrast; it was able to rely on the fact that edges in the real world almost always offer to the eye a change of lightness as well as a change of colour and so can be detected by other neural channels. But if we abut a green and a red patch of equal brightness, then we do perceive a clear and well-localised edge. For the subsystem (Fig. 4.5b) that compares the absorptions in the long- and middle-wave cones is an ancient edge-detecting system, one that still has as its main job the analysis of fine detail and so continues to be sensitive to the exact position of a boundary. It is on this detail-discriminating system that our new dimension of colour vision is parasitic.

Recognising the colours of things

I have described the trick that the visual system uses to discover the wavelengths locally present at some point on the retina. But this alone will not give us accurate and reliable colour vision. What we have evolved to do is recognise the *spectral reflectances* of objects, that is, to recognise the permanent tendency of objects to reflect some wavelengths more than others. And to achieve this, our visual system must take into account the colour of the illumination. For the spectral composition of the light that reaches us from

the object – the proportion of different wavelengths within it – depends not only on the permanent properties of the object's surface but also on the fluctuating tone of the illumination, which may change in colour from moment to moment. We enjoy a remarkable ability to make allowance for the colour of the illumination and to identify accurately the permanent colour of the object. A piece of white paper continues to look white as we pass from the red-rich light of a domestic tungsten bulb to the bluish illumination of northern daylight. 'Colour constancy' is the term traditionally used for this stability of our perception in the face of changes in the illumination[24].

Because, most of the time, our visual systems achieve colour constancy so effortlessly, direct demonstrations of the phenomenon are rather boring. It is more vividly demonstrated by the classical illusion of 'coloured shadows'[25]. Upon a screen we cast white light from one projector and pink light from a second, so that the screen appears to be flooded with a pinkish white illumination (Fig. 4.6, lower left). In the beam from the pink projector we now interpose a cardboard shape, which casts a corresponding shadow on the screen. In the area of this shadow, the screen is physically illuminated only by white light. But in fact the shadowed area looks clearly bluish green.

Such coloured shadows are usually spoken of as illusions, as tricks that the eye plays upon us; but like most illusions they serve to draw attention to the normal operations of our perceptual systems, operations that run so smoothly they go unnoticed most of the time. And the coloured shadows are an illusion only if we cast our account in terms of pink light and white light drawn from separate projectors. In fact, our visual system is making the most plausible interpretation of the stimuli presented. It assumes that the scene is lit by a single illuminant, one that is pinkish white in colour. In such an illumination, an area that does not reflect its fair share of the pink component of the illumination must be a greenish or bluish object. And that is how the shadowed area looks to us.

Figure 4.6 helps explain this point. To the left are shown two demonstrations that are normally treated as distinct – a demonstration of 'colour constancy' above, and a demonstration of a 'coloured shadow' below. The figure shows how the two demonstrations deliver the same actual stimulation to the eye (bottom right). In the case of the *'coloured shadow'*, the light reaching the eye from the tree-shaped area is white because the pink beam is obstructed. Graph (c) at the bottom right shows what is meant by saying the light is white: all wavelengths are present in equal amounts. In the case of *'colour constancy'*, the tree-shaped area is truly green: it reflects the middle wavelengths of the spectrum more than others (graph b). But when it is illuminated by a suitably chosen pinkish light (graph a), which is *deficient* in

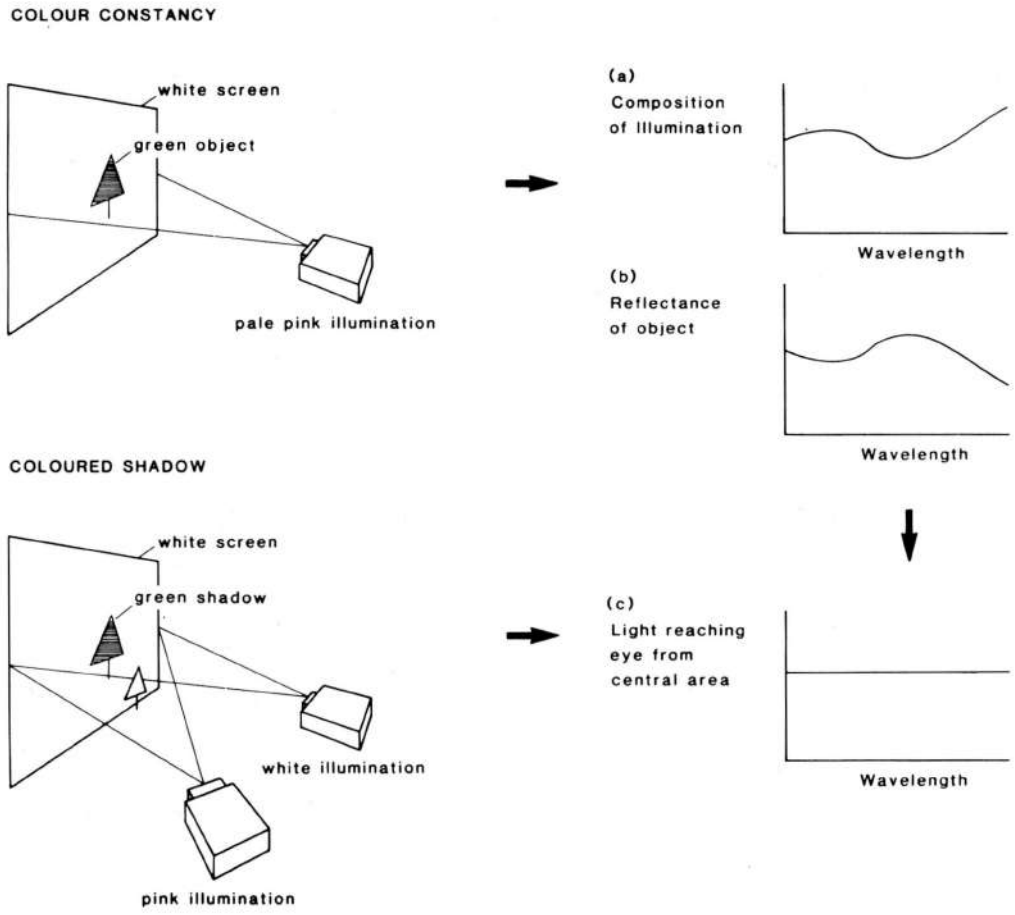


Fig. 4.6. A coloured shadow as a demonstration of colour constancy. At the bottom left a white screen is illuminated by light from two projectors, one giving pink light and one giving white light. An opaque card in the pink beam casts a shadow on the screen. This area therefore reflects only white light to the eye (c), but it appears a quite vivid, bluish green. Above is shown how a truly green object would present the same stimulation to the eye if it were presented on a white ground in a pale pink illumination (top left). If we multiply the strength of the illumination at different wavelengths (graph a) by the proportion of light reflected from the object at each wavelength (graph b), then we derive (c) the relative intensity of different wavelengths reaching the eye from the central area. Graph (c) is identical in the case of 'colour constancy' and in the case of the 'coloured shadow'. It is also the case – and this is crucial – that the light reaching the eye from the surround is identical in the two cases.

the middle of the spectrum, then the net result is that white light is reflected to the eye (graph c). What we see, nevertheless, is a green tree, because our visual system can take account of the fact that the illumination is biassed in its colour – in, that is, the relative predominance of different wavelengths. The visual system takes its estimate of the illumination from the surrounding field. Compared to the assumed illuminant, the white light from the tree-shaped area is deficient in red light: it ought therefore to derive from a

greenish or bluish surface, a surface that absorbs the red component of the light. In the case of this demonstration of 'colour constancy' we would say that our visual system is effortlessly recovering the true colours of objects, despite the biassed colour of the illuminant. In the case of the 'coloured shadow', we might speak of an illusion. But the stimulus for the eye, and the underlying process, is the same in the two cases.

How can the visual system achieve colour constancy, reconstructing the true colours of objects from the changing flux of wavelengths on the retina? Almost certainly it does it by playing its standard trick a second time, this time taking the ratio of cone signals found in a local region of the scene and comparing this ratio with the set of ratios found elsewhere in the field of view. So true colour vision depends on extracting a ratio of ratios.

The worlds of the colour blind, and of their mothers

I should like lastly to consider those who live in perceptual worlds that are different from those of the rest of us. I include here not only the 8% of men who exhibit some anomaly or deficiency of colour vision, but also the 16% of women who are carriers of colour blindness and who can be led to reveal themselves in subtle perceptual tests.

The work of Nathans and his collaborators (discussed above) has changed our views of how colour blindness arises. Most of us in this field of research thought that the various forms of red-green blindness arose from local errors in the genes for the long- and middle-wave photopigments – that is to say, from very circumscribed mutations of the sections of DNA that specify the sequence of amino acids which make up the photopigments. Instead it seems that colour deficiencies arise from a genetic phenomenon called 'unequal crossing-over', a genetic error that is encouraged by the juxtaposition, and extreme similarity, of the two genes. There is a stage in the formation of the ovum when corresponding chromosomes line themselves up, the matching strands of DNA aligning themselves at one end and working along in something of the manner of a zip fastener (Fig. 4.7A). But if one carelessly closes a zip fastener, an error of alignment sometimes occurs, with a few teeth on one side forming a little loop. In an analogous way, the middle-wave gene on one chromosome may appose itself not to its fellow gene on the second chromosome, but to the (very similar) long-wave gene on the second chromosome (Fig. 4.7B1). The slack on each chromosome will form a little loop, rather as in the case of our zip fastener.

Now, at this stage, when the two chromosomes come together, sections of the DNA of one chromosome are exchanged with corresponding sections of DNA from the second chromosome (Fig. 4.7B2). This is the process called

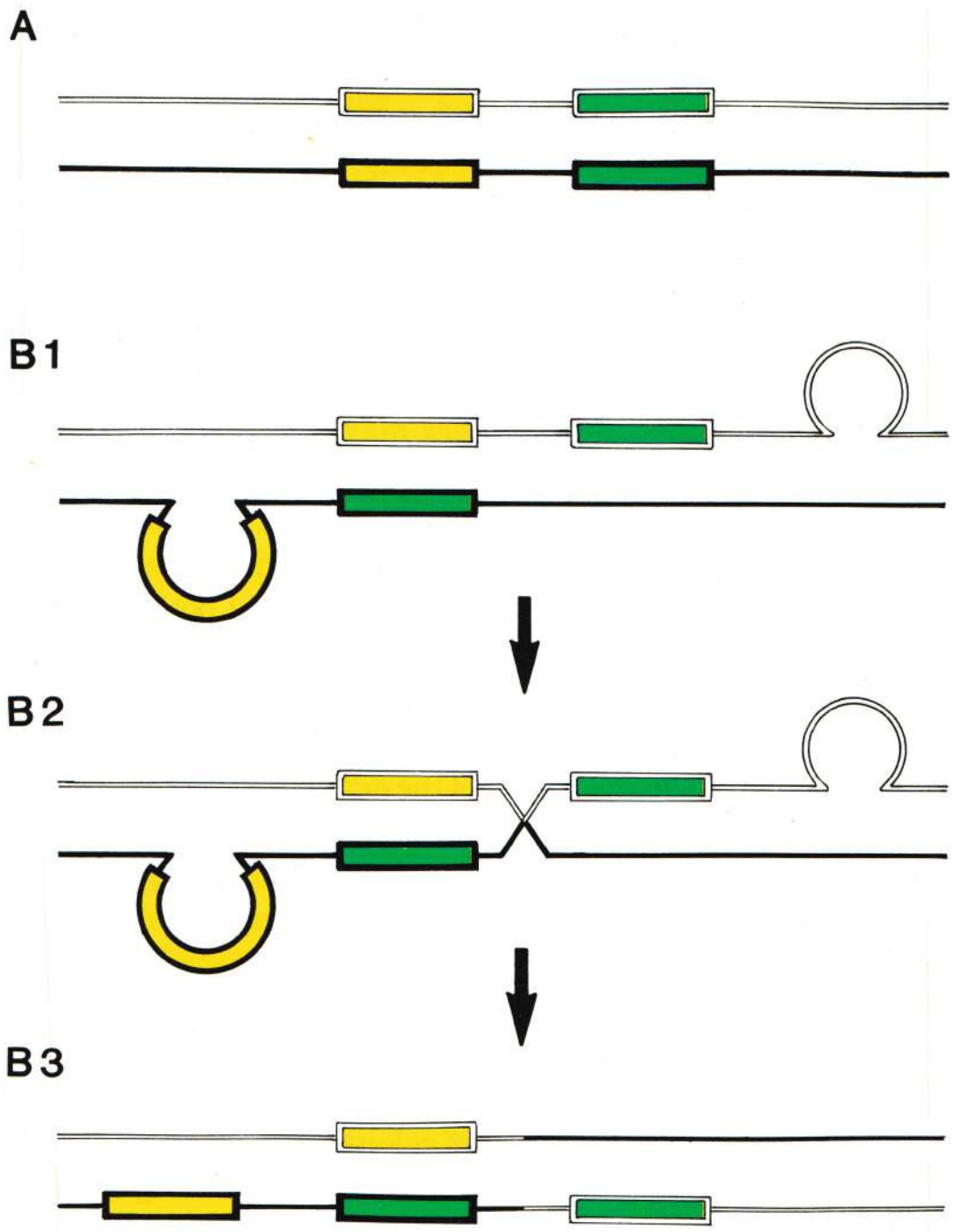


Fig. 4.7. Unequal crossing-over as an explanation of the origin of dichromatic colour vision.

(A) The normal case, where corresponding sections of two X-chromosomes are correctly aligned. The yellow boxes represent the gene for the long-wave pigment and the green boxes represent the gene for the middle-wave pigment. Between the genes lie sections of DNA that do not code for proteins.

(B1) A case where the two chromosomes have become locally misaligned, so that the middle-wave gene on one is apposed to the long-wave gene on the other.

(B2) A crossing over occurs – DNA is exchanged between the two chromosomes – and the breakpoint lies to the right of the misaligned pair of genes.

(B3) After the crossing over, one chromosome carries two middle-wave genes and the other carries none. A man who inherits the latter chromosome will exhibit the form of colour blindness called 'deuteranopia'.

'crossing-over'. If the break-point of the DNA occurs between the long-wave and middle-wave genes of one chromosome and if the genes have been misaligned, then one chromosome will lose its middle-wave gene and the other chromosome will end up with two (see Fig. 4.7B3). A man who inherits the former chromosome will lack the middle-wave pigment and exhibit the relatively severe form of colour blindness called 'dichromacy', whereas many of us with normal colour vision are carrying around extra copies of the middle-wave gene – which our forebears have unwittingly purloined from the colour blind. If, at crossing-over, the break-point occurs in the middle of two misaligned genes, then hybrid genes may be formed and these may produce hybrid photopigments with sensitivity curves different from those in Fig. 4.3. Men who inherit a hybrid gene may exhibit the milder and most common form of colour deficiency called 'anomalous trichromacy', in which one of the curves of Fig. 4.3 is displaced in its spectral position.

But what of the mothers of colour-deficient men? For some time there has been evidence for subtle anomalies in the vision of heterozygous carriers of colour-deficiency, that is, in women who have one normal X-chromosome and one that is abnormal in the region of the genes for the photopigments[26]. Consider a woman whose son exhibits either a lack of the long-wave pigment or an alteration in its sensitivity. Although her *matching* of colours is likely to be quite normal, the relative luminosities of different colours will be changed, so that red colours will look dimmer to her than to other women. This subtle change in the vision of carriers is analogous to the somewhat slower blood clotting that has been shown in carriers of haemophilia. In collaboration with J. Ellis and J. Watson, I have recently used the 'OSCAR' test of colour vision[27] to examine a group of carriers of colour deficiency. The OSCAR test measures the relative luminosities of a green light and a red light. The green and red lights are flickered out of phase and the subject is asked to adjust the depth of flicker of one so as to cancel out the flicker of the other. Subjectively, it is a matter of minimising the flicker that is seen. We tested mothers at the same time as their sons. We also tested a group of normal mothers and sons; these control subjects were all friends or neighbours of members of the first group. Figure 4.8 plots the OSCAR setting of each mother against the setting of her son: there is a very clear correlation between the two. The OSCAR test does not allow us to identify with certainty all women who are carriers; but Fig. 4.8 shows that the two groups of carriers are statistically distinct from the group of normal mothers.

The reason that visual abnormalities express themselves in carriers of colour deficiency is the phenomenon of 'X-chromosome inactivation': although a woman has two X-chromosomes, one or other of them is inacti-

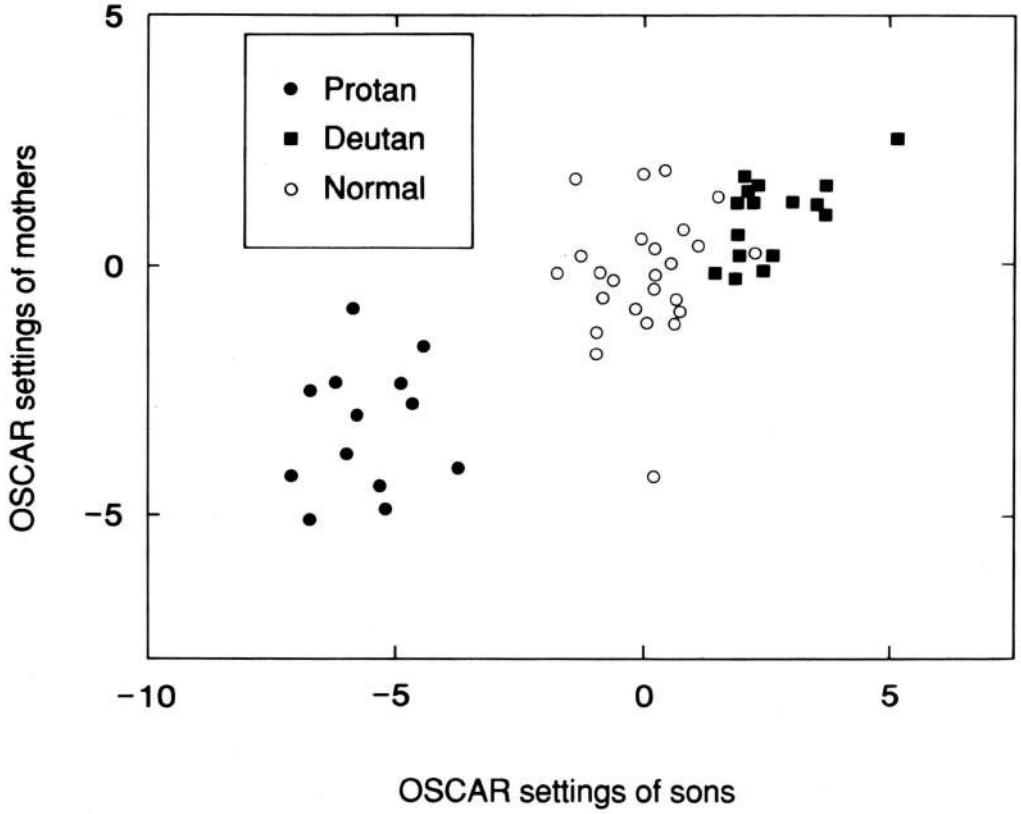


Fig. 4.8. The relationship between mothers and their sons in their settings on the OSCAR test, a test that measures the relative luminosity of red and green light. Conventionally the 3 groups of mothers would all be held to have normal colour vision, but as groups they differ in their setting according to the type of colour vision exhibited by their sons. The 'protan' sons are those who exhibit an abnormality of the long-wave cones, the 'deutan' sons those who exhibit an abnormality of the middle-wave cones.

vated in any particular cell of her body[28]. So, if a woman is a carrier of anomalous trichromacy, there will be four types of cone cell in her retina – the three types of normal vision plus the anomalous type that her son may inherit. An interesting possibility now presents itself. We might have been wrong in assuming that carriers of anomalous trichromacy merely share a little in the disability of their sons. If there are four kinds of cone in their retinae, is it possible that the carriers are tetrachromatic, enjoying an extra dimension of colour discrimination? Since some 12% of women are carriers of anomalous trichromacy, there would be a sizeable group amongst us whose colour vision was as superior to that of the normal man as his is to that of a dichromat[29]. But would visual scientists not long ago have detected a tetrachromatic minority? The truth is that the necessary experiments have never been done. Instruments for research on colour are traditionally built with three variables. If a tetrachromatic carrier said that it was difficult to get

a perfect match, a male experimenter would seldom think to offer her a fourth variable; he would accept the best match she could make and would remind himself that carriers of colour blindness are known to share in the disability of their sons.

If tetrachromatic women do exist, then a further possibility arises. Tetrachromacy may offer a biological advantage – the tetrachromat may, for example, be able to distinguish some property of complexions that is invisible to the rest of us and thus she might better choose a potential mate or better detect malaise in her infants. If the carrier did enjoy an advantage of this kind, then it might well be her advantage that maintained the high incidence of colour anomaly in the male population[30].

Conclusions

We do not all share the same world. It is certain that the tints of nature and of art look different to 8% of men; and that the relative luminosities of different colours are altered for a large minority of women. But there is also the possibility that some of these same women enjoy an additional dimension of hue discrimination, one that is unsuspected by the rest of us. Our colour vision has not yet yielded up all its secrets[31].