2

"Cherries among the Leaves": The Evolutionary Origins of Color Vision

J. D. Mollon

One of the earliest accounts of color blindness was read to the Royal Society by Huddart in 1777 and describes the case of the shoemaker Harris. In the eighteenth-century world, where the color coding of displays, electronics, and modern packaging was unknown, color deficiency was a minor handicap, but there is one task that has always challenged the color blind, and that is picking fruit. Huddart says of Harris:

> He observed also that, when young, other children could discern cherries on a tree by some pretended difference of color, though he could only distinguish them from the leaves by their difference of size and shape. He observed also, that by means of this difference of color they could see the cherries at a greater distance than he could, though he could see other objects at as great a distance as they; that is, where the sight was not assisted by the color. Large objects he could see as well as other persons; and even the smaller ones if they were not enveloped in other things, *as in the case of cherries among the leaves.* (Huddart, 1777, italics added)

A century later, Lord Rayleigh would echo Huddart, remarking that children sometimes revealed their color deficiency when gathering holly berries: "It is so difficult to see the berries among the leaves" (Strutt, 1924, p. 176). Even in a modern Australian survey, many dichromats reported difficulty in spotting "red flowers, berries, and fruit against the green foliage" (Steward & Cole, 1989).

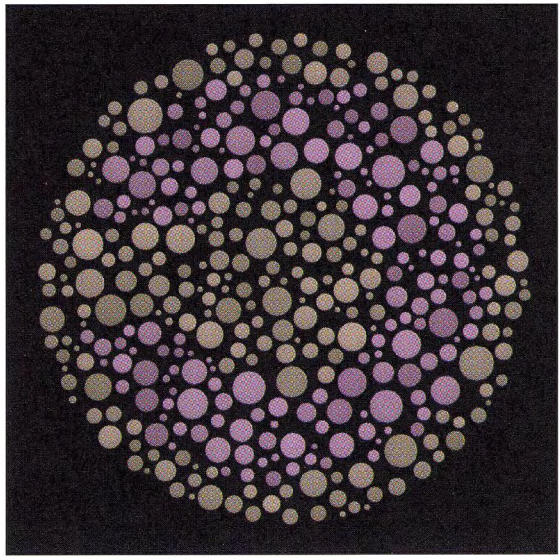
It is sometimes suggested that we need color vision in order to distinguish edges between equiluminant surfaces, that is, surfaces of different color but equal luminance (Hemlä, Reuter, & Virtanen, 1976). In fact, such edges are almost nonexistent in the natural world. Rather, we need color vision when the target is embedded in a background that is varying randomly in lightness and in form, as in the case of cherries among the leaves (Mollon, 1989).

The most effective way to isolate experimentally the color channels of the human visual system is to challenge the observer with an analogue of the natural task of gathering fruit. If we want to oblige the observer to rely on chromatic signals, it is much easier to randomize background luminance than to try to equate it for the individual subject. It was this insight that led Stilling to design the first pseudoisochromatic plates for detecting color deficiency (Stilling, 1877): the patches of the array vary randomly in lightness, and their individual contours give no clue as to the digit present. This antique principle is incorporated in a new computerized test for color deficiency that we have developed in Cambridge (Mollon & Reffin, 1989; Regan, Reffin, & Mollon, 1994): here the targets differ from the background in color, and the chromatic difference between the target and background is adjusted dynamically by the computer according to the participant's performance (Figure 1). What color does in such an array is to link elements in the field that belong together, imposing perceptual organization on the field, rather than just supporting detection.

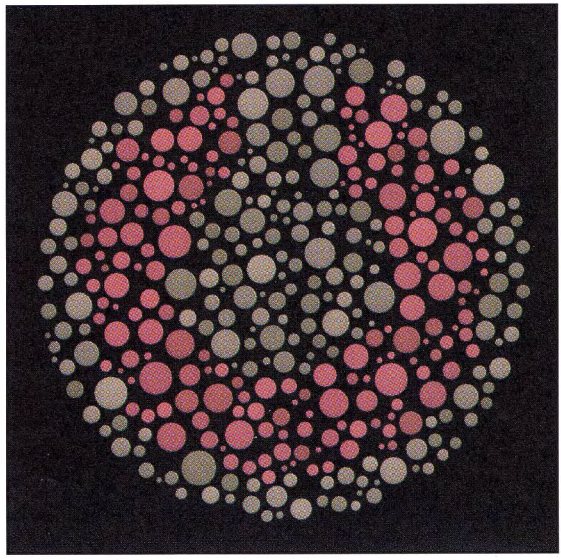
A Brief History of Primate Color Vision

To understand our own color vision, I believe we must understand how it evolved to be the way it is. I argue that it depends on two distinct subsystems, a relatively recent one overlaid on a phylogenetically ancient one (Mollon & Jordan, 1988). Almost certainly, fruits played a role in the later stages of this evolution.

The lowermost panel of Figure 2 represents the absorption curves of the three classes of light-sensitive cone cells in a typical human retina. The peak sensitivities lie in the violet, the green and the yellow-green (and much conceptual mischief is done by calling them red, green and blue). Each class of cone obeys the Principle of Univariance (Rushton, 1972): although the input to the cone can vary in radiance and wavelength, the output signal is one-dimensional, depending only on the total number of photons absorbed by the molecules of photosensitive pigment that are packed into the membranes of the cone. Graphs such as those of Figure 2 simply show, for a given type of cone, how the probability of a photon being absorbed varies as wavelength varies. So an individual cone, or an individual class of cones, cannot discriminate colors. What do vary with wavelength are the ratios of the photon catches in different classes of cone, and our color vision depends on neural channels that extract these ratios: by drawing inputs of opposite sign – excitatory and inhibitory – from different types of cone, a higher-order nerve cell becomes color-specific, responding with excitation only to that part of the spectrum where the excitatory input exceeds the inhibitory.



(a)



(b)

Figure 1. Stimulus arrays used in the Cambridge Colour Test. The subject's task is to report the orientation of the C-shaped figure. In (a) the target lies on a tritan confusion line (i.e., it is signalled by the older subsystem of colour vision). In (b) the target lies on a deutan confusion line (i.e., it is signalled by the newer subsystem of colour vision). Owing to the inaccuracies of colour reproduction, these figures should not be used for test purposes.

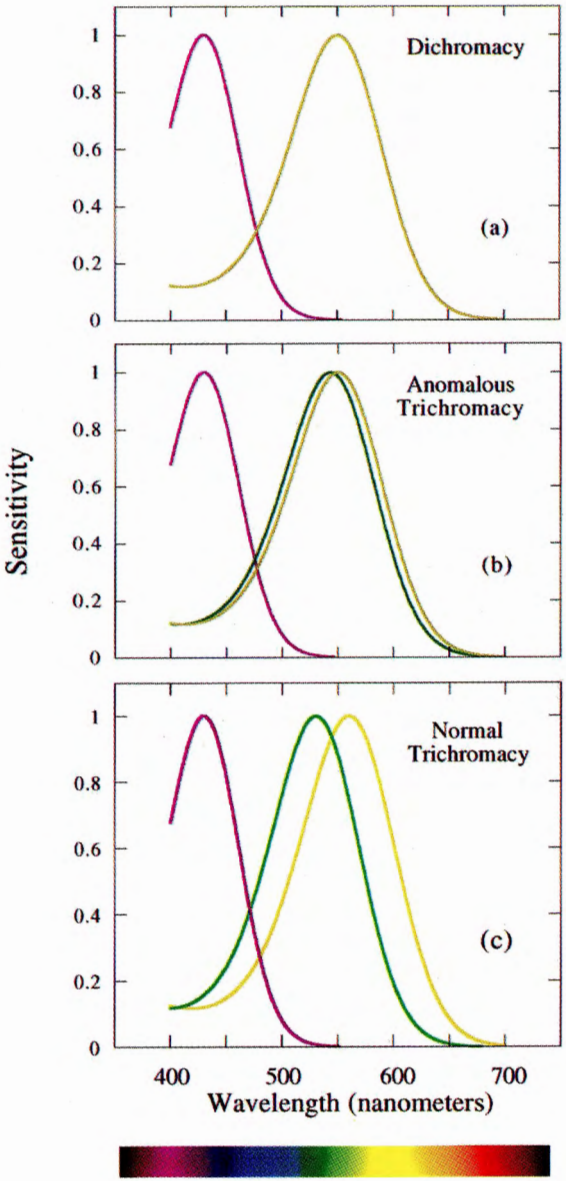


Figure 2. Stages in the evolution of colour vision. Panel (c) shows the spectral sensitivity curves of the three photosensitive pigments in the normal human eye. Ancestral mammals are thought to have had dichromatic vision (a). At an intermediate stage (b) the spectral separation of the long-wave/middle-wave pigments may have been small, and dependent on a single amino-acid difference between the two proteins.

Many fish and birds today have at least four different types of cone, a richer complement than has been found in any mammal (Bowmaker, 1998). The mammals began to diverge from reptilian ancestors as long as 300 million years ago. The photopigment complement of these ancestors is unknown, and for much of their history mammals are thought to have been nocturnal (Martin, 1990). Nevertheless, two types of cone pigment – or at least the genes that encoded them – appear to have survived continuously through this nocturnal period; for there are recognizable similarities between the DNA sequences that encode these pigments and the corresponding genes found today in other classes of vertebrates. Most modern mammals retain these two retinal pigments (Jacobs, 1993), and the two different pigments are usually segregated in different cone cells.[1] One type of cone, with peak sensitivity in the middle of the visible spectrum (500 to 570 nm), subserves the main business of vision – the detection of flicker, movement, form. But a basic color vision is achieved by comparing the photon catches in these cones with the photon catches in a second, minority type of cone that has a peak sensitivity at short wavelengths, either in the violet or in the ultraviolet (Figure 2a). I shall use the term *ancient subsystem* for the neural channel that carries this chromatic signal.

Sometime after the emergence of the primates, there appears to have been a duplication of the gene on the X-chromosome that coded for the long-wave pigment. These two X-chromosome genes diverged in their sequence until they coded for the present long-wave (L) and middle-wave (M) pigments. The phylogenetically recent subsystem of color vision depends on comparing the photon catches in these L and M pigments. But still some 8% of men (and 0.4% of women) exhibit a deficient form of color vision, resembling the earlier evolutionary stages – either dichromacy or anomalous trichromacy.

Let me elaborate in turn on each of the two subsystems and discuss the extent to which they remain separate within the primate visual system.

The Ancient Subsystem of Primate Color Vision

In primate retinas, the short-wave cones are rare, as in all mammals. In the talapoin, an Old World monkey, where James Bowmaker and I were able to measure microspectrophotometrically each cone in small patches of fovea, the short-wave (S) receptors were found to constitute 3% of all cones (Mollon & Bowmaker, 1992), while earlier measurements for Man suggested a figure of 8% (Dartnall, Bowmaker, & Mollon, 1983).

These rare S cones supply an ancient subsystem of color vision that remains morphologically distinct throughout the early stages of the visual system (Figure 3). Thus the S cones appear to have their own distinct bipolar cell, which resembles the common type of midget bipolar cell, but

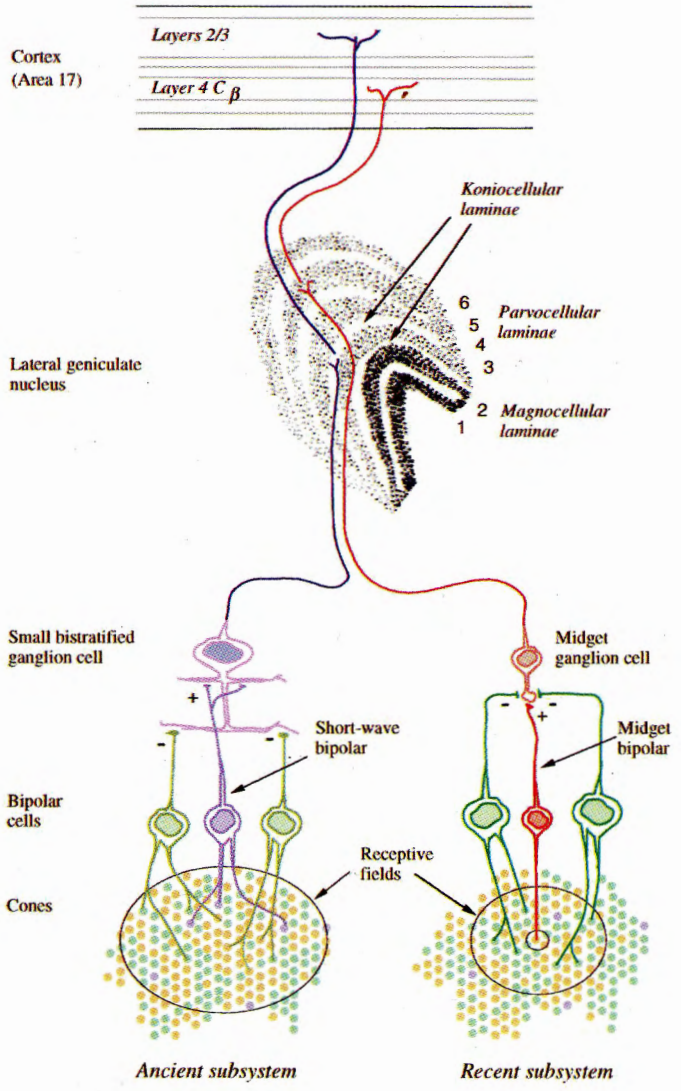


Figure 3. The two subsystems distinguished in the text. The arrays of small discs at the bottom of figure represent the distribution of cones in a small region of the fovea. On the left, the short-wave cone signal is carried forward by the short-wave bipolar and the small bistratified retinal ganglion cell, and onwards to the koniocellular layers of the lateral geniculate and the *blobs* of layers 2 and 3 of the striate cortex. On the right, an ON-centre midget ganglion cell draws its centre input from a single long-wave cone, via a midget bipolar; its signal projects to one of the ON parvocellular layers of the lateral geniculate and thence to Layer 4Cβ of the striate cortex.

makes contact with two or three separated cones rather than only one (Mariani, 1984). The signal is thought to be carried onward by a distinct, minority type of ganglion cell, the small bistratified cell recently described by Dacey: the dendritic fields of these cells are larger than those of midget ganglion cells, and lie in two separate planes at the inner and outer limits of the Inner Plexiform Layer of the retina (Dacey, 1993). Recording *in vitro* from cells of this type, Dacey and Lee have been able to show that the response is strong when the stimulus light modulates only the signal of the S cones – a stimulus that leaves silent the more common midget and parasol types of retinal ganglion cell (Dacey & Lee, 1994). These small bistratified cells constitute about 3% of all ganglion cells.

A closely similar system of bipolars and small bistratified ganglion cells has recently been described in the common marmoset, *Callithrix jacchus*, a New World species of primate (Ghosh, Martin, & Grünert, 1997). Here, too, the system appears to take its origin from cones that are labelled by antibodies for the S photopigment. Since the Old World and New World monkeys diverged some 30 million years ago, the morphological similarity of this pathway in the two primate lineages supports the idea that it is the substrate of a primordial color system, a system of some antiquity, and one that remains independent in the early stages of our own visual system (Mollon & Jordan, 1988).

In both the Old and New World cases, the axon terminals of the putative S cone bipolars end in the stratum of the retina that contains the inner dendrites of the small bistratified ganglion cells; and so this is thought to be the route of ON signals from the S cones. It is likely that inputs of the opposite sign are drawn from the L and M cones via diffuse types of bipolar cells (DRB2, DRB3), which synapse with the outer dendrites of the bistratified cell in the OFF stratum of the inner plexiform layer (Calkins, Tsukamoto, & Sterling, 1998).

This primordial color system may remain distinct at later stages of the visual pathway. Thus there is evidence that the small bistratified ganglion cells project not to the main parvocellular layers of the lateral geniculate nucleus (as traditionally thought) but to the so-called interlaminar or *koniocellular* zones, which are neurochemically distinct[2] and contain very tiny cell bodies (Calkins, Meszler, & Henry, 1998; Martin, White, Goodchild, Wilder, & Sefton, 1997). Specifically, in catarrhine (Old World) primates the small bistratified ganglion cells project to konoiocellular layers K3 and K4. These layers in turn exhibit a direct projection (Hendry and Yoshioka, 1994) to the *blobs* or puffs in layers 2 and 3 of the primary visual cortex – regions that stain strongly for the mitochondrial enzyme, cytochrome oxidase, and which contain cells that are selective for color but not for orientation. Ts'o and Gilbert (1988) report that a subset of the blobs, about 1 in 4 of them, are selective for what I am calling the ancient subsystem of color vision.

About 2% of men are dichromats and retain just this ancient subsystem, which in their case compares the signal of the S cones either with the signal of the L cones or with that of the M cones. Although such men are conventionally called color blind, they are far from living in an achromatic world. Some idea of the private color world of the dichromat can be gained from recently published simulations (Brettel, Viénot, & Mollon, 1997; Viénot, Brettel, Ott, Ben M'Barek, & Mollon, 1995). We cannot of course share the sensations of others, and in particular we cannot know whether - as traditionally supposed - the residual hues of the typical dichromat do correspond to the normal's blue and yellow; but the published simulations give an estimate of the range of the dichromat's sensations.

The sparseness of the S cones must necessarily limit the spatial resolution of our ancient subsystem of color vision. Moreover, the chromatically opponent inputs to the small bistratified ganglion cell are coextensive: the receptive field is not divided into the concentric, spatially antagonistic regions that characterize the large majority of ganglion cells. However, it would be wrong to say that the S cones have no role at all in spatial vision. To make use of the color information, we must be able to associate the color signal with a particular local region of space, and thus with a particular object. And certainly the signal of the ancient subsystem can support perceptual segregation, the linking of elements in the field that share a common color. Thus, the normal observer can readily read pseudoisochromatic plates for detecting tritanopia (color tests in which the small patches that make up the target digit are differentiated from the background patches only by the signal of the S cones; see Figure 1a). Yet in the detection of fine texture and of local discontinuities, the S cones have little role. When contrast sensitivity is measured for gratings detected only by the S cones, the peak sensitivity lies at a spatial frequency of slightly under 1 cycle per degree of visual angle, and the highest frequency that can be resolved, at maximum contrast, is only 10 cycles per degree (Cavonius & Estévez, 1975). The poor resolution of this neural channel can strikingly be seen in the effect described by Susanne Liebmann in her Berlin thesis (Liebmann, 1927; West, Spillmann, Cavanagh, Mollon, & Hamlin, 1996): if a form and its background are of equal brightness but of different hue, and if the two colors lie on a tritan confusion line – if, that is, they are differentiated only by the signal of the ancient subsystem – then the form melts into the background, its contour becoming vague and labile. The S cones do not seem to support the recognition of edges in the visual scene. It is safe to say that this sparse population of cones exists primarily for color vision, provided that one does recognize that they support those aspects of spatial vision that can be secondary to color vision, such as perceptual organization and the identification of particular objects.

Not only does the older color subsystem exhibit a distinct morphology and physiology, but it can also be impaired selectively by certain toxins and pathologies. In the nineteenth century, the anthelmintic drug Santonin was already known to produce a reversible tritanopia, temporarily reducing normal vision to a state of dichromacy in which only the L and M cone signals are available (Helmholtz, 1867). Santonin was used to kill intestinal worms, but it is listed today only for veterinary use, and Helmholtz rehearses the unpleasant side-effects: so the experiment is not recommended to the reader. However, it is interesting that a pharmaceutical company that began by manufacturing Santonin has now given us sildenafil citrate (marketed as *Viagra*), which is similarly reported to produce a short-lived impairment of the older subsystem.

Clinical evidence for the independence of this color system is offered by the case of a 40-year-old woman who suffered a binocular tritanopia that spread upwards across both visual fields in the course of four weeks (Jordan, Sarkies, & Mollon, 1990). Clinical color tests, discrimination ellipses, and increment-threshold measurements show that she has lost all access to the signals of the S cones (Regan, Reffin, & Mollon, 1994). Yet her color discrimination remains exquisite when it depends on the relative excitation of the L and M cones, and she shows no sign of retinal disease. We believe that an immunological reaction has selectively attacked the older subsystem of color vision.

The Divergence of the Genes Encoding the Long- and Middle-Wave Photopigments

Much of our recent knowledge of the evolution of color vision has come from the sequencing – by Jeremy Nathans and his colleagues – of the genes for the protein parts of the visual pigments (Nathans, Thomas, & Hogness, 1986). Figure 4 shows, to the left, the outer segment of a cone cell, with its multiply infolded membrane, in which the photopigment molecules are embedded. The protein parts of the photopigments are called *opsins*, and are members of the super-family of G-protein-coupled receptors or *heptahelicals* (Mollon, 1991): each consists of seven helices that cross the membrane and form a palisade surrounding the chromophore, 11-*cis*-retinal, which is bound to a lysine in the seventh helix of the protein (Figure 4, bottom right).

Whereas the amino acid sequence of the S cone pigment, inferred from its gene, is very different from that of the L and M pigments, the latter two are 96% homologous and they lie close together on the X-chromosome (Nathans, Thomas, & Hogness, 1986). This has been taken as primary evidence for the recent divergence of the two genes that underlie the newer subsystem of color vision. It is thought that the duplication

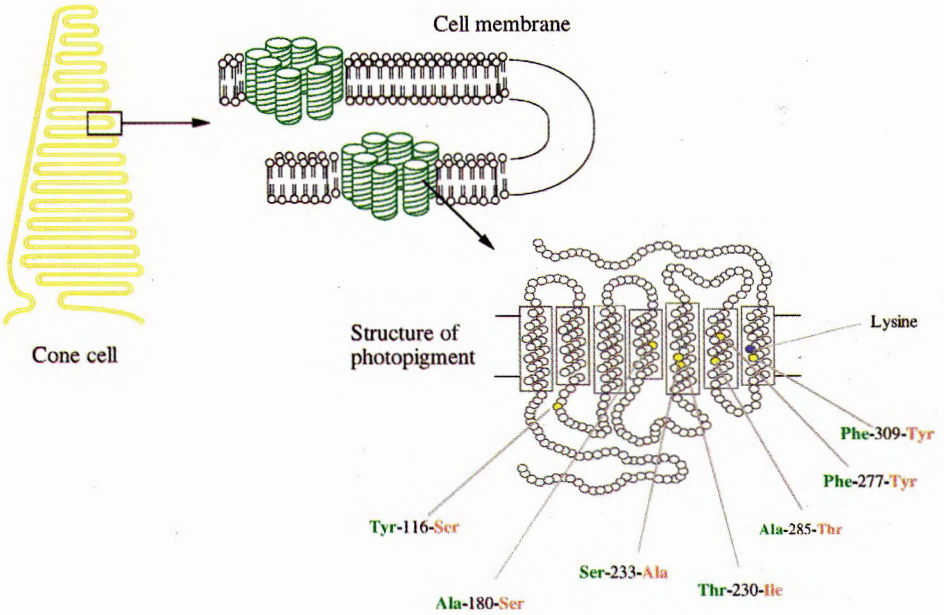


Figure 4. The enfolded membrane of the outer segment of a retinal cone (upper left), with its embedded opsin molecules. Each of the latter consists of seven helices, which span the membrane and are linked by loops outside and inside the membrane. To the lower right is represented the sequence of individual amino acids that make up the opsin, and indicated in yellow are the small number of amino acids that determine the difference in wavelength sensitivity between the long-wave and middle-wave versions of the photopigment. The coloured lettering indicates the alternative amino acids at each numbered position: in each case the amino acid shown in green is the alternative that shifts the peak sensitivity of the molecule to shorter wavelengths, while that in red is the one that gives a long-wave shift.

of the single ancestral X-chromosome gene occurred during the early history of the catarrhine primates. Either the two resulting genes then diverged so as to code for the present L and M pigments, or the original single gene was already polymorphic, as it is today in many New World species, and two alternative forms of the gene became established on a single chromosome (Mollon, 1989).

That the new color system arose after a gene duplication within the primate lineage does seem inescapable. I myself, however, have doubts about attempts to use present-day gene sequences to estimate when this duplication occurred in the phylogenetic sequence (e.g., Nei, Zhang & Yokoyama, 1997). My doubts arise from the homology of the genes' introns.

Consider the gene array on the X-chromosome that codes for the L and M cone pigments (Figure 5). Typically, there is a single copy of the L gene, but there may be several copies of the M gene – the number differs among individuals. Shown expanded below are the positions of *introns* and *exons* within each gene. The relatively short exons are the regions of the genes that code for the opsin, whereas the introns are non-coding – ostensibly regions of junk DNA. Now, in the case of Man, there is one provocative feature of the introns that is little spoken about: the introns are more similar in their nucleotide sequence than are the exons. Thus intron 2 of the L and M genes differs only at 6 nucleotides out of 1,987, whereas the 1,552 nucleotides of intron 4 are identical for the two genes (Shyue, Li, Chang, & Li, 1994). Independent data from intron 5 of a different individual shows only two differences in 2,282 nucleotides (Zhao, Hewett-Emmett, & Li, 1998). Yet there ought to be little selection pressure on these "non-coding" regions to stop them diverging. Granted, one might expect *restricted* regions of the introns to be conserved, for there is certainly evidence in other genes that introns often contain enhancer sequences and have a role in the control of expression (Storbeck, Sabourin, Waring, & Korneluk, 1998). But the near identity of whole introns strongly implies that there has been frequent and recent gene conversion in this region. Indeed, it becomes remarkable that the exons, the coding regions, of the L and M genes have retained their separate identities when the much longer introns that surround them have been homogenized. The implication is that *all* the amino acid differences between the L and M genes are under selection pressure, if not because they affect spectral tuning, then because they affect the stability and function of the molecule (Williams, Hunt, Bowmaker, & Mollon, 1992). At any rate, it is rash to infer evolutionary history from the coding regions when the introns do not tell the same story.

There is one other curious complication that is not yet well known. The intervals between the L and M genes have been thought to be filled with non-coding DNA. But it turns out that the opsin array is not the

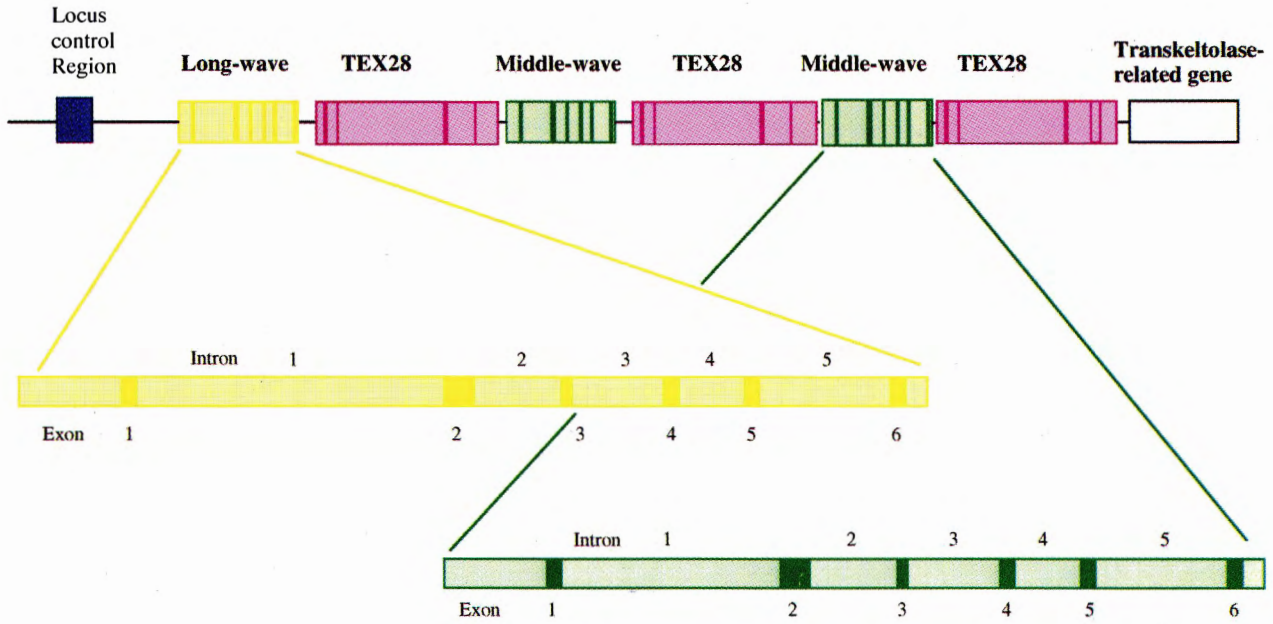


Figure 5. A simplified representation of the array of opsin genes on the q arm of the X-chromosome, with the detailed structure of the long-wave gene (yellow) and the middle-wave gene (green) expanded below. Exons are indicated by saturated colours and introns by desaturated colours. The total number of opsin genes varies among individuals. Interdigitated with the opsin genes are multiple copies of the TEX28 gene (purple in the diagram), which read in the opposite direction. Exon 1 of the TEX28 may be present only in the rightmost copy of the gene, and it is possible that only this copy is expressed (Hanna, Platts, & Kirkness, 1997).

private domain of visual science: the intervals between the L and M genes, and between successive copies of the M gene, are filled by copies of an alien gene, which reads in the opposite way and is duplicated as the opsin genes are (Hanna, Platts, & Kirkness, 1997). It is depicted in purple in Figure 5. The function of this gene is not yet known, but it is primarily expressed in the testis and its discoverers have provisionally called it TEX28. The intimate interdigitation of the opsin genes and the TEX28 gene hints that some human color anomalies might be associated with other traits.

Morphological Basis of the Second Subsystem

The newer subsystem of color vision compares the photon catches in the L and M cones. What is the morphological basis for this subsystem in the early stages of the visual system? Most commentators believe that its signals are carried by the midget bipolars and midget ganglion cells of the retina, which in turn project to the parvocellular laminae of the lateral geniculate nucleus.[3] Midget ganglion cells have tiny dendritic fields and are numerically the most common type of primate ganglion cell, constituting 60% to 80% of all retinal ganglion cells. Unlike that of the small bistratified cell, the receptive field of the midget ganglion cell is divided into distinct excitatory and inhibitory zones, making the cell sensitive to spatial contrast. In the foveal region of the retina, the midget bipolar cell contacts only a single cone; and the midget ganglion cell appears to draw its centre input from a single bipolar cell (Figure 3, bottom right). Electrophysiological recordings reveal that the centre input is from either an L or an M cone. The antagonistic input from the concentric surround of the receptive field is either drawn from cones of the opposite type, as held by Reid and Shapley (1992), or from a combination of L and M cones, as suggested by Lennie, Haake and Williams (1991).

The midget ganglion cells project to the parvocellular laminae of the lateral geniculate nucleus, two laminae being drawn from each eye. The parvocellular laminae, in their turn, project predominantly to layer 4Cβ of the striate cortex, and thence to layers 2 and 3 of the cortex – rather than directly to layers 2 and 3, as is the case for the koniocellular laminae (Figure 3).

Mollon and Jordan (1988) suggested that the second subsystem of color vision was parasitic upon an already existing neural system. Cells resembling the midget bipolars and midget ganglion cells of the macaque are present in male platyrrhine monkeys, whose retinas contain only a single class of long-wave cone (Goodchild, Ghosh, & Martin, 1996). Between dichromatic and trichromatic members of such species, there is no difference in the density or the morphology of midget gan-

glion cells. Moreover, it is implausible that 60% to 80% of retinal ganglion cells, and the four parvocellular layers of the lateral geniculate nucleus, should be given over to a system that exists only for color vision. So a plausible guess is that the parvocellular system originally existed for encoding other surface properties, notably lightness (but perhaps also texture). It is odd that lightness and color have so often been treated as independent properties in recent discussions. There are strong intrinsic correlations between chromaticity and lightness. For example, real-world colors of high purity (i.e., surfaces that reflect only a narrow band of wavelengths) can never exhibit a high lightness, unless they correspond to wavelengths near the peak of the photopic luminosity function. And subjectively, wavebands that do lie near the peak of the luminosity function (i.e., greens and yellows) change their quality as well as their apparent lightness when they are darkened, giving olives and browns.

The Role of Frugivory in the Evolution of the Second Subsystem

Although, as I have argued, we do not know the exact antiquity of the critical gene duplication, all the evidence suggests that the second subsystem of color vision is confined – within the mammals – to primates (Jacobs, 1993). I should like to develop the argument that this second subsystem co-evolved with frugivory, fruit eating. The idea that primates (and birds) are to colored fruit as bees and butterflies are to flowers is a nineteenth-century one; it was the thesis of a book by Grant Allen at a time when Darwinism was still young:

> The contrast between nuts and fruits is exactly parallel to the contrast between the wind-fertilised and the insect-fertilised flowers ... Some gay and striking tint, which may contrast strongly with the green foliage around, is needed by the developing fruit, or else its pulpiness, its sweetness, and its fragrance will stand it in poor stead beside its bright-hued compeers ... I have given this large amount of space to the consideration of fruits, because I believe we can hardly over-estimate their importance in quickening the color-sense of the higher animals, and, above all, in settling the aesthetic tastes of birds, quadrumana, and men. (Allen, 1879, p. 110 ff.)

This nineteenth-century hypothesis has become much more plausible, and more specific, as a result of work by French ecologists. Both in the Old World rainforest and in the New, there exist an important subset of trees that are disseminated exclusively by monkeys (Charles-Dominique, 1993; Gautier-Hion, Duplantier, Quris, Feer, Sourd, Decoux, et al., 1985; Julliot, 1992). In South America, for example, these include trees

of the Sapotaceae family. The fruits of such trees typically weigh from 5 to 50 g (being thus too big for birds), and often have a tough outer pericarp (restricting access to disseminators with strong teeth). There may be a single large seed, or a small number, surrounded by sweet pulp. Most notably for our present interest, these *fruits de singe* are usually orange or yellow in color. Against the dappled and variegated background of the canopy, such fruit is visible only to a trichromatic disseminator. In many cases, the color signal not only serves to allure the monkey from a distance, but also serves locally to distinguish unripe fruit from those fruit whose seed is viable, ready to be disseminated, and whose pulp has a high sugar content.[4] The tree provides the monkey with a nutritious pulp, and in return the monkey either spits out the seed at a distance or defecates it later, together with fertilizer.

In collaboration with French colleagues, Benedict Regan and I recently studied the spectrophotometric properties of fruit signals in primary rainforest in French Guiana (Regan, Julliot, Simmen, Viénot, Charles-Dominique, & Mollon, in press). We asked quantitatively whether the retinal photopigments of primates are optimized for the discrimination task of spotting fruit against a background of foliage, where lightness varies randomly. One primate species that we particularly studied was the red howler monkey, *Alouatta seniculus*, a species whose color vision is known to resemble that of a normal human trichromat (Jacobs, Neitz, Deegan, & Neitz, 1996).

We have been concerned to measure spectroradiometrically the fruits taken by *Alouatta* under natural conditions in intact primary rain forest and to measure at the same site the background foliage against which fruit signals must be discriminated. By following troops of howlers, we were able to obtain fresh samples of fruits that were actually harvested by these monkeys. Two of the most common fruits eaten by *Alouatta* in French Guiana are those of *Chrysophyllum lucentifolium* and *Pouteria guianensis*, and it is significant that primates are essentially the sole disseminators for these trees. These favourite fruits are typical *fruits de singe*, becoming yellow or orange when ripe. But we allowed in our sample a much wider range of fruits, including most of the fruits eaten by *Alouatta* during the period of our study. We also obtained analogous spectra for many samples of background foliage.

The monkey's foraging task, of finding fruits embedded in foliage, can be regarded as a signal detection problem, and in the analysis of our spectroradiometric data, we asked how the monkey's cone pigments should be placed in the spectrum to maximize the signal-to-noise ratio of fruit against foliage. Such an analysis is facilitated by the fact that the absorbance curves of different photopigments can be described by a single polynomial, if they are expressed in terms of log frequency rather than wavelength (Baylor, Nunn, & Schnapf, 1987). Knowing the reflec-

tion spectra of our samples, and also having measurements of the illuminant in the forest canopy, we can calculate the relative photon catches that a given sample would produce in any possible set of retinal photopigments (Regan, Julliot, Simmen, Viénot, Charles-Dominique, & Mollon, 1998).

Our analysis offers an explanation of why primate photopigments are so asymmetrically placed in the spectrum, with their peaks near 430, 530 and 560 nm. It turns out that such an arrangement maximizes the signal-to-noise ratio of fruit to foliage in the newer subsystem of color vision, the subsystem that draws its opposed inputs from the 530-nm and 560-nm pigments. With these pigments, the standard deviation of the noise distribution – the range of neural signals produced by the foliage – is minimized and does not overlap with the signals produced by ripe fruit. If the cone pigments were more evenly placed in the spectrum, if the newer subsystem took its input from, say, pigments peaking at 485 and 560 nm, then the neural signals produced by the foliage would have a broader distribution and would overlap with those of the fruit.

If the second subsystem evolved for frugivory, we can readily understand why its spatial properties differ from those of the older subsystem. The midget retinal ganglion cells, drawing their centre inputs from a single L or M cone, seem optimized for detection of nearly punctate fruit signals at a distance.[5]

Coevolution of Fruit Signals and Primate Color Vision?

To those who study the cycle of regeneration in the rainforest, it is easy to see primates as unconscious orchardists, disseminating the tree species that provide their staple nourishment. From a humbler viewpoint, we might conceive of the whole primate lineage as an invention of trees for propagating themselves. The evolutionary relationship between primates on the one hand, and trees with big fleshy fruits on the other, is a large and unresolved issue, but it lies behind the more specific question considered here. We have seen that monkey photopigments are optimally placed in the spectrum for discriminating fruit from foliage. This might be mere coincidence, but, assuming that there is a causal relationship, what is the direction of causality? It is possible either that trichromatic primate color vision evolved for detection of pre-existing fruit signals, or that signalling fruits adapted themselves to pre-existing properties of primate color vision. For example, primates might have evolved trichromatic color vision in order to detect small red fruits primarily disseminated by birds; and these trees, or others, might later have adjusted their fruits so as to specialize in dissemination by primates.

But an interesting possibility is that primate trichromacy and *fruits de singe* coevolved. Suppose that 30 million years ago, occasional sports occurred among the fruits of a given species of tree, sports that were slightly yellower than the foliage in which they were embedded; and suppose in the same forest there occasionally arose ancestral primates whose vision resembled that of anomalous trichromats and depended on the presence of two slightly different forms of a long-wave photopigment (as, say, in Figure 2b). The yellower fruits would enjoy an advantage by their greater visibility to the anomalous monkeys, and the anomalous monkeys would enjoy the advantage that came from being able to spot such fruit sports. As the two mutants rose in frequency, we might expect new variants to arise - fruits that were still yellower and retinal photopigments that differed by more than one amino acid and so exhibited greater spectral separation. As each trait became more marked, in the plant and in the animal, the advantage of the complementary trait would be enhanced, and so the two advantages would reinforce each other.

How would a minimally trichromatic monkey arise in the first place? It is easy to imagine how unequal crossing-over of DNA during meiosis might place on one X-chromosome two allelic forms of one opsin gene; but this would not be enough, since presumably the two corresponding photopigments would need to be expressed in different cones if their photon catches were to be compared for the purposes of color vision. Perhaps what we now term the *locus control region* (see Figure 5) was already part of the machinery of expression of the single ancestral gene. If the duplication event included duplication of the promoter region just upstream of each gene, then random coupling of the LCR to one or other promoter region would ensure that the alternative pigments were segregated in different subsets of cones. Once such segregation was achieved, chromatically opponent ganglion cells might have emerged automatically, if (as I have suggested above) there was already a system of midget ganglion cells that drew their centre input from a single cone and their surround input from unselected cones in a concentric annulus. Such is thought to be the route to trichromacy in female squirrel monkeys, where the segregation of alternative pigments in different cones is achieved by random inactivation of one or other X-chromosome (Mollon, Bowmaker and Jacobs, 1984). The cerebral cortex is thought to be designed for recognizing inputs that are correlated in time (Weliky & Katz, 1997). Once there are different subsets of midget ganglion cells, with centre inputs driven by single L or M cones, it is easy to imagine how Hebbian processes would lead to color-specific units in the cortex (i.e., units that collected their inputs from midget ganglion cells with correlated responses).

Conclusion

I have argued that our color vision depends on two subsystems that remain separate at early stages of the visual pathway. The older subsystem compares the photon catch in short-wave cones with that in long- and middle-wave cones; this subsystem's spatial resolution is poor, its signals are carried by morphologically distinct neurons, and it probably antedates the mammals. The second subsystem compares the photon catches of the L and M cones; its spatial resolution is good, it probably evolved with frugivory in monkeys, and it may have been parasitic upon an existing parvocellular system in the primate visual pathway. In acquired or inherited pathologies, we may examine the properties of each subsystem in isolation; and in the normal observer, the two neural channels can be independently adapted by psychophysical methods (Krauskopf, Williams, & Heeley, 1982). Yet our subjective experience of color exhibits no discontinuities: our sensations range seamlessly over the full gamut of hues – and lightnesses. Moreover, the yellow-blue and red-green axes of our subjective color space do not map conveniently onto the two chromatic signals discussed in this chapter (Mollon & Jordan, 1997). These things are mysteries.

Acknowledgment

I am grateful to Dr. B. Regan for comments on the text.

Notes

1 For an exception see Röhlich, van Veen, & Szél (1994).
2 Unlike the parvo- and magnocellular laminae, they are immunoreactive for the α subunit of type II calmodulin-dependent protein kinase (Hendry and Yoshioka, 1994).
3 The minority view of R. W. Rodieck has to be acknowledged here. He has proposed that all color information is conveyed by a group of ganglion cells that show a bistratified morphology and that synapse on to a minority population of LGN units that were termed Type II by Wiesel and Hubel (Rodieck, 1991). The latter cells show chromatic opponency but not spatial opponency: the opposed inputs, from different classes of cone, are coextensive. Clearly Rodieck has turned out to be correct in the case of the ancient subsystem, but he has few followers in the case of the newer subsystem. For one argument against Rodieck's position, see Mollon (1996).
4 "... it is noticeable that fruits themselves are sour, green, and hard during their unripe stage, that is to say, before the seeds are ready to be severed from the mother-plant; and that they only acquire their sweet taste, brilliant color, and

soft pulp just at the time when their mature seeds become capable of a separate existence" (Allen, 1879, p. 110).

5 "The primary necessity which led to the development of the sense of color was probably the need of distinguishing objects much alike in form and size but differing in important properties, such as ripe and unripe or eatable and poisonous fruits, flowers with honey or without, the sexes of the same or of closely allied species. In most cases the strongest contrast would be the most useful, especially as the colors to be distinguished would form but minute spots or points when compared with the broad masses of tint of sky, earth or foliage against which they would be set" (Wallace, 1889).

References

Allen, G. (1879). *The colour-sense: Its origin and development.* London: Trübner.

Baylor, D. A., Nunn, B. J., & Schnapf, J. L. (1987). Spectral sensitivity of cones of the monkey *Macaca fascicularis. Journal of Physiology, 390,* 145–160.

Bowmaker, J. K. (1998). Evolution of colour vision in vertebrates. *Eye, 12,* 541–547.

Brettel, H., Viénot, F., & Mollon, J. D. (1997). Computerized simulation of color appearance for dichromats. *Journal of the Optical Society of America A, 14,* 2647–2655.

Calkins, D. J., Meszler, L. B., & Henry, S. H. C. (1998). Multiple ganglion cell pathways provide input to the koniocellular neurons of the macaque LGN. *Investigative Ophthalmology and Visual Science, 39,* S238.

Calkins, D. J., Tsukamoto, Y., & Sterling, P. (1998). Microcircuitry and mosaic of a blue-yellow ganglion cell in the primate retina. *Journal of Neuroscience, 18,* 3373–3385.

Cavonius, C., & Estévez, O. (1975). Contrast sensitivity of individual colour mechanisms of human vision. *Journal of Physiology, 248,* 649–662.

Charles-Dominique, P. (1993). Speciation and coevolution: An interpretation of frugivory phenomena. *Vegetatio, 107/108,* 75–84.

Dacey, D. M. (1993). Morphology of a small-field bistratified ganglion cell type in the macaque and human retina. *Visual Neuroscience, 10,* 1081–1098.

Dacey, D. M., & Lee, B. B. (1994). The "blue-on" opponent pathway in primate retina originates from a distinct bistratified ganglion cell type. *Nature, 367,* 731–735.

Dartnall, H. J. A., Bowmaker, J. K., & Mollon, J. D. (1983). Human visual pigments: Microspectrophotometric results from the eyes of seven persons. *Proceedings of the Royal Society of London B, 220,* 115–130.

Gautier-Hion, A., Duplantier, J. M., Quris, R., Feer, F., Sourd, J. P., Decoux, G., Dubost, G., Emmons, L., Erard, C., Hecketsweiler, P., Moungazi, A., Roussilhon, C., & Thiollay, J. M. (1985). Fruit characters as a basis of fruit choice and seed dispersal in a tropical forest vertebrate community. *Oecologia, 65,* 324–337.

Ghosh, K. K., Martin, P. R., & Grünert, U. (1997). Morphological analysis of the blue cone pathway in the retina of a New World monkey, the marmoset *Callithrix jacchus. Journal of Comparative Neurology, 379,* 211–225.

Goodchild, A. K., Ghosh, K. K., & Martin, P. R. (1996). Comparison of photoreceptor spatial density and ganglion cell morphology in the retina of human, macaque monkey, cat, and the marmoset *Callithrix jacchus. Journal of Comparative Neurology, 366,* 55–75.

Hanna, M. C., Platts, J. T., & Kirkness, E. F. (1997). Identification of a gene within the tandem array of red and green color pigment genes. *Genomics, 43*, 384–386.

Helmholtz, H. (1867). *Handbuch der Physiologischen Optik* (1st ed.). Leipzig: Leopold Voss.

Hemlä, S., Reuter, T., & Virtanen, K. (1976). The evolution of colour-opponent neurones and colour vision. *Vision Research, 16*, 1359–1362.

Hendry, S. H. C., & Yoshioka, T. (1994). A neurochemically distinct third channel in the macaque dorsal lateral geniculate nucleus. *Science, 264*, 575–577.

Huddart, J. (1777). An account of Persons who could not distinguish Colours. *Philosophical Transactions of the Royal Society, 67*, 260–265.

Jacobs, G. H. (1993). Mammalian colour vision. *Biological Reviews, 67*, 413–471.

Jacobs, G. H., Neitz, M., Deegan, J. F., & Neitz, J. (1996). Trichromatic colour vision in New World monkeys. *Nature, 382*, 156–158.

Jordan, G., Sarkies, N., & Mollon, J. D. (1990). Clinical evidence for the independence of two sub-systems of colour vision. *Perception, 19*, 360.

Julliot, C. (1992) *Utilisation des ressources alimentaires par le singe hurleur roux,* Alouatta seniculus (Atelidae, Primates), *en Guyane: Impact de la dissémination des graines sur la régeneration forestière.* Unpublished doctoral dissertation, Université de Tours.

Krauskopf, J., Williams, D. R., & Heeley, D. W. (1982). Cardinal directions of color space. *Vision Research, 22*, 841–856.

Lennie, P., Haake, P. W., & Williams, D. R. (1991). The design of chromatically opponent receptive fields. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 71–82). Cambridge, MA: MIT Press.

Liebmann, S. (1927). *Über das Verhalten farbiger Formen bei Helligkeitsgleichheit von Figur und Grund.* Thesis, Friedrich-Wilhelms-Universität, Berlin.

Mariani, A. P. (1984). Bipolar cells in monkey retina selective for the cones likely to be blue-sensitive. *Nature, 308*, 184–186.

Martin, P. R., White, A. J. R., Goodchild, A. K., Wilder, H. D., & Sefton, A. E. (1997). Evidence that blue-on cells are part of the third geniculocortical pathway in primates. *European Journal of Neuroscience, 9*, 1536–1541.

Martin, R. D. (1990). *Primate origins and evolution.* Princeton: Princeton University Press.

Mollon, J. D. (1989). "Tho' she kneel'd in that Place where they grew . . .". *Journal of Experimental Biology, 146*, 21–38.

Mollon, J. D. (1991). Hue and the heptahelicals. *Nature, 351*, 696–697.

Mollon, J. D. (1996) The evolution of trichromacy. In N. Elsner & H.-U. Schnitzler, *Brain and Evolution* (pp. 124–139). Stuttgart: Georg Thieme.

Mollon, J. D., & Bowmaker, J. K. (1992). The spatial arrangement of cones in the primate fovea. *Nature, 360*, 677–679.

Mollon, J. D., Bowmaker, J. K., & Jacobs, G. H. (1984). Variations of colour vision in a New World primate can be explained by polymorphism of retinal photopigments. *Proceedings of the Royal Society, London, B, 222*, 373–399.

Mollon, J. D., & Jordan, G. (1988). Eine evolutionäre Interpretation des menschlichen Farbensehens. *Die Farbe, 35/36*, 139–170.

Mollon, J. D., & Jordan, G. (1997). On the nature of unique hues. In C. Dickinson, I. Murray, & D. Carden (Eds.), *John Dalton's colour vision legacy* (pp. 381–392). London: Taylor and Francis.

Mollon, J. D., & Reffin, J. P. (1989). A computer-controlled colour vision test that combines the principles of Chibret and of Stilling. *Journal of Physiology, 414*, 5P.

Nathans, J., Thomas, D., & Hogness, D. S. (1986). Molecular genetics of human color vision: The genes encoding blue, green, and red pigments. *Science, 232*, 193–202.

Nei, M., Zhang, J., & Yokoyama, S. (1997). Color vision of ancestral organisms of higher primates. *Molecular Biology and Evolution, 14*, 611–618.

Regan, B. C., Julliot, C., Simmen, B., Viénot, F., Charles-Dominique, P., & Mollon, J. D. (1998). Frugivory and colour vision in *Alouatta seniculus*, a trichromatic platyrrhine monkey. *Vision Research., 38*, 3321–3327.

Regan, B. C., Reffin, J. P., & Mollon, J. D. (1994). Luminance noise and the rapid determination of discrimination ellipses in colour deficiency. *Vision Research, 34*, 1279–1299.

Reid, R. C., & Shapley, R. M. (1992). Spatial structure of cone inputs to receptive fields in primate lateral geniculate nucleus. *Nature, 356*, 716–718.

Rodieck, R. W. (1991). Which cells code for color? In A. Valberg & B. B. Lee (Eds.), *From pigments to perception*. New York: Plenum.

Röhlich, P., van Veen, T., & Szél, A. (1994). Two different visual pigments in one retinal cone cell. *Neuron, 13*, 1159–1166.

Rushton, W. A. H. (1972). Pigments and signals in colour vision. *Journal of Physiology, 220*, 1–31P.

Shyue, S.-K., Li, L., Chang, B. H.-J., & Li, W.-H. (1994). Intronic gene conversion in the evolution of human X-linked color vision genes. *Molecular Biology and Evolution, 1*, 548–551.

Steward, J. M., & Cole, L. (1989). What do color vision defectives say about everyday tasks? *Optometry and Visual Science, 66*, 288–295.

Stilling, J. (1877). *Die Prüfung des Farbensinnes beim Eisenbahn- und Marine-personal*. Cassel: Theodor Fischer.

Storbeck, C. J., Sabourin, L. A., Waring, J. D., & Korneluk, R. G. (1998). Definition of regulatory sequence elements in the promoter region and the first intron of the myotonic dystrophy protein kinase gene. *Journal of Biological Chemistry, 273*, 9139–9147.

Strutt, R. J. (1924). *John William Strutt: Third Baron Rayleigh*. London: Edward Arnold.

Ts'o, Y., & Gilbert, C. D. (1988). The organization of chromatic and spatial interactions in the primate striate cortex. *Journal of Neurophysiology, 8*, 1712–1727.

Viénot, F., Brettel, H., Ott, I., Ben M'Barek, A., & Mollon, J. D. (1995). What do colour-blind people see? *Nature, 376*, 127–128.

Wallace, A. R. (1889). *Darwinism*. London: Macmillan.

Weliky, M., & Katz, L. C. (1997). Disruption of orientation tuning in visual cortex by artificially correlated neuronal activity. *Nature, 386*, 680.

West, M., Spillmann, L., Cavanagh, P., Mollon, J., & Hamlin, S. (1996). Susanne Liebmann in the critical zone. *Perception, 25*, 1451–1495.

Williams, A. J., Hunt, D. M., Bowmaker, J. K., & Mollon, J. D. (1992). The polymorphic photopigments of the marmoset: Spectral tuning and genetic basis. *EMBO Journal, 11*, 2039–2045.

Zhao, Z., Hewett-Emmett, D., & Li, W.-H. (1998). Frequent gene conversion between human red and green opsin genes. *Journal of Molecular Evolution, 46*, 494–496.