Colour Vision: Why Are We Primates Unique?

Petroc Sumner

Colour is such an integral part of our visual experience that most people assume that the colours we see are physical properties of objects themselves – a certain flower looks blue because the flower is blue. In one sense this is approximately true: each object has a reflectance spectrum, which specifies what proportion of each wavelength of light will be reflected by this object. If more photons in the shorter wavelength region of visible light are reflected we perceive the object as bluish, whereas a high ratio of longer wavelengths is seen as red. However this simple correspondence is only part of the story. While reflectance spectra are the property of the objects themselves, the colours you perceive are created in the brain, determined by many factors such as your previous experience, the environment the objects are seen in, and above all, the way your visual system samples the light it receives (i.e. the spectral sensitivities of your photoreceptors).

In order to have colour vision, an animal must possess photoreceptors with different absorption spectra, because for any single type of photoreceptor, any incoming spectrum can be made to look like any other by suitable adjustments of intensity. If an animal has two types of cone photoreceptor, the creature can discriminate lightness and one dimension of hue, and its vision is said to be dichromatic. For example, human dichromats may be able to make two blue flowers that look identical to us, may, for example, be perceived as completely different colours by birds. Mammals, however, seem to have retained only one class of cones.

Evolution of photopigments

The earliest vertebrates arose about 520 million years ago and it is thought that they possessed a single photopigment in cone-like photoreceptors that possessed some rod-like features. By the time the earliest tetrapod developed, around 350 million years ago, the single pigment had diverged into five distinct families. Figure 1 shows typical sensitivity curves, a simplified phylogeny and alternative naming conventions of these five classes of pigments. Various species of teleost fish, amphibians, lizards and birds have been found to possess all five pigment families, normally one in rods and four in cones. This means that they have the potential for tetrachromatic vision and consequently, two blue flowers that look identical to us, may, for example, be perceived as completely different colours by birds. Mammals, however, seem to have retained only three classes of pigment: the rod class and the longest- and shortest-wavelength-sensitive cone classes. Ancestral mammals were mostly active at night or at dawn/dusk, and not in full daylight, so a trade-off between colour vision and absolute sensitivity occurred. In many nocturnal and aquatic mammals the short wavelength cone has also been sacrificed. Thus most mammals are dichromatic or monochromatic (no colour vision) because they have either two or one class of cones.

An interesting ecological mystery is that several primates with polymorphic colour vision sport orange fur on at least part of their bodies (e.g golden lion tamarins and squirrel monkeys). We have measured the spectral reflectance of this fur and found that for dichromatic primates it would lie within the range of foliage colours. This is illustrated in the photographs above. In A) the chromaticities are directly reproduced from the original slide, whereas in B) the chromaticities have been calculated to represent the appearance to a dichromatic primate, according to the algorithm of Brettel et al. (1997). If the printing could reproduce the exact chromaticities, the two pictures would appear identical to a human deuteranope (who lacks M cones). The photograph is of Cercopithecus ascanius in Kibale forest, Uganda, and this species actually has uniform trichromacy, but its orange tail shares the spectral properties of fur of several polymorphic species. Why would a monkey evolve fur colour potentially conspicuous to predators, but cryptic to most members of its own species? To help solve this puzzle we will have to discover what colour vision is possessed by the monkeys’ predators, such as eagles.
different cones, creating three types of cone in total. This leads to the interesting situation in which, within each species, some females are trichromatic, but the other females and all males are dichromatic.

In the ancestor of African and Asian monkeys and apes (including humans), a gene duplication gave rise to two distinct long-wavelength pigment genes arranged in tandem on the X chromosome. All individuals could therefore possess three distinct types of cone: long-, middle- and short-wavelength, or L, M and S cones. Our L and M pigments thus belong to the same class, and are in fact very similar, differing from each other only by 15 amino acids, which shift the $\lambda_{\text{max}}$ from 560 nm to 530 nm. The tandem arrangement of two genes with 98% homology has been a fertile field for genetic mischief in humans. Unequal crossover at meiosis has produced many hybrid genes in the population, and the number of genes now in the array on our X chromosomes varies from one to more than five. This is why around 8% of men have ‘anomalous trichromacy’ or are dichromatic. In fact the polymorphism in our population is such that even amongst ‘normal’ trichromats, only about 60% of L pigments have the longest known $\lambda_{\text{max}}$, so there is really no such thing as ‘normal colour vision’. Whether the primates possess one or two types of

---

**Figure 1A**

The five families, or classes, of vertebrate visual pigments. A: characteristic sensitivity functions. B: simplified phylogeny and nomenclature. The nomenclature of the five classes has become confusing for three reasons:

1) the first animal to have its photopigments identified, namely humans, lacks two of the classes altogether, but has two types of the longwave family;
2) the range of sensitivities of the families overlap, so pigments from both M/LWS and RH2 classes have been called ‘green’, and pigments from SWS1, SWS2, RH1 and RH2 classes have been called ‘blue’;
3) there is not exclusive correspondence between each family and rod or cone receptor types: RH2 pigments, and even an M/LWS pigment, has been found in rods (e.g. geckos), and an RH1 pigment can be found in some cones of a chameleon.

It is therefore recommended that we dispense with the misleading colour names ‘red’, ‘green’ and ‘blue’, and follow the nomenclature used by Yokoyama and Ebrey and Koutalos as in Figure 1A.

---

**Figure 1B**

---

\[ \lambda_{\text{max}} \]
Fruits of *Ficus asperifolia*, a fig species known to be eaten by chimpanzees and red-tailed monkeys in Kibale Forest, Uganda. The fruits are much more conspicuous in the coloured than in the black and white version of the photograph, illustrating the importance of colour in a visual search task where lightness and form vary unpredictably.

**Ecology of colour vision**

When one asks why primate colour vision is the way it is, there are two distinct types of answer. One type deals with the details of how the visual system works: molecular genetics; cell biology; anatomy; physiology etc. Another type of answer concerns why primates with that kind of colour vision survived and bred successfully enough to be our ancestors, while primates with alternative forms of colour vision did not. Here, I offer the latter type of answer.

In asking why primates evolved trichromacy, there are actually two questions 1) Why have trichromacy at all? 2) What advantage does our form of trichromacy have over other possible forms? More specifically, 1) what was the selective advantage in having two distinct long-wave cones? And, 2) given that altering even a single amino acid can shift the $\lambda_{\text{max}}$ several nm$^{17}$ and a wide range of $\lambda_{\text{max}}$ are known for M/LWS cones in other mammals,$^{7}$ why do all fully trichromatic primate species have virtually identical L cones, with $\lambda_{\text{max}}$ at 560 nm, and virtually identical M cones, with $\lambda_{\text{max}}$ at 530 nm?

1. **Why do we have trichromacy?**

The hypothesis that primate trichromacy evolved for frugivory (fruit eating) is at least a century old$^{16,18,19}$. Many primates rely heavily on fruit, and those species that now eat a lot of leaves are believed to have had more frugivorous ancestors.$^{20}$ There are other useful cues beside colour, such as shape, position, smell, touch, or sound (alerting from birds for example), but this would not stop trichromacy being selected if it gave a primate a competitive edge. Spotting fruit in foliage is in fact one of the few natural tasks at which human dichromats are very impaired.$^{16,21}$ The difficulty is caused by the ‘visually busy background’ that has dappled patches of different luminance and varying shape (Figure 2). Laboratory visual search tasks$^{22}$ have reliably found that colour serves as an efficient cue for detecting a target amongst distracting objects. So the crucial advantage of trichromacy to the primate ancestors may have been an enhanced ability to detect fruit in the forest. An alternative, or additional advantage may have been better discrimination between ripe and unripe fruit (Figure 3).

A different hypothesis is that the advantage of trichromatic vision lay in the detection of red or reddish/brown colouration in the edible young leaves of some tropical plants.$^{23,24}$ Many primates rely on leaves to some extent, whether they eat them year-round or just when fruit is unavailable. They are selective about which leaves of particular plants they eat, and preference has been found to correlate with the protein/fibre ratio, which is generally higher in young leaves than in older leaves. In most species the young leaves are lighter than the mature leaves and slightly yellowish (owing to greater carotenoid/chlorophyll ratio). However, in a significant proportion of plants the young leaves in the tropics are reddish-brown. Dichromatic animals might confuse these young leaves with mature leaves because lightness cannot be used as a cue, but a trichromatic primate would be able to spot the young leaves without difficulty. Even for the yellowish young leaves, hue difference might be a more efficient means of detection than lightness difference, because the mature leaf background has dappled variation in luminance owing to patches of differing local illuminant, the angle of leaf blade and patches of sky showing through gaps. The folivory hypothesis appears to be supported by the fact the that a fairly folivorous genus of South American monkeys, howler monkeys, has become uniformly trichromatic by means of a gene duplication similar to that which occurred in the ancestor of Old World monkeys and apes.$^{25,26}$

How can we test whether primate trichromacy is significantly better than dichromacy for finding fruit in the forest, for discriminating fruit ripeness, or for finding edible young...
Photographs and reflectance spectra of four species of fruits commonly eaten by primates at Kibale Forest, Uganda. The unripe fruits appear on the left of the photographs and their reflectance spectra are drawn in green. The ripe fruits are on the right of the photographs and their reflectance spectra are drawn as solid black lines. The dashed lines are for mid-ripe fruits. Note that primate long-wave (L) cones have extremely low sensitivity beyond 700 nm, and therefore the sharp rise in many spectra at around 700 nm is of very little importance to extant primates.

Figure 3

A Chrysophyllum Albidum

B Ficus exasperata (Moraccae)

C Diospyros abyssinica (Ebenaceae)

D Uvariopsis congensis (Annonaceae)
leaves? The approach taken by Ben Regan, John Mollon and myself has been as follows:

1) We measured the spectral properties of fruit and leaves in the diets of various primates: howler monkeys, spider monkeys and capuchins in French Guiana, chimpanzees, colobus monkeys, guenons and mangabeys in Uganda. We needed to obtain ‘stimulus spectra’ that might reach a primate’s eye in its natural environment. This was done in two ways: a) by directly making spectral measurements in situ in a forest canopy; b) by measuring the reflectance spectra of collected fruit and leaves, and combining them with a measurement of the natural lighting conditions in the forest canopy.

2) We obtained also stimulus spectra, by the same two methods, for the mature leaves that form the natural background amongst which the fruit or young leaves must be found.

3) Each stimulus spectrum was filtered as it would be by the optical properties of a primate’s eye, before it reaches the retina. The important ocular filters are the lens and macular pigment.

4) The filtered stimulus spectra could then be multiplied by cone sensitivity functions and integrated over the visible range of wavelengths in order to calculate the quantum catch (S, M or L) in each class of cone.

5) We calculated chromaticity coordinates that represent the signals in the post-receptoral pathways: for dichromatic primates, S/L for the colour pathway and L for luminance; for trichromats, S/(L+M) and L+M for the equivalent colour and luminance pathways, and L/(L+M) for the new colour pathway. We could then assess whether the new pathway of trichromats offered an advantage, in terms of signal to noise ratios or signal spread, over the colour and luminance pathways of the possible forms of dichromacy.

The results indicated that the ability to compare signals from distinct L and M cones offered an advantage for finding most fruit amongst forest foliage, for discerning the ripeness of many of these fruit species, and also for finding the young leaves of many species (Figure 4). Therefore any or all of these important tasks might have been important influences in the evolution of trichromacy.

2. Why do we have our particular form of trichromacy?
If having trichromacy is helpful in many tasks, how might we distinguish between the fruit-finding, fruit-ripeness

---

Figure 4
Chromaticity and luminance of fruit and leaves calculated for a primate with L, M and S cones with λ values of 561, 531 and 430 nm. In the left hand diagrams, the two colour subsystems that together produce trichromacy are represented separately by the vertical and horizontal axes (the curve marks part of the locus of monochromatic lights). The corresponding right hand plots represent, for the same fruit or leaf spectra, luminance and the ancient chromaticity channel common to dichromats and trichromats.

A: Unripe (filled circles), mid-ripe (open squares) and ripe (filled triangles) fruits of Diospyros abyssinica (see Figure 3). As this fruit ripens its signal in the ancient colour channel (S/(L+M)) does not change much, but the signal increases dramatically in the new colour channel of trichromats (L/(L+M)), accompanied by a decrease in luminance (L+M). For nearly all fruit whose reflectance spectra change as they ripen, there is an increase in signal in the new colour channel, and for some fruit there is no change. The same is true for luminance. The trichromat therefore has the advantage of a generally larger signal in a consistent direction.

B: The top ten fruit species in the diet of chimpanzees (black squares), and the leaves amongst which these fruits must be found (grey circles). For a dichromat, the fruit colours are indistinguishable from the leaves (the patterns of S/L values for all known types of primate dichromacy are very similar to the pattern of S/(L+M) values shown). The luminance of many fruit is higher than that of most leaves, but the leaves were measured in situ in the forest (in cloudy conditions), whereas the fruits were collected. In the forest much fruit would lie in shadow, and thus have lower luminance, making them harder to detect. In the new colour channel of the trichromat, however, the fruit clearly stands out.

C: Young leaves eaten by black and white colobus (black squares), and the mature leaves amongst which they must be found (grey circles). Many young leaves have reflectance spectra that are a subset of the mature leaves, but for those that are different at all, the results are similar to those for the fruit.
and leaf-finding theories? This brings us to our second question: why do all fully trichromatic primate species have virtually identical L cones, with $\lambda_{\text{max}}$ at 560 nm, and virtually identical M cones, with $\lambda_{\text{max}}$ at 530 nm? In the absence of other constraints, we might expect the L/M channel to be optimized for the task that most influenced its evolution. Therefore we asked whether changing the spectral sensitivity of the M or L cones could provide an even greater advantage for any of the tasks: what vision would be optimal for the signals present in the forest?

For the task of discriminating fruit ripeness, we calculated which combination of all possible M and L cone spectral sensitivities would maximise the differences between the chromaticities of unripe and ripe fruit.29 We found that the optimal pair of cone pigments depends on the fruit species in question, but is never what we actually possess. For most fruit, moving the $\lambda_{\text{max}}$ of the L cone to wavelengths longer than 565 nm would be an advantage. Therefore we can conclude that although the ability to discriminate fruit ripeness may have been a factor in selecting trichromacy over dichromacy, this task has not determined the exact form of trichromacy primates possess.

For the tasks of finding fruit or young leaves, we calculated signal-to-noise ratios for detecting target food items amongst foliage for all possible combinations of M and L cone spectral sensitivities. Regan and Mollon27,28 performed this modelling first for the fruit diets of the three South American primates, and then we addressed separately the fruit and leaf diets of the six African primates studied in Uganda.29 We found that in all cases the maximum signal-to-noise ratios were yielded by a pair of pigments very close to what the primates actually possess: about 530 nm and 560 nm (Figure 5). The two sets of targets, edible leaves and fruit, produce virtually identical results because, we discovered, it is not the exact spectral properties of the targets themselves, but rather the properties of the mature foliage background that determine which spectral positions are optimal: the pair of $\lambda_{\text{max}}$ values that yield maximum signal-to-noise is determined chiefly by minimising the variance of the chromaticities of mature leaves (Figure 6).30 So although these results cannot differentiate between the folivory and frugivory hypotheses, they do, we believe, explain why all extant trichromatic primates have similar photopigment tuning despite varied diets.

**Conclusions**

In answer to the question what was the selective advantage in having two distinct long-wave cones, we can say that as soon as a primate has two cone classes with peaks between 500 nm and 600 nm it gains an advantage in finding fruit, discriminating fruit ripeness, and in finding young edible leaves.

In answer to the question why do all fully trichromatic primate species have virtually identical L cones, with $\lambda_{\text{max}}$ at 560 nm, and virtually identical M cones, with $\lambda_{\text{max}}$ at 530 nm, we can say that these L and M cone pigments are nearly optimal for detecting anything against a background of tropical forest leaves (i.e. leaf chromaticity noise is minimised).

In answer to whether primate colour vision has evolved for folivory or frugivory, we would have to say both may have been important drivers, but fruit-finding was probably more important in the past, with leaf-finding growing in importance as primates became larger and more folivorous. It is a much simpler genetic step to create the polymorphism than the viable gene duplication needed for uniform trichromacy. The polymorphic state in lemurs and South American monkeys may be maintained by ‘pure heterozygous advantage’,31 and it seems likely that the ancestors of humans and other Old World primates, and those of howler monkeys, were just lucky. We can only guess what differences it would have made to the evolution of humans and our culture if all men and a proportion of women were dichromats.

---

**Figure 5**

Results of the signal-to-noise analysis for detecting A) the fruit diet of chimpanzees, and B) the leaf diet of black and white colobus. The abscissa shows the wavelength of peak sensitivity ($\lambda_{\text{max}}$) of one putative cone photopigment (labelled ‘M’ for convenience) and the ordinate shows the $\lambda_{\text{max}}$ of the other putative pigment (‘L’) that subserves the colour channel, L/(L+M). Each pixel, corresponding to one pigment pair, shows the mean of the signal-to-noise ratios for each target item (fruit or young leaf) against the foliage background, plotted in a grey scale as a proportion of the maximum ratio produced by any pigment pair. The white area shows the $\lambda_{\text{max}}$ combinations that produced over 90% of that maximum, and the filled square marks the pigments actually possessed by the primates, which seem to be the optimal choice for any pair separated by 30 nm. As pigments become more spectrally separated, spatial vision suffers because of chromatic aberration, and increased noise in the luminance channels.
Figure 6 illustrates how the chromaticities of mature leaves change for different M and L pigments with \( \lambda_{\text{max}} \) separated by 30 nm. For pigments with \( \lambda_{\text{max}} \) values of 531 and 561 nm, the leaves form a vertical distribution in the chromaticity diagram, with very little spread of signal in the L/M colour channel. This tight distribution of leaf chromaticities would give targets maximum chance of standing out. However, if the pigment sensitivities are moved to shorter or longer wavelengths, the distribution tilts, and also broadens in the latter case, producing more variance in the L/M pathway, and making it harder to detect fruit, young leaves, or other target items.

References