Trichromatic Color Vision in Primates

Michael H. Rowe

Department of Biological Sciences and Quantitative Biology Institute, Ohio University, Athens, Ohio 45701

Trichromatic color vision is rare among mammals, occurring only in some primates. Recent work has elucidated the adaptive behavioral significance of trichromacy as well as its underlying genetic and neurophysiological mechanisms. These studies reveal a complex neural system whose design and operation apparently does not conform to rigid deterministic principles.

The images formed by vertebrate eyes consist largely of luminance fluctuations, which are caused by differences in reflectance among the various objects in the external environment. Retinal circuitry in all vertebrates creates an extremely efficient representation of these images by generating neural signals proportional to the luminance contrast at each point in the image, defined as the difference between the luminance at that point and the average luminance in the area around that point, normalized to the average luminance. This is a form of signal compression known as opponent processing, and it is accomplished via the center-surround organization of retinal ganglion cells (Fig. 1). Its efficiency is due to the fact that the subtraction of the surround signal from the center signal minimizes luminance values that are common to both signals, allowing the dynamic range of the neuron to be used to represent local variations in luminance.

For most diurnal vertebrates, the visible part of the electromagnetic spectrum includes wavelengths between ~400 and ~650 nm. Natural objects do not reflect this range of wavelengths in equal proportion, so the light at various points across an image can vary in spectral composition as well as intensity. Thus the spectral content of natural images is also a rich source of information about the natural environment. The spectral content of most terrestrial images can be resolved into two principal components, one representing chromatic contrast between the short and middle wavelength portions of the spectrum and one representing chromatic contrast between the middle and long wavelength portions of the spectrum. For an animal to extract and use this information, there are two basic requirements: *1*) the visible portion of the wavelength domain must be adequately sampled, and *2*) retinal circuitry must allow the spectral information to be efficiently represented and transmitted to the brain. Animals whose visual systems meet these requirements have what is commonly known as color vision.

Most mammals have two cone pigments, one maximally sensitive to wavelengths at the short end of the visible spectrum (blue) and one maximally sensitive to wavelengths near the middle of the spectrum (green/yellow). The former are generally referred to as S cone pigments, the latter as M cone pigments. These apparently diverged from a common ancestral pigment ~500 million years ago. The peak sensitivities of these pigments only differ by ~100 nm, but because of the limited extent of the visible spectrum, these two sampling points are adequate for extracting spectral information from most natural images. Retinal circuitry in these animals uses signals from the two cone types to create an opponent signal proportional to spectral contrast along an S-M or blue-yellow axis, in addition to the signal proportional to the luminance contrast in the image. Thus the optic nerve of most mammals includes a chromatic information channel in addition to a luminance channel.

Exceptions to dichromacy are rare. Many marine mammals and a few nocturnal rodents, carnivores, and primates have secondarily lost the S cone pigment and become monochromats (4, 11). Many diurnal primates, on the other hand, have acquired a third cone pigment, the L cone pigment, which is maximally sensitive to the longer visible wavelengths (red). Furthermore, retinal circuitry in these primates creates an L-M (red-green) information channel in their optic nerves in addition to the blue-yellow [S-M in dichromats; S-(M + L) in trichromats] and luminance channels found in other mammals.

This review will highlight recent progress in understanding three aspects of trichromatic color vision in primates: *1*) behavioral significance of trichromacy, *2*) evolution and genetics of photopigments, and *3*) retinal circuits that create the red-green and blue-yellow color channels in the optic nerve.

Behavioral significance of primate trichromacy

Most primates are frugivorous to some extent; when available, fruit can constitute as much as 90% of the diet of both old world (catarrhine) and new world (platyrrhine) primates (reviewed in Ref. 12). Leaves and other foliage typically are eaten when fruit is scarce. Fruits eaten by primates differ from those eaten by other animals in having yellow, orange, or red coloration, and this has led to the suggestion that L cone pigments and primate trichromacy evolved as an adaptation for detecting fruit against a background of green foliage. This hypothesis has received considerable support over the last decade. Although it was first advanced to account for the origin of trichromacy in old world primates, its generality has recently been extended to include one new world species, the howler monkey (12). Howler monkeys are the only species of new world primates in which a form of trichromacy similar to that seen in old world primates has so far been observed; it is uniformly present in all members of the species, and both the spectral sensitivity and genetic basis of the L and M cone pigments are similar to those of old world primates (see below). Regan et al. (12) examined the spectral reflectance of fruit

0886-1714/02 5.00

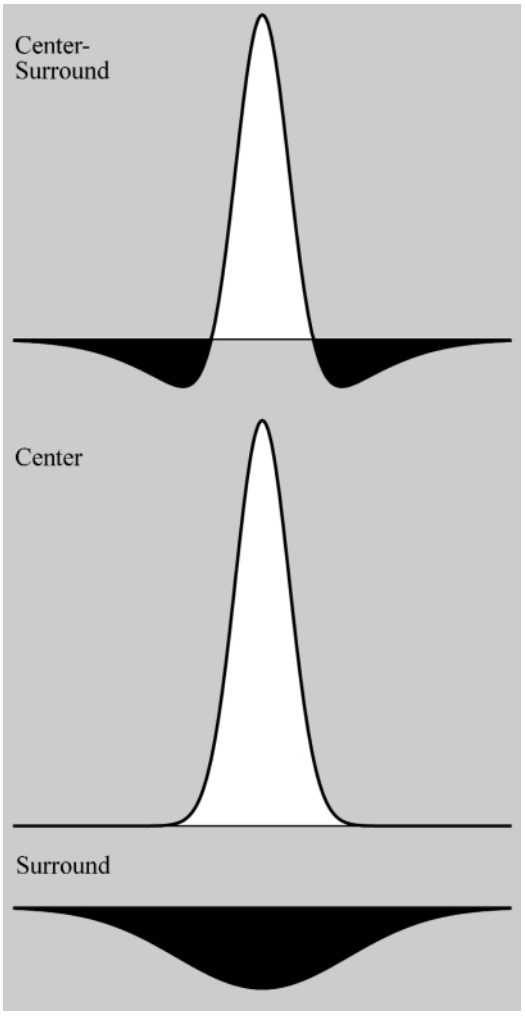

FIGURE 1. Luminance contrast signals in retinal ganglion cells are a consequence of the center-surround organization of their receptive fields. This organization is already present in the bipolar inputs to the ganglion cell. Center components represent summation of signals from a small number of cones that are directly connected to the bipolar cell. Surround components are opposite in polarity and reflect inhibitory signals originating from a much larger number of cones, extending over a larger area than the center component, and conveyed directly to the central cones via horizontal cells.

eaten by howler monkeys in their natural habitat and report that, as for old world primates, the spectral properties of their L and M cone pigments are well suited for detecting these fruits against a background of spectrally noisy green foliage.

An alternative hypothesis, that leaf consumption rather than fruit consumption is the principal factor in maintaining trichromacy, has also received support in a recent paper by Dominy and Lucas (3). They measured the spectral reflectance of the fruit and leaves that four old world species, with different specific dietary preferences, eat in their natural habitats. They then used simple calculations to assess the discriminability of preferred fruits and leaves in terms of their color. They report that the fruits eaten by all four species could be discriminated from background foliage not only along a red-green color axis but also along a blue-yellow color axis as well as a simple luminance axis. Thus, although trichromacy may provide some advantage in the form of an additional signal, it is not a requirement for discriminating preferred fruits from foliage. In contrast, they found that young leaves, which are slightly redder in color than mature leaves, could only be discriminated from mature leaves along a red-green color axis. The significance of this color difference among leaves is that younger leaves have a higher protein content and are less tough than mature leaves, making younger leaves a higher quality food item. Old world primates must rely on leaf foraging during periods when fruit is unavailable, and this is apparently also true for howler monkeys. In light of this, Dominy and Lucas argue that leaf foraging, rather than fruit foraging, is the critical factor in maintaining trichromacy in primates.

Only further tests can determine whether the leaf foraging hypothesis proves to be a refinement or a wholesale replacement of the fruit foraging hypothesis. In either case, the foraging behavior of trichromatic primates stands out as a rare example of a naturally occurring mammalian behavior for which some of the neural as well as genetic bases can be identified.

Evolution and genetics of photopigments

Recent work has greatly refined our understanding of the genetic basis of primate trichromacy (9, 12). All living primates appear to have an autosomal gene that produces an S cone pigment and an X-linked gene that produces a pigment with peak sensitivity at medium-long wavelengths (M/L pigment). [Nocturnal primates that lack an S cone pigment (see above) do so because the autosomal gene is not expressed.] Thus it is probable that the earliest primates were dichromatic, like most other mammals.

Old world primates. Sometime after the divergence of old and new world primate lineages, the M/L pigment gene was duplicated in old world primates, resulting in two copies of the gene on a single X chromosome. It is believed that at the time of this duplication, the two genes produced identical pigments, diverging subsequently to produce distinct M and L cone pigments. Except in humans, the resulting trichromacy in old world primates is considered to be uniform in the sense that the same three pigments are present in all members of a species.

One of the most remarkable findings to emerge from recent work is the degree to which human color vision is not uniform (10). Human L and M pigment genes are located next to each other on the X chromosome and are 98% identical, with the result that hybrid or chimeric genes are commonly created. These produce a variety of slightly different pigments. Amino acid dimorphisms that are capable of producing significant spectral shifts can occur at seven different positions along the >300-amino acid chain comprising the opsin component of the pigments (which the X-linked genes produce). The peak spectral sensitivities of L and M pigments are separated by ~30 nm, and most of this separation is due to amino acid substitutions at two of the seven positions (positions 277 and 285); substitutions at any of the other five positions produce smaller spectral variations within each class of pigment. These amino

acid dimorphisms in humans are the same as the amino acid shifts that underlie the polymorphic color vision of new world primates (see below).

Our view of the two forms of anomalous trichromacy, protanomaly (reduced sensitivity to long wavelengths) and deuteranomaly (reduced sensitivity to medium wavelengths), has also changed significantly. The classic view was that protanomalous and deuteranomalous individuals suffered from reduced or abnormal function of L or M pigments, respectively. New evidence suggests, however, that protanomalous trichromats are missing the L pigment gene and instead have two M pigment genes that produce pigments with amino acid substitutions that are not sufficient to support normal trichromacy. Similarly, deuteranomalous trichromats have two L pigment genes that produce slightly different pigments. These individuals are not missing the M pigment gene, however; it is present, but for reasons that are not understood, it is not expressed.

New world primates. Most new world primates lack a second X-linked pigment gene, but their single M/L gene is polymorphic, with as many as five different alleles in some species, producing pigments with different spectral sensitivities. Thus the potential for trichromacy exists in females of these species (with different alleles on each X chromosome), and behavioral tests in squirrel monkeys indicate that this potential is realized in about two thirds of the female population. Polymorphism of the M/L pigment gene has also recently been reported in a number of prosimians, raising the possibility that it may have been present in the ancestor to all primates.

The only known exception to this polymorphic form of trichromacy in new world primates (excluding the monochromatic species in which the S cone pigment is not expressed) occurs in howler monkeys. These animals have two X-linked pigment genes, producing distinct M and L cone pigments, which give them uniform trichromacy similar to that of old world primates. However, the duplication event that produced the extra gene occurred independently in this species and is thought to be more recent than the duplication that occurred in old world primates. Furthermore, unlike the situation in old world primates, this duplication is believed to have resulted in two different alleles of the polymorphic M/L gene being placed on each X chromosome, creating the immediate potential for uniform trichromacy in both males and females.

Retinal circuitry and color channels in the optic nerve

It has been recognized for over a century that the human visual system encodes colors in the form of red-green opponent and blue-yellow opponent signals, and it has been almost half a century since such color opponent signals were first recorded in the retina. Nevertheless, the neural basis of the red-green, blue-yellow, and luminance opponent channels identified psychophysically has remained elusive and controversial. Current accounts of primate retinal ganglion cell physiology divide them into three broad groups whose labels reflect their axonal destinations within the lateral geniculate nucleus (LGN): P cells terminate in parvocellular layers, M cells in magnocellular layers, and K cells in koniocellular zones intercalated between magnocellular and parvocellular layers. Recent data have helped to clarify the pattern of correspondence between these physiological groupings and the three psychophysical mechanisms.

Luminance channel. It is generally believed that M cells provide at least part of the neural basis for the luminance channel, since M cells do not show any evidence of spectral opponency in their receptive fields (13). However, the spatial resolution of human observers for achromatic stimuli is much higher than that of M cells, so it would appear that P cells, which do have sufficiently high achromatic spatial resolution, also contribute to the luminance channel.

Blue-yellow channel. S cones are morphologically and biochemically distinct, and they are sparsely distributed, comprising ~10% of the cone population. The circuitry that uses the S cone signal to create the blue-yellow opponent channel has been greatly clarified by recent experiments using an in vitro primate retina preparation (2). Ganglion cells with blue-on, yellow-off receptive fields are small, bistratified cells with dendritic arbors in both inner and outer sublaminae of the inner plexiform layer (IPL). These ganglion cells have axons that ramify in the koniocellular zones of the LGN and are therefore members of the physiological koniocellular channel (5). Their inner dendritic arbor receives synaptic input from a bipolar cell that is selectively connected to S cones (Fig. 2). The light responses of the S cone bipolars have not yet been recorded, but since it is well established that bipolar cells whose axons ramify in the inner sublamina of the IPL are depolarizing or

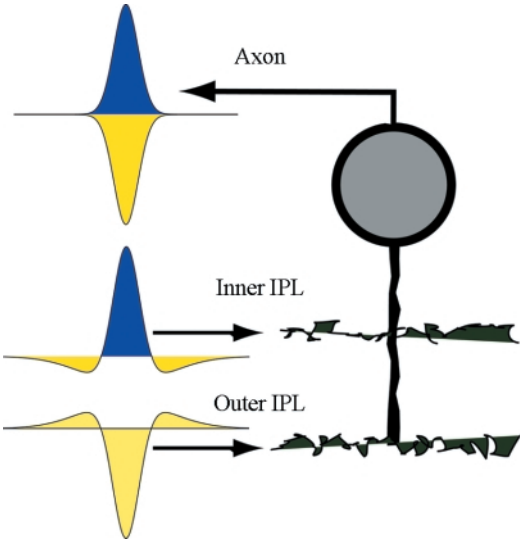

FIGURE 2. Bistratified ganglion cells have dendrites that stratify at both the inner and outer levels of the inner plexiform layer (IPL). Inner dendritic branches receive input from bipolar cells that are selectively connected to blue cones and have receptive fields with blue-on centers and nonselective-off surrounds. The outer dendrites receive input from bipolar cells that are nonselectively connected to red and green cones (in trichromats) and have receptive fields with yellow-off centers and yellow-on surrounds. Summation of these two bipolar inputs produces a ganglion cell receptive field with a spectrally opponent, blue-on, yellow-off center component. The surround components of the bipolar inputs largely cancel each other, so the ganglion cell receptive field lacks spatial antagonism.

"on" bipolars, the S cone bipolar is presumed to be the source of the blue-on response of the bistratified ganglion cell. It is also presumed that S cone bipolars resemble other bipolar cells in having center-surround receptive fields. The most likely source of the surround is a particular type of horizontal cell, designated H2, in which L, M, and S cone responses are equally weighted. Thus the signal conveyed to the bistratified ganglion cells by the S cone bipolar would be spectrally opponent. Furthermore, experiments using pharmacological agents known to block transmission from photoreceptors to depolarizing bipolars eliminate not only the blue-on response but also much of the yellow-off response of bistratified ganglion cells. Thus it appears that the S cone bipolar signal to the inner dendrite can account for most of the blue-yellow opponency observed in the response of the bistratified ganglion cell.

The part of the bistratified dendritic tree in the outer sublamina of the IPL receives input from diffuse bipolar cells that contact both M and L cones nonselectively. These bipolar cells are presumed to have a surround component to their receptive fields mediated by H1 horizontal cells that also contact L and M cones. Since bipolar cells whose axons ramify in the outer sublamina of the IPL are hyperpolarizing or "off" bipolars, the diffuse bipolar input is a likely source of at least part of the yellow-off (i.e., L + M) response of the bistratified ganglion cell.

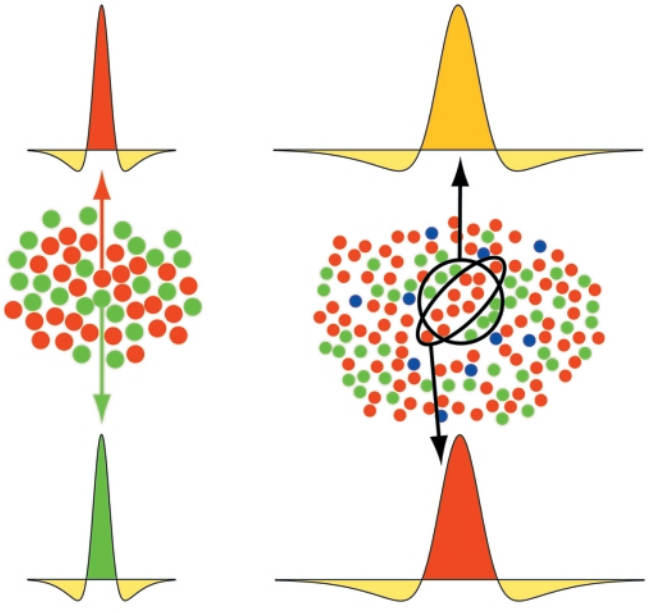

FIGURE 3. *Left*: midget ganglion cells (P cells) serving foveal vision are connected to single red (*top*) or green (*bottom*) cones via midget bipolars (represented by the red and green arrows). Blue cones are absent in the fovea. The surrounds of the bipolar cells are nonselectively connected to both red and green cones. Consequently, the ganglion cell receptive fields are automatically spectrally opponent with red or green centers and yellow surrounds. The two ganglion cells illustrated are both on-center, but off-center varieties also exist. *Right*: beyond perifoveal regions, P cells are connected to 15–30 cones via multiple midget bipolar cells. If these connections are random, the spectral composition of the center signal would be expected to be shifted toward yellow and the spectral opponency of the receptive field reduced (*top*). However, the combination of patchy red and green distributions and elliptical receptive field centers makes it possible for cone purity of the center signal, and thus the spectral opponency of the receptive field, to be retained (*bottom*).

The receptive field center of the diffuse bipolar could serve to create a push-pull signal that renders the bistratified ganglion cell more sensitive to blue-yellow differences in the image. The surrounds of the diffuse bipolar would be expected to largely cancel the surrounds of S cone bipolar inputs, thereby explaining the lack of center-surround spatial antagonism in the receptive fields of the bistratified cells (2). A similar lack of spatial antagonism has been a feature of blue-yellow opponent ganglion cells in primates since their earliest descriptions.

Ganglion cells with spectrally opponent, blue-off responses have also been reported, but they appear to be much more rare than blue-on cells and have not been studied with intracellular methods. However, in addition to the S cone bipolar cells described above, S cones also contact a subset of midget bipolars similar to those contacted by L and M cones (see below), so the blue-off pathway probably involves P cells, but the reasons for this difference in circuitry and for the numerical imbalance between the two pathways are not clear.

Red-green channel. L and M cones collectively form a fairly regular hexagonal sampling area in and around the fovea. When considered separately, however, L and M cones appear to be distributed randomly, with the result that small patches of retina can contain clumps of pure L or pure M cones. Furthermore, L cones generally outnumber M cones by about 2 to 1, although the ratio can vary by a factor of 4, even among individuals with normal color vision. These observations have significant consequences for understanding the circuits responsible for red-green opponency and challenge the assumption that a stereotypical pattern of connections between cones and postreceptoral neurons is required.

Abundant physiological and anatomic evidence implicates P cells as the cellular correlate of the red-green opponent channel, but the exact circuitry remains controversial. Much of the controversy revolves around the question of whether or not the opponency observed in P cells requires that the opponent mechanisms, which are usually equated with the center and surround of the receptive field, be selectively connected to either L or M cones. Early models proposed that center and surround were selectively connected to opponent cone types. Recent evidence challenges this idea for both center and surround and supports an alternative model, known as the random connection model, that does not require either the center or the surround to be selectively connected to cones of one type (2).

SURROUND SPECIFICITY. Extracellular recordings from P cells, or their target neurons in the LGN, have yielded inconsistent results; some authors report that surrounds are cone specific, others that they are not. However, recent intracellular analysis has shown that the horizontal cells that generate the surround responses of midget bipolars, and presumably of P cells, are nonselectively connected to L and M cones (2) but that the ratio of L:M input can vary considerably. Results also differ in regard to surround size; some results indicate that they are spatially coextensive with the center, others that they are larger than the center. One way to reconcile these discrepancies has been suggested by Martin (5), who argued that mechanisms such as shunting inhibition could cause the effective space constant of horizontal cells to be much smaller than the size of

their receptive field as measured at the soma. Given the patchy distribution of L and M cones, smaller space constants increase the likelihood that the signal from the horizontal cell will be dominated by one cone type. This proposal also accounts for the observation that surrounds that are cone specific are smaller and more likely to be spatially coextensive with the receptive field center.

CENTER SPECIFICITY. In regions of retina in and around the fovea, P cells are known to receive input from single midget bipolar cells, which are in turn connected to single L or M cones. Thus cone specificity of the center mechanism is automatic in central retina (Fig. 3, *left*). This combination of single cone input to the center and mixed cone input to the surround virtually guarantees that the center and surround of central P cells will be spectrally opponent.

At greater distances from the fovea, although midget bipolars continue to receive input from single cones, P ganglion cells receive input from increasing numbers of midget bipolars and thus increasing numbers of cones. If P cells are randomly connected to midget bipolars, and thus to L and M cones, the cone specificity of the receptive field center would be expected to decline with increasing distance from the fovea (Fig. 3, *top right*). Indeed, human chromatic sensitivity does decline significantly with increasing distance from the fovea, and Mullen and Kingdom (8) have shown that the slope of this decline closely matches the decline in the cone selectivity of P cell centers that is predicted by a random connection scheme. Nevertheless, there is some uncertainty as to whether or not the chromatic opponency exhibited by peripheral P cells is sufficient to sustain peripheral color vision, and some authors have suggested that perceptual red-green discriminations could be mediated instead by a mechanism similar to the blue-yellow circuitry described above (1). In this view, P cells are required only for providing high spatial resolution. So far, however, there is no evidence to support the existence of red-green opponent ganglion cells other than P cells.

On the other hand, a recent paper by Martin et al. (6) suggests that chromatic opponency in peripheral P cells may in fact be more than adequate to support perceptual performance. These authors report that, even as far as 10 mm from the fovea, P cells in macaque monkeys can be as sensitive to red-green contrast as their foveal counterparts. They provide an explanation, furthermore, that is entirely compatible with the random connection model. The receptive field centers of P cells in their sample were typically elliptical in shape, with average circular anisotropy of 1.8. This, in combination with the patchy distribution of L and M cones, creates a situation in which the shape of the receptive field could be matched to that of local patches of L or M cones, resulting in a relatively high degree of cone specificity (Fig. 3, *bottom right*). Using simulations based on actual L and M cone distributions in human and macaque retinas, they report that even small anisotropies can dramatically alter the ratio of L and M cone input to the receptive field center of P cells. They conclude that a high degree of chromatic opponency in peripheral P cells is not inconsistent with the random connection model and that there is no need to postulate additional mechanisms to support peripheral color vision. These results, however, do present a dilemma in that the sensitivity of peripheral P cells to red-green contrast is as much as 10 times higher than the sensitivity of human observers to the same stimuli presented at comparable locations in the visual field. Thus the degree to which the signals from chromatically opponent peripheral P cells are used by the brain remains to be determined.

Concluding remarks

The genetic and neurophysiological evidence reviewed above suggests that the construction and operation of even complex neural systems need not follow rigid, deterministic rules. Random and stochastic processes can also be involved and may actually be better suited for providing the natural variation that is a necessary substrate for the evolution of complex systems. Thus the random connection model of P cell chromatic (red-green) opponency is attractive not only because it is neurophysiologically and computationally plausible but also because it provides a scenario for understanding how functional trichromacy might have evolved (7). The appearance of the L cone pigment gene in old world primates apparently occurred after the divergence of the old world and new world lineages. In contrast, midget bipolar and ganglion cells are present in new world primates and prosimians as well as old world primates, and 1:1 convergence ratios from cones to midget ganglion cells are also present in new world primates (14). This clearly suggests that P cells evolved as a system for high spatial resolution before the old world/new world divergence and that the circuitry required for creating a red-green opponent signal in perifoveal P cells was probably already present when the L cone pigment first appeared. If cone-specific connections to receptive field center and surround were not required, as the random connection model suggests, then no additional specialization was needed to achieve functional trichromacy at the retinal level. Important remaining questions involve how the chromatic and spatial signals carried by P cells are separately decoded centrally and what additional evolutionary changes in central pathways were needed to make functional trichromacy possible. Parallel convergence into two populations of central neurons, in which spatial detail is preserved at the expense of chromatic specificity in one population and chromatic specificity at the expense of spatial detail in another, is one possible decoding mechanism. The fact that functional trichromacy is present in those female squirrel monkeys that happen to have both L and M cone pigments suggests either that the wiring of central pathways in these animals is already sufficient to support trichromatic processing or that the presence of the third pigment can induce

"...the construction and operation of even complex neural systems need not follow rigid, determanistic rules. Random and stochastic processes can also be invloved...."

developmental changes in central wiring that are appropriate for processing the additional signal. Given the recent pace of discovery, answers to such questions should soon be forthcoming.

Due to space limitations, many individual contributions to the literature could not be directly cited; citation preference was given to reviews and recent papers. Interested readers can obtain a more complete reference list for the work summarized here from the author: rowe@ohiou.edu.

References

1. Calkins DJ and Sterling P. Evidence that circuits for spatial and color vision segregate at the first retinal synapse. *Neuron* 24: 313–321, 1999.
2. Dacey DM. Parallel pathways for spectral coding in primate retina. *Ann Rev Neurosci* 23: 743–775, 2000.
3. Dominy NJ and Lucas PW. Ecological importance of trichromatic vision in primates. *Nature* 410: 363–366, 2001.
4. Jacobs GH, Neitz M, Deegan JF, and Neitz J. Trichromatic colour vision in New World monkeys. *Nature* 382: 156–158, 1996.
5. Martin PR. Colour processing in the primate retina: recent progress. *J Physiol* 513: 631–638, 1998.
6. Martin PR, Lee BB, White AJR, Solomon SG, and Ruttiger L. Chromatic sensitivity of ganglion cells in the peripheral primate retina. *Nature* 410: 933–936, 2001.
7. Mollon JD. "Tho' she kneel'd in that place where they grew...": the uses and origins of primate color vision. *J Exp Biol* 146: 21–38, 1989.
8. Mullen KT and Kingdom FAA. Losses in peripheral colour sensitivity predicted from "hit and miss" post-receptoral cone connections. *Vision Res* 36: 1995–2000, 1996.
9. Nathans J. The evolution and physiology of human color vision: insights from molecular genetic studies of visual pigments. *Neuron* 24: 299–312, 1999.
10. Neitz M and Neitz J. Molecular genetics of color vision and color vision defects. *Arch Ophthalmol* 118: 691–700, 2000.
11. Peichl L, Behrmann G, and Kroger RHH. For whales and seals the ocean is not blue: a visual pigment loss in marine mammals. *Eur J Neurosci* 13: 1520–1528, 2001
12. Regan BC, Jullio C, Simmen B, Vienot F, Charles-Dominique P, and Mollon JD. Fruits, foliage and the evolution of primate colour vision. *Phil Trans R Soc Lond B Biol Sci* 356: 229–283, 2001.
13. Shapley R. Visual sensitivity and parallel retinocortical channels. *Ann Rev Psychol* 41: 635–658, 1990.
14. Silveira LCL, Lee BB, Yamada ES, Kremers J, and Hunt DM. Post-receptoral mechanisms of colour vision in new world primates. *Vision Res* 38: 3329–3337, 1998.