Almost Reason Enough for Having Eyes

Jay Neitz, Joseph Carroll and Maureen Neitz

It is a commonly held misconception that dogs are color blind. The truth is that a dog’s ability to see colors is similar to that of most other mammals,1 although different from that of humans with normal vision. Cross-species comparisons of differences in color vision have helped us appreciate our own ability to see colors, as well as to understand the underlying mechanisms of color vision. Among humans with normal color vision, there are differences in color perception that can also help us understand the fundamental processes of color vision.2,3 Over the past decade, comparisons and contrasts made at the cellular and molecular level among different color vision systems have brought us to a new understanding of color vision, one that differs in many unexpected ways from the tenets of conventional wisdom. With this understanding comes a new appreciation of the fact that wavelength sensing, in the form of color discrimination, is a fundamental visual capacity that serves to greatly expand the eye’s information-gathering power. Color vision has played an important role in how the biological machinery for seeing has developed over the course of evolution.
Neural mechanisms for seeing color

The underlying machinery of a biological system for seeing color is composed of two stages: the first consists of light-sensitive receptors and the second consists of the neural components needed to process, partition, and encode information about wavelength that the photoreceptors collect. Our understanding of how visual information is encoded in humans has developed from studies of human perception coupled with research on the physiology and behavior of animals. The central idea is that, from the retinal image, neural elements extract the relative responses of neighboring receptors and process the information through a sequence of stages to ultimately encode the information using a system of “labeled lines.”

The concept can be illustrated in the case of the percepts of black and white that, at higher levels of visual processing, are carried by a pair of labeled lines. Consider the small cluster of photoreceptors that are illuminated by a bright star as we fixate on it: multiple neural output lines gather information from the illuminated receptors, as well as from their immediate neighbors. In the brain, one set of nerve fibers that carry output from this cluster of receptors constitute the labeled line for “whiteness;” this line is activated when the white light from the star stimulates the photoreceptors. Activation of this labeled line is responsible for the perception of white. Another line, the labeled line for “blackness,” carries output collected from the very same photoreceptors. However, this line derives from neural circuits with physiologically different components that invert the signal relative to the whiteness labeled line. Because the signal carried by the blackness line is sign-reversed, its activity is maximally silenced when the star illuminates the photoreceptors. Conversely, the neural line labeled “blackness” is activated when the contrast is reversed, for instance when we fixate on a printed black dot against the background of a white page. Since one of the two output lines is inverted relative to the other (although both access the same photoreceptors), the activities of the two lines are always opposed. The activity in these two parallel output lines is presumably responsible for the percepts not only of black and white, but also of all of the intermediate shades of gray. There is not general agreement about how the activities of black and white labeled lines are related to percepts of gray. However, one theory is that there is no labeled line for gray. Rather, “gray” is the percept that occurs when neither line is active. A series of sensations from light gray to white would correspond to increasing levels of activity, ranging from zero input to maximum activity, in the white labeled line; sensations from dark grays to black would correspond to graded levels of activity in the black labeled line.

Hue, which is determined by the wavelength content of colors, is the property that allows us to perceive colors as ranging from red through yellow, green and blue. In the human visual system, the second...
COLOR VISION

stage for encoding hue has two components: each is responsible for a pair of sensations, and just as in the case of the black-white system, the sensations in each pair are opposed to one another. One hue system is responsible for perceiving the hue-pair blue and yellow; the other is responsible for the hue-pair red and green. Each of these two systems is believed to share properties with the black-white system: each presumably draws from a common set of photoreceptors, but outputs the information through different neural components to constitute the different labeled lines.

The blue-yellow and red-green systems are remarkably different and separate. Together, the two systems extract information used for color vision from the three traditionally recognized classes of cone photoreceptors: short-, middle-, and long-wavelength-sensitive, abbreviated S, M, and L.

However, the two systems diverge at the level of the cone photoreceptor output: one system draws output from the S cones, comparing it to the L plus M cone responses; the other extracts wavelength information from a comparison of the relative L and M cone responses. The fact that only one system depends on S cones is significant because S cones differ from M and L cones both in physiology and in retinal distribution. We have also learned that the underlying neural circuits of the systems are quite different in design, operation, and vulnerabilities: the blue-yellow system is more likely to succumb to insults, such as toxic exposure, eye disease or a trauma such as retinal detachment; the red-green system, less susceptible to acquired disorders, is subject to an astonishing number of congenital defects that rarely occur in the blue-yellow system. A comparative approach has provided clues toward understanding these two systems from an evolutionary perspective. In fact, with the exception of a few differences driven by utility, in each case the differences between the systems can be explained by a surprisingly unique evolutionary history.

The blue-yellow color vision system

Among mammals, trichromatic color vision is found only in humans and in a subset of the primates. Some mammals, probably a minority, are monochromats that lack color vision. The common form of color vision among mammals is dichromacy, like that found in dogs, in which only one of the two color vision systems—the one homologous to the human “blue-yellow” system—is present. In mammals, this system makes use of two different photoreceptor types: one is related to the human S cone, the other to the human M and L cones. The photoreceptor basis of wavelength discrimination mediated by this system is illustrated in Fig. 1. Over a wide band of the spectrum, each wavelength of light produces a unique ratio of activation in the two different photoreceptors. The neural circuits responsible for color vision compute the ratio of absorption of the two types of cone to provide a continuum of wavelength information through the region where the absorption spectra of the pigments overlap. Figure 1 represents an attempt to illustrate the color world of creatures with dichromatic color vision. Wavelengths in which the relative absorption by the S cones is highest are shown as appearing blue to the animal (by analogy to the human blue-yellow system) and wavelengths in which the relative absorption by the other cone type (L in the dog) is highest are shown as appearing yellow. From the middle of the spectrum to the ends, changes in color are seen solely as changes in saturation (i.e., vividness of hue, or degree of difference from gray). In the middle there is a wavelength for which the outputs of the two cones are balanced and the appearance is colorless (shown as gray in Fig.1), which corresponds to minimum saturation of both cone types. Thus, over an interval of about 100 nm, dichromats can discriminate among 20 to 30 gradations of monochromatic light. Presumably, the neural circuit that compares the responses of the S and L cones to produce an output with the sign “plus S/minus L,” constitutes the neural substrate for a labeled line for “blueness;” one in which short-wavelength (blue) light produces positive activity. The same inputs fed through a sign-reversing circuit would constitute the line for “yellowness,” which is silenced by blue and activated by light at the other end of the spectrum. In physiological studies, the neural substrate for the “blueness” line has been well characterized in the retina, but interestingly, the sign-reversed “yellowness” line has proven more elusive.

Photopigments and their genes

Vision is initiated by the absorption of light quanta by visual pigments within the photoreceptors. These photopigments are composed of a chromophore (11-cis-retinal) encased by, and covalently bound to, a protein component (opsin). In terrestrial animals, the chromophore is the same for
all pigments but the opsins vary, providing different surroundings for the chromophore that tune the absorption maxima of the photopigment to distinct spectral positions. All opsins are believed to have evolved from a single common ancestor;11 they belong to an enormous family of receptor molecules that may presumably be traceable in evolutionary origin to a single progenitor.

Over the last 15 years, molecular genetic methods have been used to deduce the amino acid sequences of the photopigment opsin from humans and a wide variety of animals.12-15 The deduced amino acid sequences are strikingly different for the two pigment classes that underlie dichromatic color vision: over 50% of the amino acids that make up each pigment are different. The two pigments are nearly identical in their purpose except that in order to provide the basis for color vision, they have to be tuned to different regions of the spectrum. Producing a spectral separation requires a relatively small number of amino acid differences. We assume that the high degree of difference is an indication that, in evolutionary terms, the S and L photopigments are not closely related.

The evolution of color vision

It is believed that humans evolved from an ancestor with only one type of photoreceptor.16 As discussed above, color vision systems work by comparing the relative number of photons absorbed by different photopigments tuned to distinct regions of the spectrum. The three classes of modern human cone pigments are thought to have evolved from the processes of gene duplication and divergence.17 Duplicated copies of an original cone pigment gene would have been free to mutate separately and diversify their absorption spectra, providing the basis for color vision. Presumably it is the selective advantage provided by color vision that has served to maintain the separate S and L pigment genes in the genome. Not all the amino acid positions involved in spectral tuning between S and L pigments have been identified; however, it is known that only seven amino acid changes are required to make the 30 nm difference between M and L photopigments, and that just two changes account for most of that shift. Extrapolating from this finding, we can predict that approximately 6% difference in amino acid sequence would be required to make the 100 nm shift between S and L. We attribute the remaining 48% amino acid difference to the effects of the genetic drift that has occurred over the eons since the original duplication event. This degree of difference can be used to estimate the amount of time that has elapsed since our ancient ancestors evolved two pigments.

If molecular divergence of the pigments is our correlate of evolutionary time, then an estimate of how long ago the two pigments diverged could be made if we had a calibration for the molecular clock. One available data set that can be used is the sequences of the photopigments contained in the rod photoreceptors responsible for low light vision. These have been extensively studied in a wide variety of species,15 and presumably rod and cone pigments would be subject to the same evolutionary constraints.

In Fig. 2, the percentage identity between human rhodopsin and rhodopsins from a variety of different species is plotted against estimates of the time elapsed since each animal and humans shared a common ancestor.18-21 The relationship is relatively linear (Fig. 2) as would be expected if the differences were indeed due to a genetic drift that had proceeded at a constant rate. The best fitting regression line shows a change of about 4% per 100 million years. Extrapolation using this scale places the time elapsed since divergence between the S and L/M cone pigments at more than a billion years. Allowing reasonable confidence limits for the extrapolation would place the likely time of divergence somewhere in the interval between 800 and 1100 million years ago (MYA). The first appearance of the photoreceptive structures that were the precursors to the earliest eyes probably appeared then.

Even on the vast scale of evolution, 800-1100 million years is a very long period of time. Mammals first appeared about 220 MYA; the earliest land animals crawled out of the sea 370 MYA; the earliest vertebrates appeared about 495 MYA. The fossil record doesn’t go back much further than 600 million years, but some of the earliest fossils do show the presence of eyes.16 Although we don’t have direct evidence predating the fossil record, it is presumed that eyes or light-sensitive precursors to eyes have been present throughout the history of the existence of free-moving organisms that lived in environments with sufficient available light.22 The important point is that it appears that the divergence of the S and L photopigments must have been very nearly coincident with the appearance of the first eyes. It is logical to assume that a photosensing organ with only one receptor type must have emerged first, but the implication from examination of the molecular data is that duplication and divergence of spectrally separate photopigments occurred virtually immediately after emergence of a photosensing organ. Adding the second pigment confers the advantage of an extended spectral range, even without the neural circuitry for extracting wavelength information. However, we argue here that the advantages to having color vision are great and there would be strong selective pressure favoring evolution of those circuits posthaste. Thus, color vision, at least in some crude form, may be as old as vision itself.

The importance of color vision

There is a historical tendency to focus on the visual system’s function as light detec-
COLOR VISION

Figure 3. The geometric expansion of visual capacity that accompanies each additional spectrally different photopigment. Creatures with only one type of photopigment are color blind: their visual world is restricted to a number of distinguishable steps of gray on the order of $10^2$ (left panel). Adding a second spectral type of cone photoreceptor and the appropriate neural connections adds another dimension to vision, expanding the number of intensity and wavelength combinations that can be discriminated to about 10,000 (middle panel). Adding a third photopigment makes trichromatic color vision possible and expands the number of different colors to over one million (right panel).

The information contained in light from the world around us is encrypted in the infinite combinations of spectral content and intensity. How much information can be extracted is determined by the ability of the visual system to discriminate among the different combinations. At any given adaptation level, the human eye can discriminate nearly 200 different levels of gray (Fig. 3, Panel 1), which is small considering the infinite number of possible information-carrying combinations of wavelength and intensity. The presence of color vision tremendously expands the amount of extractable information. Consider the difference when just one photopigment is added, such as when we compare monochromatic color vision to dichromatic color vision. Starting with one photopigment and adding a second spectral type does not just add a color, it adds an entire dimension of vision. For the dichromat, changes in wavelength produce up to 50 discernible steps. However, in the visual system, since black-white and blue-yellow are carried in separate parallel systems, for each gray level there will be nearly 50 different possibilities on the blue-yellow scale. Thus, adding the second pigment and the appropriate processing machinery increases the discriminable combinations geometrically from near 200 to near 10,000 steps (Fig. 3, Panel 2). It is little wonder that there would have been a strong evolutionary push to adopt a multiple pigment system even in the earliest organisms. We have to appreciate that wavelength sensing is as fundamental to vision as is light detection.

The red-green color vision system

The estimated time of divergence for the S and L/M opsins suggests that a biological apparatus capable of wavelength discrimination may have been among the earliest sensory systems to emerge. The human blue-yellow color vision system has been called ancient. Here, it is argued that its roots may be traceable to a wavelength sensor in very early organisms dating back a billion years. Interestingly, the same kind of analysis indicates that the red-green system is as new as the blue-yellow system is old. In humans, the L and M photopigments are individually polymorphic but on average they differ by about 15 amino acids. Using the same calibration as for the S vs. L/M pigments, and subtracting seven amino acids known to be involved in the spectral difference between L and M, leaves eight that might be attributable to genetic drift ($8/364 = 2\%$). At 4% per 100 million years, the divergence of the L and M genes would be estimated to have been about 50 MYA. We don’t want to place too much confidence in the accuracy of this estimate; however, it is consistent with recent calculations that place the split between New and Old World primates at about 60 MYA. The divergence of the L and M pigments can be argued to have occurred after that split since New World primates usually have only one gene encoding a photopigment in the middle-to-long wavelength range, while our near relatives, the Old World monkeys, have both L and M genes.

The addition of the third cone pigment gene was a required step in achieving a functional red-green color vision system. From the standpoint of being able to extract the information encoded in the wavelength content of light, the addition of another pair of neuronal lines in parallel with the black-white and blue-yellow lines represents an enormous gain. Recall that since the lines are added in parallel, the addition of each pair expands the number of discriminable wavelength combinations geometrically. Humans can distinguish close to 100 steps of spectral change contributed by the activity of the redness and greenness labeled lines. Multiply that times the approximately 10,000 colors that can be distinguished...
using the combination of the other systems, and the addition of the red-green system boosts the number of “colors” we can see to upwards of one million (Fig. 3, Panel 3).26

The geometric expansion of visual capacity that accompanies each added cone receptor type is of inestimable significance. To fully appreciate it, one need look no further than a comparison of normal and defective color vision in humans. Congenital red-green color vision defects are extremely common, affecting about 8% of males and 0.4% of females in the United States.27 About 25% of the people with red-green color vision defects are dichromats; most of them suffer from what is essentially a reversal of the gene duplication that was originally responsible for allowing primate trichromacy. A gene deletion has left the human dichromat with a single photopigment gene on the X-chromosome. Human dichromats are commonly referred to as “color blind,” although their blue-yellow system is intact. They enjoy color vision similar to that of many of our mammalian relatives, but far more limited than that of trichromats: compared to the trichromat the dichromat’s palette is estimated to be “missing” nearly one million colors. As would be expected, this loss disadvantages the dichromat in many everyday tasks; selected examples are illustrated in Fig. 4.

**Beyond trichromacy**

The ability to distinguish two million colors may seem impressive, but it is still small compared to the infinite combinations of wavelength and intensity in our environment. By extrapolation from the dramatic expansion in discrimination capacity accompanying the evolution from dichromacy to trichromacy, we can guess that the geometrical increase in distinguishable combinations would also be impressive if an organism had four cone pigments wired in a system to provide a fourth dimension of color vision. Perhaps to a tetrachromat, a mere trichromat would seem as color blind as a dichromat does to a trichromat. It has become increasingly apparent over the last 20 years that most non-mammal diurnal vertebrates, such as birds and fish, have four cone pigments. They have an added pigment that is sensitive to ultraviolet light. Some species have been demonstrated to achieve tetrachromatic color vision.28 The explanation usually given for why other vertebrates have more advanced color vision than mammals is that the earliest mammals appeared at the height of the dinosaurs’ dominion of the earth; the mammals found a niche available in being nocturnal, and they developed a retina that was highly adapted for vision under very low light levels.23 In optimizing the eye for night vision, two of the four photopigment genes found among the other vertebrate lines were apparently lost in mammals. Modern birds, reptiles, and fish enjoy color vision that is rooted in a multi-cone color vision system that predates the divergence of those vertebrate classes. The nocturnal ancestors of modern primates were reduced to dichromacy, and the blue-yellow system is the only color vision machinery they share with other vertebrates. Primates achieved trichromacy by inventing it separately.

**Unique processes create the neural circuits for red-green color vision**

An obvious question at this point is why trichromatic color vision has evolved only in the primates when other mammals have adopted diurnal lifestyles. If added dimensions of color vision confer such a tremendous advantage, why wouldn’t all diurnal mammals have evolved several higher dimensions of color vision? Diurnal mammals probably emerged within about the last 100-150 million years. This is a very short period of time in which to evolve a higher dimension of color vision, particularly if many statistically unlikely mutational changes are required to develop the neural circuits required to process and partition color information.

Primates have unique retinal features, evolved for different purposes, that may have afforded a shortcut unavailable to other mammals for the evolution of color vision.29 Diurnal primates have highly acute spatial vision. The high spatial resolution is served by a retinal pathway that makes use of neurons with unusually small receptive fields, called “midgets,” each of which directly contacts a single cone. Through connec--
tions that presumably evolved to amplify spatial contrast, a signal of opposite sign is collected from the cone’s surrounding neighbors. If a new spectral class of cone were added to the retina, the output collected by the midget system would automatically be one that compared responses from spectrally different cone types, exactly as is necessary for color vision. If the newly added cone type were randomly interspersed in the photoreceptor mosaic, any given cone of the new type would have some probability of having all of its neighbors be of the opposing type. Thus, for example, if the central cone were M and the surrounding cones L, a circuit for red-green color vision would be created serendipitously, without a delay stemming from the evolution of genetically specific neurons to wire the appropriate connections in the retina. Even in cases in which the surrounding cones were mixed in type, the averaged signal from the surround would be different enough from the central cone to provide a usable chromatic signal. In contrast, in mammals other than primates most neurons responsible for spatial contrast connect to a larger number of cones. Random connections that collect from a large number of cones would produce very similar average responses for both center and surround, and spectrally opponent cells would not be likely to arise by chance.

In primates, it seems plausible that evolution exploited existing circuits for spatial analysis to create the first stage of color processing that occurs in the retina. But how is the partitioning maintained and the processing extended at higher visual centers in the cortex? One possibility is that once again the newly evolved primate red-green color vision system may have taken advantage of pre-existing properties of the nervous system in order to avoid the necessity of creating genetically specified circuits that would exclusively recognize and connect to outputs carrying specific chromatic information. The circuits of the mammalian visual cortex are known to be molded by visual experience. This neural plasticity has been best studied in its role in the development of circuits responsible for coordinating input from the two eyes. The principle is that several synaptic inputs might initially converge on a single neuron in the cortex, but that over time the synapses are strengthened for those inputs that are concurrently active, while those inputs that are not well correlated with the consensus of activity may be selectively weakened or lost. It is possible that the nervous system could use a similar preexisting capacity for “neural learning” to extract information from our chromatic experience to mold the higher cortical circuits for red-green color vision.

In the future…

It may be possible to use gene therapy to replace missing photopigments in the eyes of color blind humans.

Directions in color vision research

Although vision scientists have yet to produce definitive evidence for the wiring of color vision being done by the opportunistic use of chance connection and neural learning, several very exciting lines of research related to these possibilities are underway. These experiments take advantage of naturally occurring large variations in the genes for the human L and M pigments.

Evolution is opportunistic. The gene duplication that eventually gave rise to the human L and M photopigment genes placed the duplicated gene adjacent to the original, producing a highly unstable genetic arrangement. If the evolutionary process had been more long-sighted, this probably would not have been the chosen arrangement. The configuration allows the L and M genes to misalign during meiosis and recombine, consequently intermixing their sequences. This has formed chimeric genes, which can produce photopigments with absorption spectra that are intermediate between those of the original L and M genes. In the normal population, variants are particularly common for the L gene. Thus females, who have two X chromosomes, can have genes encoding spectrally different L pigments on each X-chromosome. The process of X-chromosome inactivation insures that the two L pigment variants are expressed in separate populations of cones, making it certain that some females do indeed have four spectrally different cone types and the photoreceptor basis for tetrachromatic color vision. If the mechanisms described above are responsible for creating the neural circuits for a third dimension of primate color vision, then theoretically they should work the same way to provide a fourth dimension of color vision for women heterozygous for the L pigment gene. Is it true that some women have super color vision? From the description of the geometric increase in capacity that comes from an additional pigment one might think that if such women were among us, we would surely know about it. However, there are reasons this might not be the case. One is that the most common variants of the L pigment are only separated by a few nanometers; color vision based on such small spectral differences would produce a limited number of additional distinguishable steps. Alternatively, it is possible that such small differences may not be enough to drive cortical neural learning mechanisms when they have to compete with the existing well-separated L vs. M inputs. There is a chance that in some women, the two L pigment variants will be separated by about 10-12 nm. These women are the best candidates for having true tetrachromatic color vision, and technology is now available that would allow them to be identified by direct examination of their photopigment genes. There are some women who claim to have a superior color sense, and to be able to make color distinctions lost to other people. It would be an easy matter to examine the genes of these women to see if they encode spectrally well-separated L pigments. Jordan and Mollon, in a search for an extra dimension of color vision among female carriers of color vision defects, a group that is expected to have a high frequency of tetrachromats, found one woman in a sample of 31 who behaved as would be predicted if she were a tetrachromat. It would be interesting to know if her genes predicted the presence of pigment variants that are particularly well separated spectrally.

There is amazing variation in the ratio of L to M cones in the human retina. This was demonstrated recently in a dramatic fashion when adaptive optics was used to obtain the first images of the cone mosaic in the living human eye. One might predict that these differences in reti-
nal architecture would lead to dramatic differences in color vision. However, two people who had very different L/M cone ratios (one was 1:1 and the other 4:1) did not have measurably different color vision for parameters predicted to be most affected by the ratio difference. One explanation offered was that plasticity has allowed the nervous system to use information gathered from experience to make compensating changes for the differences in cone ratio. This lends credence to the idea that neural plasticity is responsible for setting up the cortical connections for color vision. In some of the best-studied systems, neural plasticity disappears after a critical period during early childhood. However, in recent years there has been growing evidence for plasticity of the adult cortex. If the circuits responsible for red-green color vision are plastic, is it possible that they could be among those that remain plastic throughout life? Recently, Yamauchi et al. presented evidence in favor of that theory. They reported long-term changes in color vision in subjects who had chromatically altered their visual environment for relatively short periods of time.

The future of technology and color vision

The development of color monitors for computers has recapitulated the evolution of human color vision. Monitors were originally monochrome. Although they did not evolve through a two phosphor stage that would have been the exact equivalent of dichromacy, they did evolve through stages in which the number of different colors was very limited. Today, the most highly evolved monitors, capable of simultaneously producing millions of colors, are matched with human color vision capacity. Evolution can occur quickly when a change is highly favored. In our survey of the local computer monitor population, all are of the millions-of-colors type. We might infer from the stunningly rapid change in the monitor population that in the modern world, a computer monitor’s ability to produce millions of colors has an extremely high adaptive significance. It is possible that in the future, technology may be available to do for the eye what has been done for color monitors. If the neural circuits for color vision are sufficiently plastic, it may be possible to use gene therapy to replace missing photopigments in the eyes of color blind humans. Furthermore, if the neural circuits can handle even higher dimensions of color vision that could come from artificially adding a fourth cone type, it is possible that gene therapy could also be used to extend normal human color vision. From witnessing how strongly people are driven to have a monitor that can output the highest amount of color information, we expect that if there were not associated risks, a therapy for color blindness would be widely adopted. Would trichromats have their vision expanded to tetrachromacy if a safe procedure were readily available?

Acknowledgment

This work was supported by NIH grants EY09303, EY09620 and EY01921, and Research to Prevent Blindness. We would like to thank P. M. Summerfelt for her invaluable help in preparing the manuscript and figures.

References


Jay Neitz, Joseph Carroll and Maureen Neitz work in the Department of Ophthalmology and in the Department of Cell Biology, Neurobiology and Anatomy at the Medical College of Wisconsin in Milwaukee, Wisconsin. Jay Neitz’s e-mail...