The Shape of Color: Retinal Cones and Spectral Dispersion

John A. Medeiros†, COLSA Corporation, Huntsville, AL & Nancy E. Medeiros, Retina Specialists of North Alabama, Huntsville, AL

Abstract

Why are the retinal color receptors cone-shaped? This is not a trivial question: the cone shape is evidently a universal feature of the color receptors while the achromatic rod receptors are always rod-shaped. What might be behind this dichotomy has not previously been explored in any meaningful way. We suggest here that the cone shape is not an incidental feature, but actually integral to cone function. We describe a waveguide mode cut-off effect that can physically separate light into its spectral colors along the length of a small, tapered waveguide, i.e., a cone. This effect converts the spectrum of incident light into position-dependent information along a tapered fiber; long-wavelength red light is excluded from the cone first, and the shortest-wavelength, blue light last. The retinal cone optical dimensions are apparently ideally tuned to exhibit this spectral dispersion. Converting this length-correlated information into a readable color code can be accomplished through translation into a time-correlated code through the microsaccadic movements of the eye to synchronize the time-delayed electrical signal generated by photoabsorption events from different positions along the receptor. Such a time code explains the existence of subjective color effects such as that induced by Benham’s Top. Detecting color information through such a waveguide effect also explains the Stiles-Crawford Effect of the Second Kind (SC II) whereby the apparent color of monochromatic light depends on its angle of incidence at the retina. The long puzzling similarity of violet and purple is directly explained by excitation of second-order waveguide mode propagation for short-wavelength light in this model. This dynamic model of color vision also accounts for the breakdown of statically-established metameric color matches under dynamic presentation, a feature of color vision that contradicts the basic assumptions of the standard Young-Helmholtz trichromatic model. Further underscoring the utility of the our proposed model is that it explains the major features of common forms of color blindness as a consequence of “mistuning” of the cone parameters.

† Corresponding Author: john.medeiros@mac.com, 256-426-2797
Introduction

We ask what is arguably one of the simplest questions one can ask about the photoreceptors providing color vision. Namely, why are they shaped the way they are?

Cone Shape

Consider that the human retina, and indeed the retinas of virtually all vertebrates, has a dual system of photosensitive elements. This duality, know as the Duplex Theory of Vision and first articulated by Schultze (1866), describes a retina with two separate sets of photoreceptors; rods that mediate monochromatic vision at low light levels and cones that mediate color vision at higher light levels (Weale, 1961). The disparate nature of the two receptor types is well-illustrated in Shultze’s original drawing, shown as Figure 1. The rod and cone receptors are so-named because of the shape of their photosensitive outer segments; the rod outer segments (ROS) being rod-shaped and the cone outer segments (COS) being cone-shaped. This dichotomy is apparently absolute; so this begs the question of why do the cones in fact have a cone shape?

This is not a trivial issue. There are numerous examples throughout the biological world, from the molecular to the macroscopic, where function follows form and the shape of a molecule or a biological element directly supports its biological function. Yet, no such fundamental role for its shape has ever been proposed for the cones. However, some ancillary roles for the cone shape have been proposed. For example, Miller and Snyder (1973) proposed a secondary function for the taper in peripheral cones, suggesting that the cone shape could serve to radiate away any light not absorbed within them outwards towards the surrounding rods in order to enhance the sensitivity of rod vision.

More recently, Harosi and Flamarique (2012) suggested a set of three possible functional roles for the cone taper: (1) “that outer...
segment taper serves to compensate for self-screening of the visual pigment contained within”, (2) a proposed linkage of “outer segment taper to compensate for a signal-to-noise ratio decline along the longitudinal direction”, and (3) a proposed relationship of “outer segment taper to the optical properties of the inner compartment whereby the primary determinant is the inner segments’s ability to concentrate light via its ellipsoid”. They concluded that real data (a study of geometrical optics combined with spectrophotometric and morphological data) supports a role for both of the first two functions, although they noted that “real cones taper more than required for these compensatory roles”. In support of the third functional role, they noted that “the rod/cone ratios of primarily diurnal animals are predicted based on a principle of equal light flux gathering between photoreceptors”. They further support this third role by noting that the “ellipsoid concentration factor, a measure of the ellipsoid ability to concentrate light onto the outer segment, correlates positively with outer segment taper expressed as a ratio of characteristic lengths”. Excepting these suggestions for relatively minor roles, the cone shape has hitherto been all but totally ignored by researchers, despite being a salient, defining physical feature of the color receptors.

Further underscoring the evident importance of cone shape is the fact that the color receptors across all species of vertebrates - mammals, reptiles, amphibians, birds, and fishes - are cones (Crescitelli, 1972). The cone shape is even seen in the photoreceptor elements of invertebrates, as in bees, for example. In vertebrates the relative number of cones and rods varies enormously across the gamut of species, depending apparently on their habits and habitats, but the color receptors are always conical (Detwiler, 1940).

Rod and cone outer segments are known to undergo distinctly different renewal processes (Young & Droz, 1968; Young, 1969, 1971; Anderson, Fisher & Steinberg, 1978; Steinberg, Fisher & Anderson, 1980; Eckmiller, 1987). Rods incorporate new material into their outer segments in a fashion whereby new discs with their photosensitive rhodopsin are formed at the cilia joining with the inner segment and progressively move as refreshed, intact discs down the length of the outer segment until they are shed and phagocytized by the pigment epithelium at their apical ends. Cones, on the other hand, do not undergo such a systematic process; indeed, they could not and still maintain their conical shape. Instead, new material is incorporated diffusely along the entire length of the infolding, continuous lamina that forms their outer segment and contains the photosensitive cone pigment(s). Phagocytosis does apparently occur at the cone distal ends as lamella material is shed, but the cone shape remains robustly intact (Hogan, Wood & Steinberg, 1974; Eckmiller, 1987). Thus entirely different regenerative processes exist for rods and cones that apparently allows the color receptors to maintain their conical shape.

There is more to the issue of cone shape. Consider that in the human retina (and the primate retina in general), the cone shape varies systematically across the retina in a distinctive way. Referring specifically to the photosensitive outer segments of the cones, the structure varies from being long, slender, and barely tapered near the visual
center at the fovea, to being progressively shorter, squatter, and more obviously cone-shaped towards the far periphery. Perhaps the best depiction of this systematic variation is still that of the hand-drawn images of von Greeff (1900), shown here as Figure 2, where all the cones are displayed at the same scale. It is of further interest, that the shape and morphology of the cones in any one local area of the retina are all the same; that is, cone shape changes do not occur in a punctate fashion, but, rather, progressively across the entire retina.

The functioning provided by the cones varies systematically across the retina as well. Highest resolution color vision function occurs in the foveal center where the cones are long and barely tapered, looking (as has long been noted) rather rod-like in appearance. In this central area, hue discrimination is at its best and color vision is effectively three-dimensional, or trichromatic, as specified by the number of independent parameters needed to match lights of different spectral compositions (metameric matches).

Moving away from the central retina, color vision function systematically degrades, passing first through dichromatic color discrimination, to become near-monochromatic in the far periphery (Baird, 1905; Lythgoe, 1931; Moreland & Cruz, 1959; Moreland, 1972; Wooten & Wald, 1973; Abramov & Gordon, 1977; Gordon & Abramov, 1977; Mullen, Sakurai & Chu, 2005; Hansen, Pracehus & Gegenfurtner, 2009).

This apparent correlation between cone shape and cone function, where it has previously been noted at all, has ironically been taken as evidence discounting the importance of the cone shape (Dowling, 1965). After all, the foveal cones that provide the best color vision do look rather rod-like and are the least cone-like in appearance. However, the central cones are, in fact, conical. Direct measurements (Borwein, et al., 1980) show that the foveal COSs typically have a proximal diameter (near their junction with the inner segment) of

Figure 2. Cone shapes as drawn by von Greeff (1900). Light enters the cones from the direction of the bottom of the figure. From left-to-right the six cone types depicted are from the periphery in to the fovea. The locations are: I - edge of the ora serrata, II - 3 mm in from the ora serrata, III - midway between the ora serrata and the visual center, IV - periphery of the macula, V - within the macula, and VI - at the foveal center.
about 1.0 μm and taper to a distal diameter of about 0.6 μm. This 40% diameter change over their typical length of 40 μm, gives a full cone taper angle of but half-a-degree, very rod-like in appearance, indeed, but still conical. What we suggest here is that the salient feature of the cone shape variation is not the taper angle itself, but rather the length over which the taper is spread, varying from a maximum of 40 μm for foveal cones to some 7 μm for far-peripheral cones. We will return to examine the significance of this issue below.

Seemingly then, there is a shape to color; the conical shape of the color receptors is somehow clearly important. It is a universal, defining characteristic of the color receptors, present across the entire gamut of vertebrate species, employs a renewal process different from that of rods that maintains their conical shape, and the shape itself varies systematically across the human retina in apparent correlation with the color vision function provided. So, why the cone shape?

Discussion

Cone Size

To address this question we also note another basic characteristic of the cones: namely that they are very small. The diameter of the cone outer segment is significantly smaller than that of the cone inner segment to which it is coupled. The inner segment diameter essentially limits the cone packing density in the retina and thus the photosensitive portion of the cone is much smaller than needed for resolving image detail alone. This would seem rather odd inasmuch as a larger outer segment diameter would permit the inclusion of more photosensitive pigment in the cone and perhaps greater sensitivity. The outer segment diameters are, in fact, barely larger than that of the wavelength of visible light itself. They are also refractile, with a refractive index larger than that of their surrounding medium, the interstitial matrix. This refractile property of the photoreceptors was noted long ago (Brücke, 1843), and conceptually expanded upon by O’Brien (1946) who noted that ray optics alone is insufficient to adequately describe optical propagation in the receptors because of their small size.

We know now that optical propagation in the retinal receptors must properly be described in terms of waveguide mode theory (Snyder, 1966). The optical “size” of the receptor waveguide is scaled in terms of the waveguide parameter, \( V \), given by

\[
V = \left( \pi d / \lambda \right) \sqrt{n_1^2 - n_2^2}
\]

where \( d \) is the (local) guide diameter, \( \lambda \) is the free-space wavelength of light, and \( n_1 \) and \( n_2 \) are the refractive indices inside and outside the guide, respectively. The parameter \( V \) determines the mode propagation characteristics of the waveguide. These modes essentially correspond to the allowed angles of reflection within the guide with wave-interference effects determining which angles, or modes, are allowed. As \( V \) decreases,
the number of allowed modes also decreases until for values of $V < 2.405$ only one mode, the so-called HE$_{11}$ mode is permitted. This mode too undergoes cut off as $V$ further decreases in the form of a characteristic cut off curve. We show in Figure 3 the computed cut off curves for small values of $V$, including that of both the first-order HE$_{11}$ and the second-order HE$_{21}$ modes. The plot displays the modal efficiency (essentially the ratio of optical power propagated within the confines of the guide to the total power in the mode) as a function of $V$. As a mode cuts off, less and less of the light remains inside the guide and is shunted to the external evanescent surface wave until it is lost altogether to radiation past mode cut off. Note that the HE$_{11}$ mode never exactly cuts off to the radiation field, although the fraction of light remaining inside the guide decreases rapidly towards zero for small values of $V$.

Transmission in the form of these waveguide modes in the photoreceptors has been directly observed in freshly excised retina (Enoch 1961a, 1961b, 1963). Furthermore, waveguide propagation is the only tenable explanation for the directional sensitivity of the retinal receptors, the Stiles-Crawford Effect of the First Kind (SC I) (Stiles & Crawford, 1933; O’Brien, 1946; Snyder & Pask, 1973). The best available information on the refractive indices of the retinal cone outer segment and its surround in the human retina (Sidman, 1957; Barer, 1957) give values of $n_1 = 1.387$ and $n_2 = 1.347$, respectively, resulting in a waveguide parameter relation for retinal cones of

$$V = 1.04 \frac{d}{\lambda},$$

or, to a good approximation, a waveguide parameter that is just the receptor diameter divided by the optical free-space wavelength. Thus for a foveal cone outer segment (with diameters ranging from 1.0 to 0.6 μm) over the range of visible light (roughly 700 nm to 400 nm or 0.7 to 0.4 μm) we have predicted $V$-values ranging from 2.5 to 0.8, precisely in the region of maximum cut off of the lowest-order HE$_{11}$ mode (Figure 3).

This wavelength-dependent cut off should thus be occurring in an optimal fashion along the length of the foveal

![Figure 3. Cut off curves for the two lowest-order waveguide modes plotting modal efficiency ($\eta$) as a function of waveguide parameter ($V$). $\eta$ is the ratio of power within the waveguide to the total power in the mode.](http://dx.doi.org/10.7287/peerj.preprints.331v1)
cone photosensitive outer segments (Medeiros, Borwein & McGowan 1977; Medeiros, 2006). What will this cut off look like? For incident white light, all colors will be present in the wide entrance end of the cone outer segment, but as the light propagates along the cone, waveguide conditions become progressively more restrictive and optical power is shunted out of the cone interior in a wavelength-dependent fashion. Long-wavelength, red light is shunted out first, then successively shorter-wavelength light is shunted out of the cone until only the shortest-wavelength, blue light remains in the cone outer segment interior. Essentially, light is removed from the cone interior according to its spectrum and we thus have the cone behaving as a miniature spectrometer.

There is little doubt that this spectral dispersion effect will be present in the cones. The key question is what might be done with the available spectral information? Do the retinal cones simply ignore this color information or make use of it in some fashion? How, in fact, could this information be detected and read-out to extract the color information? Could this spectral information be read-out in a manner consistent with the existing physiology of the retina and does it bear any relation to the way color vision function operates? A very simple process to accomplish this spectral read-out has previously been described (Medeiros, 2006); one that makes use of a long-puzzling feature of human vision and directly accounts for many aspects of color perception that are otherwise difficult to understand within the context of the standard three-cone model of color vision.

Coding Color Information

Perhaps the most natural way to encode the color information available in the cone is to convert the position-dependent information into a temporal code. Consider what would be required for this to be possible. At a minimum, there are at least three requirements. For one, photo-absorption of light must be a local event. That is, absorption events must be localizable to a small region along the length of the cone. Otherwise if the cone simply acts like an undifferentiated bucket of pigment there is no hope of extracting position-dependent color information. A second requirement is that there be some means of converting spatial information into a spread of time information. A third requirement is that there be some way of synchronizing the temporal information to enable the readout of sequential color information by correlating the beginning of a read out cycle with the start of the cone’s position-correlated information.

In regard to the first, localization requirement, there is ample evidence that photo-absorption is indeed localized to a small region along the cone length. For example, using micropipette electrodes, pioneering measurements by Hagens, Penn and Yoshikami (1970) determined that light absorption in rods (of rodents) produce a photocurrent source that is localized to within 12 μm (the limit of their experimental resolution) of the longitudinal position of light absorption. Higher resolution measurements by Gray-Keller, et al. (1999) determined that the longitudinal diffusion of
both cyclic guanosine monophosphate (cGMP) and Ca\(^{2+}\) second messengers in the lizard rod was confined to within approximately 3.5 \(\mu m\) of the light absorption site. More recently, Holcman and Korenbrot (2004) conducted a theoretical analysis which they verified by direct observation, of the longitudinal diffusion of cGMP in both retinal rods and cones. They found that the longitudinal spread (along the receptor length) of cGMP in the ROS to be 3 to 5 \(\mu m\) and for the COS, independently of the cone dimensions, to be 0.7 to 1.0 \(\mu m\). Thus, for foveal cones, photo-absorption events can be expected to be localized to a longitudinal region about one fortieth to one sixtieth of the cone length.

In regard to the second requirement of spreading out the position-correlated information in time, it is clear that there will be propagation delays for the electrical signal generated in the cone outer segment to travel the length of the cone back to the bipolar cells with which it is in synaptic contact. Note that we are not talking about optical propagation delays along the cone length (times on the order of a small fraction of a picosecond) but of electrical signal nerve conduction times (on the order of milliseconds). The absorption events occurring in the proximal portion of the cone (and thus associated with long wavelength red light) are closer to the synaptic junction and would be seen first. Photo-absorption events of short wavelength blue light occurring at the distal cone outer segment tip have the additional length of the outer segment to traverse and would be seen last. Thus electrical propagation delays along the cone length will be correlated with the spectrum dispersed along the cone length.

An estimate of the propagation times involved over the length of the COS is provided by the electrical RC (resistance times capacitance) times of the cone. Measurements by Attwell, Werblin and Wilson (1982) gave cone input resistances, when the inner segment was impaled by two electrodes, of up to 270 M\(\Omega\) (a value that is expected to be an underestimate of the true membrane resistance because of losses due to the impalement) and a computed capacitance of 85 pF. This gives an RC time constant for electrical signal propagation to be a minimum of some 23 msec. While this is only an estimate for propagation along the inner segment, we can expect propagation times along the outer segment to be larger (because of the lamellar membrane structure) and dependent on the length of outer segment over which the signal travels (greater length means greater resistance and capacitance involved). For example, Schnapf (1983) measured responses to a flash of light to be slower from the tip than from the base of toad rod outer segments. Whatever the detailed timing might be for signals propagated along the COS, we can expect that it be on the order of tens of milliseconds and to be dependent on the length of COS traversed.

Our final requirement is a means of synchronizing this information - how do we start the clock to know which signals are early and which are late? It has long been known that the eye continually undergoes microsaccadic eye movements (Ditchburn & Ginsborg, 1953; Riggs, et al., 1953; Martinez-Conde, 2004). In general, microsaccades exhibit a random pattern of motion, although they have a power spectrum centered at around ten
hertz and excursions that correspond to motions of the retinal image of some ten to twenty cone inner segment diameters on the retina.

It might be rather natural to assume that such movements represent residual pointing instabilities of the eye muscles directing the visual gaze and that such motions would tend to blur or otherwise reduce visual quality. So if these residual motions could be reduced or eliminated, one might expect that visual quality should improve. What has been found, however, is that rather than improving vision, stabilizing the retinal image to remove these microsaccadic movements through various optical techniques does not improve vision, but rather degrades it significantly. On stabilization of the retinal image, depending on the completeness of the stabilization, visual function and particularly color vision is rapidly and completely lost. It only returns if stabilization is removed, image motion is imposed, or if light on the retina is flashed or otherwise temporally modulated (Ditchburn & Fender, 1955; Ditchburn & Drysdale, 1977).

While these microsaccadic eye movements are clearly necessary for vision, the reason for them has hitherto been somewhat of a puzzle. They do, however, fit in with the requirement for providing the synchronization signal needed for dynamically reading out color information from the length of the cone. At each microsaccadic movement, a different part of the retinal scene is imaged on each retinal cone. If the new input comes from a region across a color border in the scene, the cone output will change as well and a new output sequence will start. This synchronization event is global across the entire retina and a new read-out sequence can thus start on the order of every 100 milliseconds or so (for a ten Hz microsaccadic jitter).

What does the resulting temporal code look like and are the times involved reasonable for color information read-out? Are there, in fact, different time delays for different colors and does it work the way this cone spectrometer model predicts? The issue of perceptual chromatic latency has long been controversial with many conflicting results reported; some researchers suggesting that blue perception is delayed relative to red, others that red is delayed relative to blue, and yet others that there is no difference at all (Uttal, 1973; Weingarten, 1972). This state of affairs is a reflection of the various techniques employed (trying to measure differential color delays of times that are to be measured in milliseconds by differences in reaction times to cues from different colors, for example) and of the blurring of the distinction between phase delays and true time delays.

We have previously described a technique (Medeiros, Caudle & Schildt, 1982; Medeiros, 2006) where we were able to visualize the separate and simultaneous perception by the rods and cones as a function of wavelength using images of monochromator slits scanning across a dark visual field. We used the achromatic response of the rod signal, which is demonstrably wavelength-independent, as a fixed reference by which to measure the relative latency of the cone color perception. Figure 4 displays the measured relative latency of color perception by the cones for two
observed rod perception to be purely white (no blue component as has often been suggested) and that the colors seen by cones alone have a much more saturated appearance then when they are seen with rod perception admixed as is the case in ordinary, everyday perception.

Now, it has long been know that it is possible to induce the perception of color, the so-called subjective colors, using appropriately modulated achromatic light (Fechner, 1838; Festinger, Allyn & White, 1971). Perhaps the best known example of what is now typically called pattern induced flicker colors (PIFCs) is Benham’s Top. Introduced as a toy during Victorian times (Benham, 1895). Benham’s Top consists of half-black, half-white disc with a series of black arcs on the white half of the disk. A typical configuration of Benham’s Top is shown as Figure 5. When spun at speeds of seven to fifteen revolutions per second, colors are seen on the edges of the black arcs as they blur into circles with red seen on the arcs positioned early in the white phase, green in the intermediate stage, and blue in the late stage of the white phase. Thus for the disc sketched in Figure 5, when spun counter-clockwise, the arcs take on the colors red, green, and blue starting from the outermost arcs in. On reversing the direction of rotation, the colors reverse as well so that from the outer arcs inward, the perceived colors are blue, green and red.

This phenomenon, long an apparent mystery of color vision (Cohen & Gordon, 1949; Campenhausen & Schramme, 1995), clearly shows that some kind of timing effect is mimicking the operation of the human color vision system (Sheppard, 1968). Evidently, Benham’s Top is presenting optical modulation with timing comparable to that being used by the color vision system and the visual system is interpreting that modulation as a color code. Consider the timing code being mimicked here. For optimum
presentation, the top is typically spun at about ten revolutions per second (within the 7-15 Hz range). This gives the timing difference (from the start of the white cycle on the disk) for the arcs that appear blue being behind the arcs that appear red of 2/3 of one half the cycle time. This amounts to a delay of blue being induced relative to red of 33 millisecond (range of 22 to 48 msec), in excellent accord with the chromatic latency measured for human color vision of 25 milliseconds for blue relative to red.

An important aspect of how this color coding must work relates to the propagation delays induced over the cone length that is to be subsequently read out as color information. We refer to the point raised earlier about the variation in cone length with retinal position and the apparent correlation of color discrimination function with the cone length. Given waveguide mode cut off dispersing spectral information over the length of the cone, it will be read out with some limited resolution as it is converted into a time code. However the conversion occurs, the available color information can be read with greater resolution if dispersed over a longer length than over a shorter one. This directly suggests an explanation of why foveal cones should have better color discrimination capability than shorter, more peripheral cones.

The limits of resolution on color discrimination by the proposed mechanism will depend importantly on the details of the read-out mechanism, photopigment distribution in the cones, cone length, and the longitudinal diffusion of a photo-absorption event. We can get some idea of these limits from the observations of Holcman and Korenbrot (2004) where longitudinal diffusion of cGMP in the COS of less than 1.0 μm suggests that a photo-absorption event is localized to at least this extent. Given the 40 μm length of the foveal COS and the dispersion of the visible spectrum of wavelengths ranging from about 650 nm to 450 nm over the cone, this diffusion would suggest a basic potential spectral discrimination available from a single COS to be better than about 5 nm (one fortieth of the 200 nm spectral span). While measurements of human color discrimination (minimum resolvable wavelength difference) varies over the spectrum and in different parts of the retina, and between different observers for various illumination conditions, the value of 5 nm for basic discrimination of wavelength
difference corresponds rather well to the experimental measures (Wright & Pitt, 1935; Bedford & Wyszecki, 1958).

Best human color discrimination is somewhat better than this suggested 5 nm, but recall that this value is what might be available from but a single cone. Experimental techniques to measure wavelength discrimination illuminate retinal areas spanning many cones. Recruitment of more cones can be expected to improve on the discrimination available from that of just a single cone. In this context, it is notable that in measurements made with point sources, Bedford and Wyszecki (1958) found that color discrimination for 1.5’ fields (corresponding to an illuminated area on the retina of about 9 um – spanning perhaps just two or three cone inner segment diameters) resulted in wavelength discrimination of 4 to 6 nm over most of the spectral range (about 650 to 450 nm). This small-field result is in remarkably good agreement with the prediction of the model on the basis of the physical limitations imposed by the basic physiology of the cones.

Thus far, we have pointed out that the retinal cone shape and size is of ideal form to exhibit spectral dispersion due to low-order waveguide mode cut-off over the length of the cone. The cones and the visual system evidently have the required machinery to convert this length-dependent spectral information into a time code, including localized photo-absorption, transmission delays correlated with the location of the absorption, and a synchronizing mechanism in the microsaccadic eye movements to read out the temporal code. We have described measured time delays inversely proportional to the wavelength and have seen that mimicking these time delays with pure black and white modulation produces subjective colors in direct accord with this time-color code. Furthermore, the spectral dispersion mechanism will provide better resolution for longer COSs suggesting a direct explanation for the correlation between color discrimination and cone length across the retina. In addition, cGMP diffusion processes implies single cone resolution for detecting color differences to be on the order of 5 nm or better, in accord with experimental measures.

We have here proposed a cone spectrometer model (CSM) for color vision, an intrinsically dynamic model that bears little resemblance to the standard, static Young-Helmholtz three-cone model of color vision. Given the entrenched nature of the standard model, is there sufficient reason to consider such a radically different alternative? Consider that there is actually quite a lot about the operation of human color perception that is not well explained by the standard three-cone model. Indeed, there are many aspects of color vision that flatly contradict the standard model. We address a few of these issues below in the context of showing how these features of color perception are in direct accord with the cone spectrometer model.
Color Vision

We suggest that there are a number of compelling reasons to consider this proposed CSM approach as the preferred context for the understanding of human color vision. While the three-cone model does proffer an easy way to understand the approximate trichromacy of color vision (in the sense that three independent variables are needed to specify color matches) and molecular genetics (Nathans, Thomas & Hogness, 1986) has provided evidence of multiple pigments, it does not necessarily follow that the three-cone model is the only way to explain color vision.

Dimensionality of Color Vision and Photopigments

Cell-counting studies of retinal neurons (Vilter, 1949; Missotten, 1974; Ahmad, et al., 2003) show that there are three output bipolar cells for every cone where color vision function is trichromatic in the central retina. This decreases to two bipolar cells per cone in more peripheral portions of the retina where color function is dichromatic, and to one-to-one in the far periphery where color vision function tends towards monochromacy (Sheppard, 1968). This does suggest that the coarse dimensionality of color vision is more directly tied to the subsequent processing of the basic information provided by the cones than by partitioning into separate buckets of color information at the very first step of detection through three different cone types.

There is no question that pigments are critical to color vision; after all, they are required to initiate the basic events in the transduction of light into an electrical signal. However, in the context of the three-cone model, all the emphasis of color vision function has been on the pigments and none on the structure. Perhaps this emphasis has been somewhat misplaced; in addition to our suggestion that structure is a critical feature for receptor function, we might note that Kefalov, Marsh-Armstrong and Yau (2003) found that cone or rod pigments work equally well when swapped into the other receptor type; that is, cone pigments in rods or rod pigments in cones. In this case they have the properties of the pigment native to the receptor. Apparently the environment or structure in which the pigment is located contributes importantly to its properties such as photosensitivity and response kinetics. Cone pigment behaves like rod pigment when in rods and rod pigment behaves like cone pigment when in cones. This too suggests some misplaced emphasis on the relation between photopigments and receptor structure.

Consider as well the research reported by Neitz, Neitz and Jacobs (1993) where they found normal color vision in their human subjects, despite the presence of more than three photopigments. The multiple pigment variations were found through examination of the photopigment genetic markers. The net result of their study led them to the conclusion: “The extra pigment types in people with normal color vision are sufficiently different to support tetra- or even pentachromacy. The fact that they don’t indicates that trichromacy of normal vision has its origin at a level of the visual pathway beyond that of the cone pigments, likely beyond the receptors.”
Multiple Pigments and Coexpression
The existence of multiple pigments is not a contraindication to the proposed CSM approach since the basic operation of the spectral dispersion mechanism is not directly dependent on the details of the absorption spectrum of whatever photopigment(s) might be in the cones, including even different pigments in different cones or even in the same cone. Indeed, the efficiency of the CSM mechanism would actually be enhanced by the presence of multiple pigments within the same cone. This enhancement would result if one were to position a long wavelength absorbing pigment in the proximal, wider portions of the cone and a shorter wavelength absorbing pigment in the more distal, narrower portions of the cone. This would enhance the detection of red light in the wide part of the cone, the only place it would be present because of mode cut off. While the detection of short wavelength blue light in the more distal portion of the cone is unambiguous (all other wavelengths have been removed by mode cut off by that point) there would be good reason to site short wavelength absorbing pigment in that region to enhance its detection.

Significantly, it has recently become evident that pigment coexpression - the presence of more than one pigment within a single receptor - is a wide-spread phenomenon in nature, both in vertebrates and invertebrates (Kitamoto, et al., 1998; Glösmann & Ahnelt, 2002; Jacobs, Fenwick & Williams, 2002; Lukâts et al., 2002, 2005; Glösmann, 2006; Nikonov et al., 2005, 2006, 2008). By examining genetic markers, pigment coexpression has been observed in mice, hamsters, guinea pigs, and butterflies. There have even been reports of multiple pigments in human cones as observed in some of the early microspectrophotometric measurements (Marks, Dobelle & MacNichol, 1964). Pigment coexpression in the cones is very problematical in the context of a three-cone model: a model that can only function sensibly if each cone of the three cone types contains only one of three photopigments exclusively.

In the case of the mouse that apparently has middle-wavelength-sensitive (M) and ultraviolet sensitive (UV) photopigments completely coexpressed in cones, Jacobs, Williams and Fenwick (2004) were none-the-less able to demonstrate discrimination “between some pairs of spectral stimuli under test conditions where luminance-related cues were irrelevant”. Their conclusion was: “Since mice can make dichromatic color discriminations, their visual systems must be able to exploit differences in the spectral absorption properties among the cones. Complete selective segregation of opsins into individual photoreceptors is apparently not a prerequisite for color vision.”

Trichromatic color vision with less than three photopigments
In a study of five deuteranopes and three protanopes (two variants of the common form of red-green color blindness) that were demonstrably dichromatic on common tests, Scheibner and Boynton (1968) were none-the-less able to demonstrate appropriate red-green discrimination. They concluded: “that most dichromats retain a residual and variably weak functioning of the red-green chromatic system.”
In a study of protanomalous color defective subjects, Neitz, et al. (1999) determined that a subset of the subjects they tested had only one distinct photopigment in the red-green portion of the spectrum (instead of the usual two). Protanomalous color deficit vision is a condition exhibiting reduced sensitivity in the red portion of the spectrum and reduced color discrimination ability (from normal) in the red-green portion of the spectrum. It is widely assumed, in the context of the three-cone model of color vision, that in the extreme form of protanomalous vision or protanomaly, the red pigment is entirely absent and there is only one green-absorbing pigment in the red-green portion of the spectrum. In the less extreme, anomalous versions, subjects are presumed to have two different pigments in the red-green region, but ones with absorption maxima with separations rather smaller than the usual 30 nm for color normals.

Neitz, et al. (1999) found, through examination of genetic markers, that one subset of their protanomalous subjects apparently had two pigments in the green region of the spectrum. These pigments had identical spectral absorption maxima, although perhaps expressed with somewhat different optical density. These subjects, despite having only one effective pigment in the red-green portion of the spectrum, none-the-less were able to discriminate colors in the red-green portion of the spectrum on a limited basis, a result that should not be possible in the three-cone model of color vision.

A related result was found by Crognale, et al. (1998) who tested subjects that demonstrably had only a single X-linked (red-green) pigment gene. They found that such subjects were never-the-less able to make chromatic discriminations in the red-green region, particularly when large fields were used. They explored - and subsequently ruled out - possible ways other receptors such as rods or “blue” cones could have provided assistance in color discrimination in this region. Their key conclusion was “The mechanism of chromatic discrimination in the presence of a single photopigment therefore remains unknown”.

Similar results were obtained by Wachtler, Dohrmann and Hertel (2004) with tests on dichromats. Their subjects had but a single X-chromosomal gene sequence and thus had only one red-green pigment. Yet these dichromats could use appropriate color names to indicate discrimination in the red-green and blue-yellow parts of the spectrum.

What all of these reports have in common is that the accepted view that exactly three pigments in three separate cone types are necessary for color vision is not supported by experimental observation.

**All color sensations in any one cone**
In a remarkable experiment, Hofer, Singer and Williams (2005), employed adaptive optics techniques to exclusively illuminate individual cones in the eye of alert human subjects. While many have tried before to get at the spectral sensitivity of single cones through illumination of small retinal spots, it has previously proved impossible to
illuminate a single living cone because the aberrations of the eye cause spot sizes to span several cones. In addition, without compensating for the involuntary microsaccadic eye movements, positioning on a single cone would not be possible in any event. Hofer, Singer and Williams (2005) were able to deal with both problems and to illuminate individual cones. Their adaptive optics technique compensates for the blur caused by the optical path through the eye. Fixation of gaze by their subjects in conjunction with very short stimulus flashes of 4 ms were used to circumvent motion blur due to microsaccades.

The were able to shine monochromatic light of 500, 550, or 600 nm wavelength on single cones in a number of subjects in retinal regions situated about 1° from the fovea. The key result of their investigation was that any cone could report any color, including white. No cones exclusively reported red or green or blue. As they noted, “The diversity in the color response could not be completely explained by combined L and M cone excitation, implying that photoreceptors within the same class can elicit more than one color sensation.” Such a result makes little sense in the context of the standard three-cone model.

**The Ives Result: Breakdown of static color matches under dynamic conditions**

Beyond these considerations, there are many other reasons to be dissatisfied with the three-cone model. For one such very important reason, we consider how the static three-cone model does under dynamic conditions. Herbert E. Ives conducted a basic, very simple experiment almost a century ago (Ives, 1918) that has profound implications for the three-cone model of color vision. Despite the importance of the experimental result and the considerable stature of Ives in the optical community, the result has been all but ignored by vision researchers.

Ives experimentally displayed the breakdown of statically-established color matches under dynamic conditions. In a very simple experiment, Ives produced a yellow slit of light that could be scanned across the retina by simple fixation in a given direction while the slit was scanned. It is well known that yellow light can be produced in more than one way; either as a pure monochromatic yellow sliced from the visible spectrum or by combining an appropriate mixture of red and green lights. In a static presentation, the pure yellow and the red and green mixture can be made to appear virtually indistinguishable; they are metamers of each other. In the standard three-cone model of color vision they are said to appear alike because they excite the “red” and “green” cones identically.

However, when these same slits of yellow that appeared identical when static, where scanned across the retina, their appearance was different. In the case of the “mixed” yellow composed of red and green lights, when scanned, the slit was seen to have a red leading edge and a green trailing edge with yellow in the middle. While the three-cone model would explain this result by saying that this simply means that the “red” cones
must have a shorter time constant than the “green” cones, it also says that the same result **must** occur with the pure yellow since it excites the two cone types the same way.

The fact is, it does not. The moving slit of pure monochromatic yellow light does not resolve into red and green leading and trailing edges, but remains the same yellow throughout. Such a difference for the two yellows is not possible in the three-cone model since the information that the two yellows are made in different ways is discarded in the very first step of detection by the cone “types”. No amount of subsequent processing in the retinal circuitry could bring back information that has been discarded in this way. The result of Ives’ experiment flatly contradicts the basic premise of the three-cone model of color vision.

The significance of this result should not be underestimated. The experimental result is so direct and unequivocal that it should be further explored, not ignored as has previously been the case. It was, in fact, in an effort to verify (or refute) Ives’ result that we explored the use of multiple slits scanning across the retina and subsequently found the process to directly separate rod and cone perception (Medeiros, Caudle & Schildt 1982; Medeiros, 2006). We used that technique to measure the chromatic latency data discussed earlier (Figure 4). It should be noted that we too observed the breakdown of metameric matches in our experimental setup, verifying Ives’ result. It is, evidently, the differential latency of light by wavelength that gives rise to the Ives’ result. Red light has a shorter latency than green light and yellow light has an intermediate latency. A mixture of red and green lights that statically matches yellow, will resolve into the mixture while a pure yellow will not because of these differences in chromatic latencies.

So in a direct and unequivocal fashion, we have a verifiable experimental result that directly contradicts the basic premise of the standard three-cone model of color vision. This same result is in direct accord with the CSM approach discussed here since it is these very differential chromatic latencies that are at the heart of how the model reads the color information generated by waveguide mode cut-off along the retinal cone taper.

Now there are many effects and properties known and measured about human color perception that any model should be able to explain. Given the sheer number of such effects, it will not be possible to address them all here (see Medeiros, 2006; 2008 for more on these issues). Here we will focus on just two of these color phenomena, effects that have no explanation in terms of past models, but do have a very direct accounting in our proposed model: the SC II Effect and the similarity between violet and purple.

**The SC II Effect**
Interest in the optical guiding characteristics of retinal receptors can be said to have arisen in connection with the discovery of the directional sensitivity of the retina by Stiles and Crawford (1933). They discovered an effect whereby the apparent brightness of a light decreases with increasing angular deviation from nominal incidence. That this
Stiles-Crawford effect of the first kind (SC I), is a manifestation in some form of optical guiding in the receptors is no longer seriously doubted (O’Brien, 1946; Snyder & Pask, 1973a; Westheimer, 2008).

There is an analogous directional effect in color perception, the Stiles-Crawford effect of the second kind (SC II), first demonstrated by Stiles (1937). In this effect, the perceived color of a stimulus alters as its angle of incidence at the retina is altered, primarily shifting to longer wavelengths with increasing angle. Explanations of this color shift in terms of pigment self-screening effects in a three-photopigment model have previously been proposed (Walraven & Bouman, 1960; Wijngaard, Bouman, & Budding, 1974). This explanation postulates directional color shifts arising through the preferentially greater absorption with greater path lengths through a pigment, which occurs for light of those wavelengths near the pigment absorption maxima. While with a sufficiently large number of free parameters to choose from (wavelengths of three photopigment absorption maxima, photopigment densities and distribution) one can mimic some of the observed directional color shift with this approach, the pigment densities one must postulate are improbably large (Enoch & Stiles, 1961). In a review of the SC II effect, Alpern (1986) concluded “that ‘self-screening’ theory may not provide a satisfactory description of the color changes throughout the visible spectrum”. In a direct study of the SC II effect at high bleaching levels, Wooten, et al. (1978) concluded that their results “strongly indicate that self-screening is an inadequate explanation.” They further stated that: “Unless there is some factor of which we are unaware, we must conclude that the only available framework that is capable of accounting for our high bleach measurements is waveguide theory.”

An effect similar to the SC II color shift for transmitted light was also observed in bleached retinal preparations by Enoch (1961b) where “The changes in modal pattern induced by oblique irradiation result in many instances in the same physical distributions of energy which are obtained by increasing wavelength.” That this effect was observed in bleached preparations requires an explanation other than pigment self-screening in this case as well.

A basic wavelength-dependent directional property of optical waveguides is just the red shift observed as the basic feature of the SC II physiological effect (and in Enoch’s retinal preparations). An explanation of the SC II effect on just this basis has been previously proposed (Medeiros, 1979). In a dielectric fiber, the propagation constant that determines how light is guided in the fiber specifically depends not on the wave vector \( \mathbf{k} \) itself, but rather on its component in the \( z \)-direction of the guide axis (with unit vector \( \mathbf{z} \)) given by \( k \cdot \mathbf{z} = k \cos(\theta) \) (c.f., Marcuse, 1974). Thus light of physical wavelength \( \lambda \) (with wave number \( k_1 = 2\pi n_1/\lambda \)) incident at an angle \( \theta \) to the guide axis in a medium of refractive index \( n_1 \) can be thought of as equivalent to light propagating axially but with an effective or guide wavelength given by

\[
\lambda_g = \frac{\lambda}{\cos(\theta)}.
\]
The predicted shift in $\lambda_g$ is thus always to longer wavelengths, i.e., a red shift.

The simple $\lambda/\cos(\theta)$ function is directly compared to the SC II data of Stiles (1937) in Figure 6. That chart displays Stiles’ original data (data points) for the apparent change in perceived color for eleven different wavelengths. Each set of measurements is arbitrarily displaced vertically for clarity. The solid curve for each test wavelength is the $\lambda/\cos(\theta)$ function where the angle of incidence at the retina, $\theta$, is determined for different pupil entry points assuming an eye of normal focal length. Each such curve is also displaced the same amount horizontally to account for the apparent eccentric center of pupil entry. For most of the test wavelengths this simple function is a remarkably good fit over all angles, corresponding to the displaced pupil entry. Even where it does depart from this form (shifting back towards shorter wavelengths at large angles for the test wavelengths of 542, 522, and 500 nm) the data is still fit by the $\lambda/\cos(\theta)$ form at small angles in all cases.

The $\lambda/\cos(\theta)$ dependence for guide wavelength is a general property of optical waveguides and thus any color vision model based on some wavelength-dependent feature of optical transmission would predict the same basic form for the SC II effect. Where different models would vary is in their explanation of the deviations from the basic

Figure 6. A display of Stiles (1937) measurements of the SC II Effect at eleven wavelengths (dots). The solid curves are computed as the guide wavelength $\lambda/\cos\theta$ where $\lambda$ is the specified wavelength and $\theta$ is the angle of incidence at the retina for an eye of normal focal length for the pupil displacement specified in the abscissa of the plot. Each set of eleven measurements is arbitrarily displaced in ordinate for clarity.
\( \lambda / \cos(\theta) \) form in the middle wavelength range for the larger angles. We note that this \( \lambda / \cos(\theta) \) dependence is the color shift for a single, isolated cone. The human retina is, however, a closely packed array of receptors. Since light incident on a fiber at sufficiently extreme angles can escape to the external medium, then in the retina this escaping radiation can couple to the surrounding receptors. The escaping light is downward directed and is incident on the sides of the neighboring cones at a more distal region than that of the receptor from which it escaped. That there is thus more light incident more distally on the receptors than is normally the case for axially incident light would result in a blue shift for large angles since an increase in the signal generated in the narrow portions of the cone is associated with short wavelengths.

This blue-shift effect for large angles would be a rather inefficient effect since waveguide coupling to neighboring cones would similarly be rather inefficient. It is notable that Chen and Makous (1989) observed that “over half the light that cones absorb when light enters the margin of the pupil is light that has previously passed through other cones”. Whatever the details of this coupling process, it can be expected to be greatest for light of those wavelengths near the absorption maximum of the photopigment contained in that part of the cone on which the sideways-directed light is incident. The maximum of photopic sensitivity is, in fact, in just the region where high-angle blue shifting occurs (542 - 500 nm). Thus the CSM approach provides a straightforward explanation of the basic red shift of the SC II effect as well as a plausible method of explaining the large-angle blue shift near the photopic sensitivity maximum.

**Violet-Purple Similarity**

It has long been noted that short wavelength monochromatic violet light is similar in appearance to a mixture of red and blue light, the color purple. This similarity of violet and purple has been difficult to explain. Attempts to explain the effect in terms of the standard three-cone model of color vision have postulated the existence of a subsidiary maximum of the “red” absorbing pigment in the violet region of the spectrum. However, no such sub-maximum has been found for the retinal pigments and the understanding of the violet-purple similarity remains elusive.

However, in terms of the proposed cone spectrometer model, a rather direct explanation of this effect is available. Note that we displayed two waveguide mode cut off curves in Figure 3, one for the lowest-order HE\(_{11}\) mode and one for the second-order HE\(_{21}\) mode. Usually, only the lowest-order mode would be propagating in the retinal cones. However, for values of the waveguide parameter of 2.405 or greater, the second-order mode can also propagate in the cones since it will be above cut-off. Values of the waveguide parameter of \( V > 2.405 \) will occur for the shortest wavelengths propagating in the largest cone widths, i.e., larger values of \( \lambda / d \). Given an entrance diameter of 1 \( \mu \)m for the COS, \( V \) will be large enough for second-order mode propagation, for wavelengths shorter than about 430 nm (using the relation \( V = 1.04 \lambda / d \)). This turn-on of second-order mode propagation is very close in wavelength to the abrupt appearance of violet at wavelengths on the order of 460 nm or shorter. Note that, corresponding to
the abrupt cut off characteristic of the HE\textsubscript{21} mode, if one looks at the output of a variable wavelength source and that source is gradually tuned to shorter wavelengths, there is a similarly abrupt appearance of a red component defining the occurrence of violet perception.

The CSM explanation of the violet-purple similarity, then, is just due to the excitation of the second-order waveguide mode at short wavelengths. The HE\textsubscript{21} component of sufficiently short wavelength light will behave like long-wavelength, red light (attenuating rapidly after propagating a short distance at the entrance end of the cone) and will have an abrupt onset due to the very sharp cut off curve of the second-order mode. The HE\textsubscript{11} component of the short wavelength light will propagate like blue light, traversing the length of the cone. Thus violet light, entering the cone as two waveguide components will propagate as if it had both a long-wavelength and a short-wavelength component. This similarity between a mixture of red and blue (purple) and a mixture of the two waveguide modes propagating at short wavelengths is schematically illustrated in Figure 7.

To this point we have discussed what works in the CSM model and how it accounts for some color vision phenomena that have found no acceptable explanation in previous models. While hardly a complete accounting of the vast array of color vision phenomenology, this discussion has surely indicated the potential utility of this novel approach towards the understanding of some long-standing mysteries of color perception. Additional color vision effects and phenomena in relation to the proposed model (including the basic shapes of the color saturation and color discrimination curves) have been addressed elsewhere (Medeiros, 2006; 2008) and will not be addressed here. However, the present discussion cannot be complete without some examination of what happens when the waveguide dispersion mechanism fails to work correctly. Do the ways in which the discrimination mechanism can fail bear any relation to the known forms of defective color vision and can

![Figure 7. Schematic representation of purple (on the left) and violet (right) light illuminating a cone. Purple as a mixture of red and blue has a long wavelength component that exits the cone early and a short wave component exiting late. Violet light has a similar behavior with the HE\textsubscript{21} component exiting early and the HE\textsubscript{11} component exiting late.](image)
Color Deficit Vision
No discussion of color vision models can be complete without an accounting of how color vision function can go wrong and "color blindness" result. There are a number of ways in which color vision function fails, but by far the most common are the two types of red-green deficit vision, protanopia and deuteranopia. Both of these forms exhibit a gradation from nearly normal color vision through their anomalous variations (protanomalous and deuteranomalous) to the near complete loss of the ability to distinguish reds and greens effectively. These forms of color deficit vision have X-linked recessive genetics and affect around 8% of the male population and a much smaller percentage of females.

Color deficit vision is itself a major topic of investigation with a vast body of research and published literature. It is far beyond the scope of the present discussion to cover the breadth of issues involved. However, it is possible to examine some of the basic issues in the context of the proposed CSM approach. We may note that perhaps the most direct way in which the spectral discrimination mechanism will fail, or at least function anomalously, is if the cones are of the wrong "size". This could happen in either of two ways; they could be too small or they could be too large. Either would cause deficient color vision in subtly different ways. It should be noted that this incorrect sizing might have nothing to do with the cone's physical size but be caused by abnormal values of the refractive index difference between the cones and their surrounding interstitial medium. Recall that the measure of receptor "size" is the dimensionless waveguide parameter V, given by

\[ V = \frac{n_1^2 - n_2^2}{\lambda} \]

and that for optical guiding, \( n_1 \) must be larger than \( n_2 \). For a cone with core refractive index \( n_1 \) and a diameter \( d \) (at one location along its length) if the refractive index, \( n_2 \), of the surrounding medium is larger than normal, the cones will effectively be smaller than they should be for optimum operation. Conversely, if the index of the interstitial medium is smaller than normal, the cones will be effectively larger than optimal.

If the cones are too small, then longer wavelength (red) light will be inefficiently coupled into the cone's photosensitive outer segment (since V is very small). This will have two immediate consequences: a lower sensitivity to red light and inefficient use of the cone length for discrimination. Both of these characteristics are signatures of one type of color deficit vision, protanopia, where there is both low sensitivity to red light and poor color discrimination.

If the cones are too large, then all colors will couple efficiently to the cone's photosensitive outer segment (V is larger), but the spectrum will be dispersed (at least in the lowest-order, HE11 mode) inefficiently over the cone with mode cut off dispersing.
the spectrum only over the lower portions of the cone (although there will surely be complications from second-order mode dispersion in wider portions of the cone). The consequences of this mistuning will be possibly reduced color discrimination and no reduction in spectral sensitivity, both of which are signature characteristics of deuteranopia.

In the proposed CSM approach, color discrimination involves much more than just spectral dispersion along the cones. There must be a conversion of position-correlated color information into a readable temporal code. There must as well be some rather sophisticated processing by successive neural elements to dimension and parse the color information into the ways that the visual cortex eventually translates into color perception. None the less, in the context of our proposed model, if one can purposefully alter the refractive index of the cone’s surrounding medium appropriately, it is not difficult to imagine that some useful color vision function could be restored to “color blind” individuals. To “cure” protanopia one would need to increase the refractive index difference between the cones and its surround. This could be accomplished by decreasing the refractive index of the interstitial medium. Conversely, to treat a deutan defect, it would be necessary to decrease the refractive index difference between the cones and its surround by increasing the refractive index of the surrounding medium.

It may be of interest that some version of this may already have been serendipitously demonstrated. In a number of little-noted papers published in the 1970’s, Louis F. Raymond reported on the treatment of a number of patients with allergies who were also colorblind (Raymond, 1971; 1972; 1975). He cites one case of a patient who tested as red-green color deficit on both Ishihara plates and Hardy-Rand-Rittler Plates. The patient tested positive for allergies to bacterial endotoxins and airborne pollens in intradermal tests. Hyposensitization treatment, consisting of the administration of diluted antigenic solutions of the items to which he was allergic cured the colorblindness. He claimed that the patient’s color vision was normal on testing after two months, one year, and at two years later and he noted that he treated a total of 24 such cases with similar results.

There are a number of identifiable antibodies in human sera. Chemically, these are glycoproteins with molecular weights of around 150 to 200 kDa. These antibodies are a variety of immunoglobulins, including immunoglobulin E (IgE). It has been shown that IgE levels may be reduced following hyposensitization treatments. Since the immunoglobulins are chemically similar to the mucopolysaccharides that are known to be present in the interstitial medium of the retina, conceivably, the hyposensitization treatments may have altered (decreased) the refractive index of the interstitial medium and thereby restored color discrimination function.

If these results can be confirmed, and, if this is indeed a possible avenue for the restoration of color vision function, then the patients that would have benefited from this particular approach should be those with protanopia. That is, decreasing the density of
the interstitial medium through the injections should decrease the refractive index of the surrounding medium and thus increase the effective size of the cones. This could be expected to restore the subject’s color vision function so long as the appropriate retinal circuitry was functional. Unfortunately, Raymond never specified which type (or types) of the red-green color deficit vision he treated. In any event, given the large number of individuals in the general population afflicted with color deficit vision, this is clearly an area of clinical research that may be of interest to explore.

In what might be a related effect, Mancuso, et al. (2009) recently reported on experiments in which they provided gene therapy to adult male squirrel monkeys (Saimiri sciureus) with dichromatic vision (protanopia). They used sub-retinal injections of viral particle carriers of the gene for the long-pigment opsin. Quoting from their paper:

*Classic visual deprivation experiments (Wiesel & Hubel, 1963) have led to the expectation that neural connections established during development would not appropriately process an input that was not present from birth. Therefore, it was believed that the treatment of congenital vision disorders would be ineffective unless administered to the very young. However, here we show that the addition of a third opsin in adult red–green colour-deficient primates was sufficient to produce trichromatic colour vision behaviour.*

Thus they demonstrably provided trichromatic color vision where it had not been present before by means of the sub-retinal injections. Their interpretation of their results was that they had added a formerly missing, third class of cones to complete the triad necessary in the three-cone model of color vision. They expressed what was probably appropriate surprise that simply adding a “third class of cones” by this method was sufficient to provide trichromatic color vision with no extensive change of neural processing in the retina. However, we suggest that an alternative explanation of their results is that they altered the refractive index of the interstitial medium of the retina and “re-tuned” the too-small protanopic cones into a more normal range similar to what Raymond had done in the 1970’s with his human patients.

A direct consequence of our proposed cone spectrometer model is that color discrimination function at the level of the retina is impacted by alterations of the refractive index environment of the photoreceptors as well as by any physical disruption in the orientation or form of the photoreceptors. It is well known that many retinal diseases and various drug toxicities can affect color vision function (Verriest, 1963; Smith, Pokorny & Diddle, 1978; Hart, 1987; Mamor & Kessler, 1999). That these diseases and chemical agents can affect the physical form and environment of the photoreceptors suggests a mechanism by which they may disrupt color discrimination function by the cones.
Consider, for example, Central Serous Chorioretinopathy (CSC). This retinal disease is characterized by serous neurosensory detachments of macular tissue, and many patients complain of abnormal color vision in addition to other visual anomalies including blur, micropsia, metamorphopsia, and scotoma (Pokorny & Smith, 1986; Bek & Kandi, 2000; Piccolino, et al., 2005). Retinal examination of these patents reveal serous fluid elevating the macular photoreceptor layer which strongly suggests a change in refractive index in the environment of these cells. In the context of our Cone Spectrometer Model these physical alterations provides a plausible avenue for the explanation for the dyschromatopsia observed in patients with CSC.

Yet another ocular pathology is a version of color deficit vision where there is true color blindness or achromatopsia. There have been a number of cases of complete color blindness where the subjects donated their eyes for scientific research after death. Autopsy results showed, in some cases, a normal population of cones, and in others a drastically reduced number of cones (Alpern, Falls & Lee, 1960; Harrison Hoefnagel & Hayward, 1960; Falls, Wolter & Alpern, 1965; Glickstein & Heath, 1975). However, one common feature found by all the histological studies was that what cones were present were abnormal in shape with outer segments that were either grossly misshapen or abnormally squat and short. Whatever else may have been wrong in these eyes, it is quite clear that in the context of the cone spectrometer model such cones could not have provided useful color discrimination.

**Conclusion**

Our goal in this paper has been to outline the case that, in at least a figurative sense, there is a shape to color, specifically the tapered shape of the cones. The retinal cone shape is evidently a universal characteristic of the color receptors and distinguishes them from the achromatic rod receptors. The cone shape varies systematically across the primate retina in apparent correlation with the color vision function provided. There has previously been little attention paid to this dominant characteristic of the color receptors and certainly little attempt to connect their physical form to their function.

We have described a simple and direct effect of low-order waveguide mode cut off that will disperse the optical spectrum along the length of an appropriately sized cone. Data suggests that the retinal cones are ideally dimensioned to optimally exhibit this mode cut off along their photosensitive outer segments. Waveguide mode propagation has been directly observed in excised primate retina but to date no direct role has been suggested for waveguide propagation except in connection with the directional sensitivity of the receptors, the Stiles-Crawford effect of the first kind.

We have described how the existing physiology of the eye provides the basic three requirements needed for converting the position-dependent color information dispersed along the cone length into a readable temporal code. These are the localization of photoabsorption events within the cone to a small longitudinal region, the delay in
propagation along the cone length associated with the site of absorption, and the utility of the microsaccadic eye movements to synchronize the time code. We showed experimental results that, in accord with our Cone Spectrometer Model, chromatic latency is proportional to decreasing wavelength with the perception of blue (450 nm) delayed by about 25 msec relative to red (650 nm). The operation of this time code directly explains subjective color effects such as seen in Benham’s Top with the temporal delays necessary to invoke the induced colors directly matching the observed chromatic latencies of the retinal cones for operation of the Cone Spectrometer Model.

We covered a number of topics related to the phenomenology of color vision that point to the inadequacy of the current models in terms of failure to readily explain color deficit vision or pigment coexpression within the cones. Experimental demonstration that individual cones can report the sensation of any color, including white, are inconsistent with explanations of color vision based on multiple cone classes, although it is directly compatible with the CSM approach.

Most importantly, we described and confirmed the Ives’ experiment illustrating the breakdown of statically established color matches under dynamic conditions. Such a result directly contradicts the very basic premise of three-cone models although it fits in precisely with the workings of the proposed model.

Furthermore, we showed that the proposed Cone Spectrometer Model directly explains the long-puzzling color change effect with increased angle of incidence at the retina, the Stiles-Crawford effect of the second kind as a basic waveguide property due to propagation as longer apparent-wavelengths for non-axial incidence of light on the cones. The simple $\lambda/cos\theta$ behavior predicted for the model matches very well with the experimental data except for large angles near photopic maximum sensitivity. This second-order departure from the simple form is plausibly explained in terms of optical coupling to distal portions of neighboring cones.

The model also provides a direct explanation of the previously inexplicable similarity of violet and purple in terms of basic waveguide behavior. For sufficiently short wavelengths (violet) second-order mode propagation is possible in the cones. This second-order mode component will behave like “red” light since it is shunted out of the cone after only a short propagation distance in the COS while the first-order mode component propagates as blue along the full length of the cone. Sufficiently short wavelengths of light will thus behave as if two components were present: a red component and a blue component, i.e., a purple mixture.

Finally we examined the routes by which the model behavior could fail. A simple failure mechanism is the mistuning of spectral dispersion that will result for cones being the wrong “size”. As we discussed, this “size” need not be a physical size, but an optical size as determined by the refractive index difference of the cone and its surround. With cones being effectively either too small or else too large, they would exhibit anomalous
behavior similar to the two main forms of color blindness, protanopia and deuteranopia, respectively. Previous allergy treatments of human patients may have serendipitously “cured” their color blindness through coincidental manipulation of the refractive index of the medium surrounding the retinal cones. Similar alterations may have played a role in the enhancement of color vision function in recent experiments on gene therapy of protanopic monkeys.

The bottom line to all this is that the eye is indeed a marvelous instrument of seeing, perhaps even more cleverly constructed than had been previously suggested. Each human eye is composed of an array of millions of sublimely constructed spectroscopic detectors that can each resolve the world of colors in a way no mere gross partitioning into three buckets of color could ever hope to accomplish. It would seem implausibly extravagant that such an elegant mechanism would be provided by the shape of the retinal cones and yet not be used for the basic function they can provide.

We have essentially described a dynamic model of human color vision, one that is better suited to describing the facts of retinal anatomy and perceptual phenomenology than current models. The subject of human color vision encompasses a wide range of seemingly complex phenomenology and it certainly has generated a vast array of research and investigational reports. But much about color vision remains unexplained or puzzling within the context of the standard three-cone model. We suggest that the CSM approach as described here offers an opportunity to reach a more unified and coherent understanding of human color vision.

References


Moreland JD, Cruz A (1959) Colour perception with the peripheral retina. Optica Acta **5**: 117-151.


Raymond LF (1975) Physiology of color vision and the pathological changes in reversible color blindness, a deficiency disease of the retina, Ann Ophthal 7: 532-534.


