https://doi.org/10.2352/ISSN.2169-2629.2018.26.318

Can Trichromacy equal Tetrachromacy?

Thomas Bangert, Ebroul Izquierdo; Queen Mary University of London; London, England

Abstract

Although our distant ancestors once had the same visual sensor arrangement as extant birds and reptiles, colour vision in most mammals is very limited; essentially reduced to a single dimension of information produced by a pair of colour sensors, each with a fairly broad spectral response. Humans belong to a small subset of mammals that have re-developed some of the ability that was lost by their distant ancestors. They have done this by stretching the natural variation of the light absorbing pigments in one of the visual sensors, thereby adapting it toward a third colour sensor. It is generally assumed that due to the differences and limitations of this solution that colour perception in humans is fundamentally different to that of our distant relations; animals which have been able to continually extend and refine their colour vision over a period of more than two hundred million years. However, over that extended period of time the fundamentals of their colour vision have remained largely unchanged; a system of four visual sensors, each with a specific colour filter that systematically narrows the spectral response. In this paper we demonstrate that a three-sensor system with a broad spectral response (similar in spectral profile to that of the human colour sensors) can in certain conditions be equivalent to a four-sensor system (broadly similar to bird and reptile colour sensors). This suggests that human colour perception (sometimes referred to as a 'trichromatic' system) may be broadly equivalent to colour perception found in a variety of other animals such as birds or reptiles (whose colour vision is generally referred to as 'tetrachromatic'). The difference between a sensor system that uses four spectrally narrow colour sensors arranged into two opponent pairs and a three-sensor system is that the former relies on a simple difference measurement whereas the later requires the use of complex trigonometric functions – which are at least two orders of magnitude greater in computational complexity.

Introduction

The ability to see is so vitally important that it developed in the earliest of our ancestors [3, 8]. The sensors these distant ancestors developed were simple neurons with the ability to be activated (or trigger activation) by photon absorption. Once developed, these sensors have seen little change in hundreds of millions of years of evolution [2]. The sensors differ primarily in the use of the material used to absorb photons (the *pigment*), which determines the wavelength of light the sensor responds best to [5, 7]. The basis for all colour perception is this spectral difference in sensor sensitivity. Many animals (such as birds or reptiles) for whom colour vision is particularly important use a special purpose coloured oil droplet in the outward facing segment of the sensor neuron to enhance this spectral difference [3]. The *colour* of this oil droplet allows the different types of colour sensor to be readily classified. The visual sensor arrays of natural visual systems that use this arrangement of optical filter to narrow the spectral response commonly use four colour sensors, as shown in Figure 1.

The nature of the visual sensors responsible for colour vision is therefore reasonably well understood. How the information from these sensors is used to support colour perception is less well understood. Colour perception has at its core the processing of luminance information and this is better understood. Luminance measurements in the visual systems of natural organisms are never taken in isolation as a measure of a single sensor; rather, a luminance measure is always the relative difference of two associated sensors – sensors from whom a difference measure is taken and which are arranged into opponent pairs. Colour is derived from the same opponent system; the difference being that for colour an opponent pair is made up of visual sensors that are sensitive to light from different regions of the visual spectrum. Colour information is, therefore, the measure of spectrally restricted luminance difference rather than the measure of spatial luminance difference. Although a *luminance model* is trivial to define a general model of colour has been more resistant to analysis; there is currently no theory of colour that provides a general framework from which sensor values for colour may be mapped to perception, even for animals that have relatively simple visual systems. In fact, it has been suggested for birds that ...

> ... it is difficult – impossible in fact – for humans to know what their perception of colours is actually like ... They not only see in the near ultraviolet, but they also can see colours that we cannot even envision ... a variety of colours that lie beyond direct human experience. [3]

One of the difficulties in understanding colour vision is that the primary focus for colour has traditionally been on human colour perception. An introspective approach to human colour perception leads to the finding that human colour perception is organized into four perceptually unique primary colours which are arranged into two mutually opposing pairs; something which is consistent with colour as an opponent measure, and which suggests two pairs of colour sensors such as those found in birds or reptiles. On the other hand, a study of the physiology of colour vision shows that humans only have three colour sensors. Moreover, when human colour perception is studied systematically by matching colours with single wavelength monochromatic colours, the results are inconsistent with opponency (which operates on the basis of simple subtraction, just as with luminance). Human colour perception appears to be the product of a ratio of the sum of the three sensors. Nevertheless, a study of the neural processing of sensor information indicates that colour information is at some point in the processing pathway put into opponent form [4]. These inconsistencies have not as of yet been reconciled and as a result there is no generally agreed model of colour vision that reliably maps sensor values to colour perception.

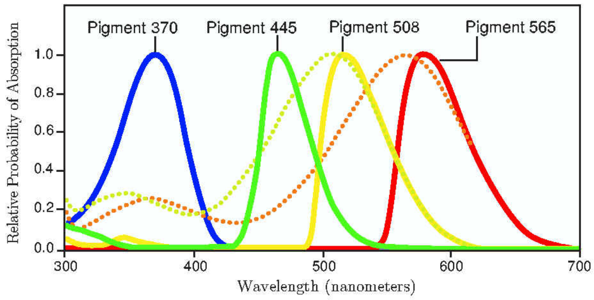

Figure 1 – Many birds have four colour sensors, each with an oil droplet at the tip which acts as a colour filter. Adapted from Goldsmith [3].

The Standard CIE Colour Model

When human subjects are presented with monochromatic stimuli (light consisting of just a single wavelength) and asked to match the colour, the functions that relate the monochromatic wavelength to the primaries used to produce the colour indicate that human colour vision is inherently ‘trichromatic’; that is, human colour vision can be described (more or less accurately) by three colour matching functions. These functions, when put into standard form, can quite readily be seen as sensor functions. Figure 2 shows the XYZ colour matching functions standardized by the CIE in 1931. A ratio of the sum of the colour matching functions (shown by the equations below, labelled as Equation 1) allows the functions to be plotted in two dimensions – which are generally referred to as chromaticity coordinates. When the chromaticity coordinates for all the monochromatic wavelengths the sensors are receptive to are plotted (as shown in Figure 3) they endeavour to enclose a region of two-dimensional space that may be seen as either somewhat circular or somewhat triangular. This ‘chromaticity’ diagram, or *colour space* is a very useful way to represent colour information. When all wavelengths within the range of the colour matching functions are equal in amplitude the colour coordinate will be (⅓,⅓), which is the centre of the colour space. Light with all wavelengths at equal energy is perceived as white light and therefore the centre of the colour space is considered the *white point*. The area between the outer boundary of pure monochromatic colours and the white centre point may be seen as a mixture of white light and a pure monochromatic light. Any arbitrary colour stimulus, irrespective of the complexity of its spectral power distribution, may be measured by the colour matching functions and the associated chromaticity coordinate will fall within the colour space, and can thereby be expressed as a mixture of a pure monochromatic colour (the *dominant wavelength*) and a white component (the *saturation* or *white level*). A fully saturated colour is equivalent to a monochromatic colour and a colour that is completely unsaturated (with a white level at 100%) is equal to white without any colour component.

$$x = \frac{X}{X+Y+Z} \ , \ y = \frac{Y}{X+Y+Z} \tag{1}$$

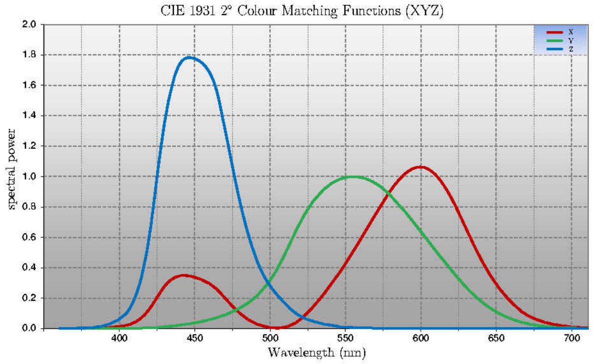

Figure 2 – CIE 1931 2° colour matching functions.

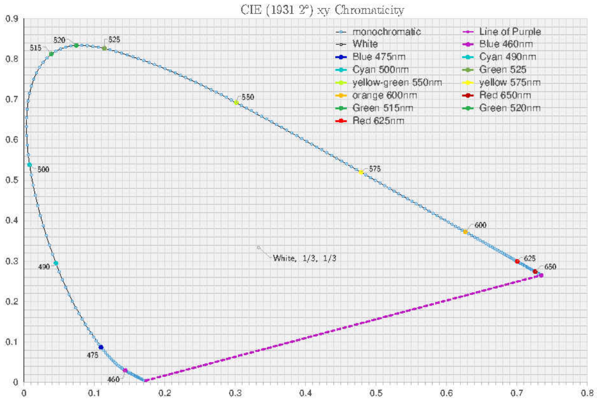

Figure 3 – CIE 1931 2° chromaticity space. Each point on the graph shows the xy coordinates for monochromatic light between approximately 400 and 700 nanometres, which are the wavelengths of light that the visual sensors of the human visual system are sensitive to.

It is known, however, that the standard CIE colour matching functions do not completely reflect the abilities of the human visual system. They are a useful approximation to human colour perception, but with minor well-known deviations from how the human visual system is known to perform. One deficiency inherent to the standard CIE colour matching functions is that the relation between wavelength and saturation is irregular, as shown by Figure 4. Furthermore, the outer boundary of the colour space itself varies irregularly between being linear and circular; and as a result, the function that maps a two-dimensional chromaticity coordinate to its respective dominant wavelength & saturation is not computable. There is also no computable inverse function to translate dominant wavelength & saturation back to a chromaticity coordinate. From an information processing point of view this is highly undesirable. It is possible to approximate these transforms with a look-up table, but this is not a practical solution that can be implemented for rapid low-level processing. The human visual system is able to efficiently produce colour perception from its visual sensors, and it follows from this that the human visual system must have available to it colour functions (and an associated chromaticity space) that are computable (provided that perception is fundamentally the product of computation). The colour functions of the human visual system cannot therefore be the same as those of the CIE XYZ system, which are not computable. The featureless irregularity of the CIE chromaticity space also leads to an inability to associate colour with chromaticity; one may calculate the angle of a chromaticity coordinate but this neither resolves the wavelength associated with that angle nor does it resolve the colour of that wavelength (hue). The ability to calculate colour may be considered fundamental to any visual system.

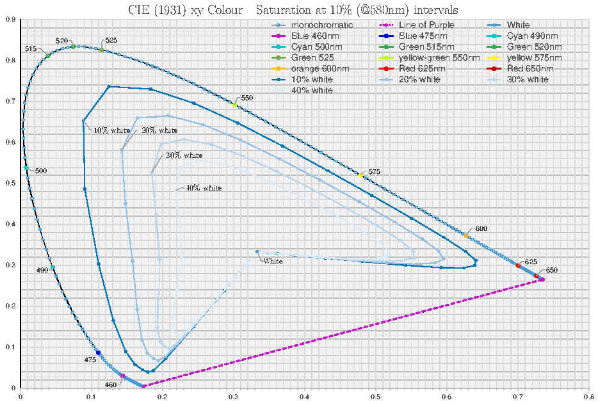

Figure 4 – CIE 1931, 2° chromaticity space with saturation. The outer boundary consists of the *pure* monochromatic colours and the central point is achromatic white. All other colours within the chromaticity space are a mixture of a monochromatic colour and white.

Dual Opponent Colour

While perceptual colour matching in humans shows that human colour vision relies on three sensors it is difficult (if not "impossible") to replicate this kind of perceptual study in animals that have four colour sensors, such as birds or reptiles. A recent study on zebra finches (a common type of songbird) that has four visual sensors (similar to that shown by Figure 1) has shown some evidence for categorical perception in birds but this was only shown for yellow-orange to red colours and is insufficient for estimating the colour space [1]. On the other hand, it is straight forward to design a colour space that relies on four sensors. Studies of human colour vision using monochromatic light indicate that the function of low-level (retinal) processing of visual information is largely to code chromaticity. This requires a sensor space that allows chromaticity coordinates to be translated to dominant wavelength and saturation. This is difficult to do with the standard CIE colour matching functions because the colour space is irregular, particularly with respect to saturation. On the other hand, a sensor colour space using four sensors equidistantly placed on the visual spectrum each with a narrowed spectral response is fully regular, both in respect of monochromatic stimuli and with respect to saturation. Figure 8 shows the sensor functions and Figure 5 shows that the colour space is regular and that it is regular with respect to saturation. The sensors are arranged into opponent pairs: Y-B and R-G. Monochromatic stimuli can be coded directly by opponent pair subtraction. If wavelength is coded relative to each sensor this has the side effect of inherently preserving the 'primary colour' of a monochromatic stimulus. The chromaticity space is regular through the entire sensor range and as a result the chromaticity space is circular. The chromaticity space is also entirely regular with respect to any level of white. The function to translate from chromaticity coordinate is simply a change from cartesian coordinates to polar coordinates.

A perceptual system that codes colour using a sensor design shown by Figure 8 may be referred to as *tetrachromatic*. It not only employs four sensors but codes colour in terms of four primaries; with all colours coded either as a primary or in transition between two adjacent primaries. A notable feature of this design is that it treats the visible spectrum as circular, which is reflected by a lack of a gap in the chromaticity diagram (Figure 5). Unlike the CIE chromaticity diagram which has a large 'purple gap' where monochromatic colours are undefined, the tetrachromatic chromaticity diagram is defined over the entire visible spectrum without any gap. It is consistent even directly on the meridional line where long wavelengths transition to short wavelengths. The chromaticity diagram is itself also circular, which means wavelength can be easily calculated from a chromaticity coordinate. Saturation is fully regular, so colour is simply a function of angle and distance to the central white point.

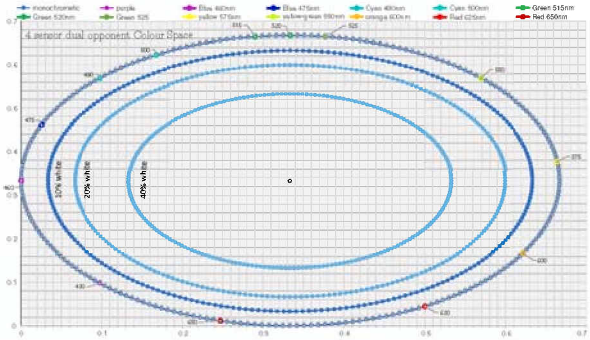

Figure 5 – Chromaticity space for opponent sensor functions in Figure 8, showing chromaticity coordinates for pure monochromatic light on the outer circle and then with white levels at 10%, 20% and 40%.

Dual Opponent Colour using Three Sensors

The key feature of using opponent sensors to measure the wavelength of monochromatic light is that one sensor of an opponent sensor pair is held constant while the second sensor of the opponent pair varies linearly by wavelength. This allows wavelength to be calculated by simple subtraction. When the colour filter that is responsible for this flat vs. linear spectrally narrowed response is removed, the sensors will have a much more broad spectral response and will no longer have an area of flat response but will instead respond linearly over the entire spectral range. Using these simple assumptions, it is possible to design a colour system using just three sensors, using only sensors that are closely related to the opponent sensors shown in Figure 8. This system will retain the property of having a chromaticity space capable of resolving all monochromatic light within the visual range of the four-sensor system. Figure 6 shows the sensor functions; with the red (X) and green (Y) sensor functions being sensor functions with a very broad linear response over most of the visual spectrum and with the blue (Z) function retaining most of the characteristics of the original opponent sensor. The red and green sensors have the interesting property of having an identical response, with one being simply a spectrally shifted version of the other.

If the chromaticity is calculated by taking the ratio, as with the CIE XYZ sensor system, this will produce the colour space shown in Figure 7. The resulting chromaticity space is regular and defined over the full range of wavelengths, but has a *square* shape. It is also fully regular with respect to saturation. Despite the sensor system having only three sensors, the chromaticity space is inherently tetrachromatic. Opponent coding can be preserved by simply taking note of the nearest *corner* of a chromaticity coordinate. A simple trigonometric transform allows the respective distance to a *corner* to be translated into the equivalent angle of the circular opponent colour space. The *corners*, therefore,

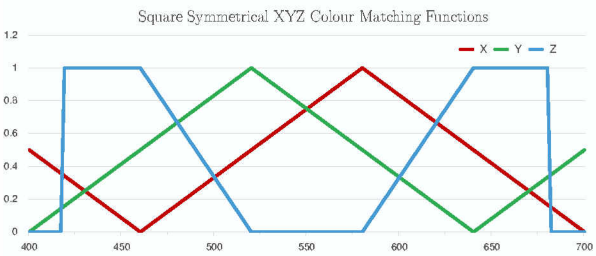

Figure 6 – Three colour matching functions (sensor functions) whose associated colour space has the properties of being complete (no 'extra-spectral' colours), computable and regular with respect to saturation.

preserve colour information. The four corners represent the primary colours and the relative distance between these primaries is the hue (or dominant wavelength) component of the colour, with the distance to the centre being the white (or *saturation*) component. The *square* chromaticity space may therefore be considered equivalent of the circular opponent chromaticity space.

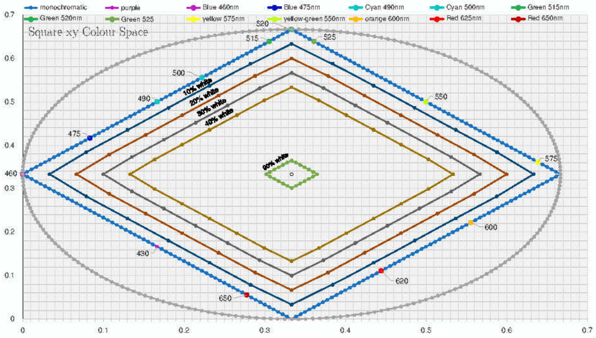

Figure 7 – The chromaticity space for the sensor functions shown in Figure 6. The chromaticity space is 'square' rather than circular. The translation from 'square' to circular requires the square root function.

A three-sensor system similar to that of the natural human sensor system can therefore be demonstrated as being equivalent to a four-sensor opponent system similar to that found in birds or reptiles. Moreover, the sensors of the three-sensor system can be derived from the four-sensor system by a sequence of simple modifications, the most important of these being the removal of the spectral filter. The difference between the two sensor systems is simply computational complexity, with both sensor systems being able to resolve monochromatic wavelength (and white level) with equal precision. This additional complexity, however, has an advantage in that it provides some degree of simplification to the calculation of white level when stimuli are in non-canonical form.

Natural Colours

The CIE XYZ colour matching functions, the dual opponent sensor functions shown in Figure 8 and the sensor functions shown in Figure 6 inherently map monochromatic light stimuli within the range 'visible' to the sensors to a chromaticity coordinate. Assuming that all stimuli are maximally saturated (that is, they are without any degree of white) coding colour is equal to coding wavelength. Natural stimuli, however, are rarely monochromatic. Figures 9 to 11 show the spectral power distribution of light reflected by a variety of flowering plants commonly found in a natural environment. Flowers are most commonly white or yellow, but variation is quite common and this is mostly between red and yellow. Figure 9 shows the spectral power distribution of variations of the colour yellow produced by a variety of sunflowers (genus *Helianthus*). Figure 10 shows the spectral power distribution of variations of the colour red produced by *Pelargonium*. Figure 11 shows the spectral power distribution of a variety of colours produced by the commonly cultivated rose (genus *Rosa*).

It may be observed that a rose or pelargonium flower reflects wavelengths of the visible spectrum precisely equally to produce the colour white. This will produce null chroma values for the CIE XYZ functions, the dual opponent functions and the square colour functions. The red of a rose or pelargonium flower on the other hand is at the extreme upper end of the visible spectrum and although the light reflected is not of a single wavelength, it will appear to be monochromatic to all three sensor functions due to most of the light reflected being out of the sensor range. The yellow of the sunflower, however, is not monochromatic, instead reflecting light very broadly; all the light from green region of the spectrum to the red region is reflected. It may be observed from the yellow sunflower and the yellow rose that light in the blue region of the spectrum is absorbed with considerable precision and that the light that is reflected is reflected with equal precision to white light. For the opponent sensors this means that blue will have a zero value because there is no light within its spectral range to activate the sensor, yellow will have a maximum value because nothing will be subtracted from its opponent pair, and the green and red pair will subtract to zero because they will be equally activated. For the dual opponent system, therefore, the yellow produced by a sunflower is equivalent to a monochromatic yellow.

For colours between yellow and red, the rose (Figure 11) shows that these colours are achieved by simply shifting the point between absorption and reflection within this spectral range. Figure 12 shows the chromaticity coordinates for the CIE XYZ colour sensors in the ideal case of shifting the absorption/reflection point from red to yellow and then to blue. It may be observed that as the point between absorption and reflection is shifted from red toward yellow that the chromaticity remains equivalent to monochromatic, and this continues until deep into the green wavelengths, but that there is a sharp boundary which when crossed leads to a sudden shift towards white. All colours above this boundary may be considered equivalent to monochromatic. Since flowers primarily produce colours within this range and as they are a very common source of colour in nature, it follows that a broad range of colours found in nature are equivalent to monochromatic colours. Visual sensor systems such as the dual opponent system and its three-sensor equivalent that are designed specifically for monochromatic light may therefore be usefully applied to a wide variety of natural colour stimuli.

Discussion

We have demonstrated that it is possible to emulate four sensor opponent chromaticity with just three sensors. These are artificial visual sensor systems specifically designed to resolve monochromatic light rather than to model how colour information is processed by natural visual systems. Nevertheless, the underly-

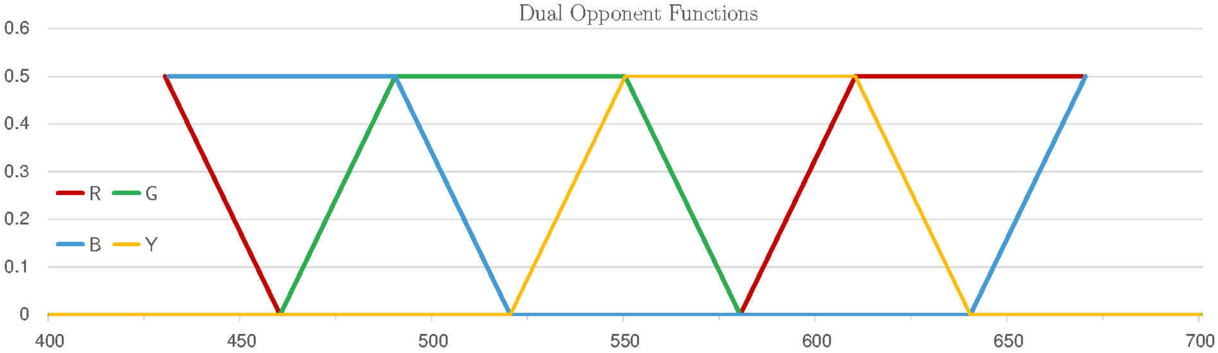

Figure 8 – Four sensor functions arranged into two opponent pairs; R-G and B-Y. The sensors can accurately code monochromatic stimuli in the range of 430nm to 670nm.

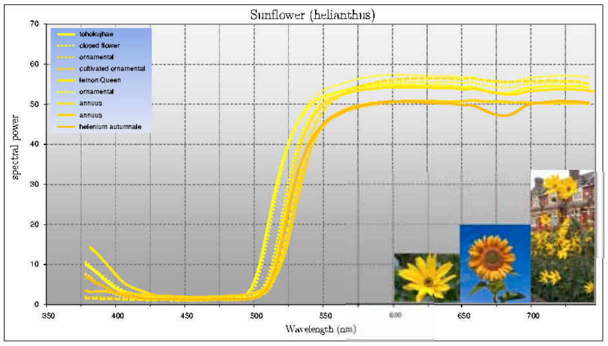

Figure 9 – Yellow colours from sunflower petal (genus *Helianthus*).

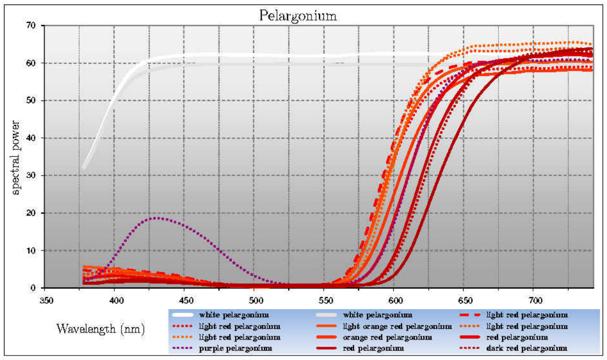

Figure 10 – Red colours from pelargonium flower petal (genus *Pelargonium*).

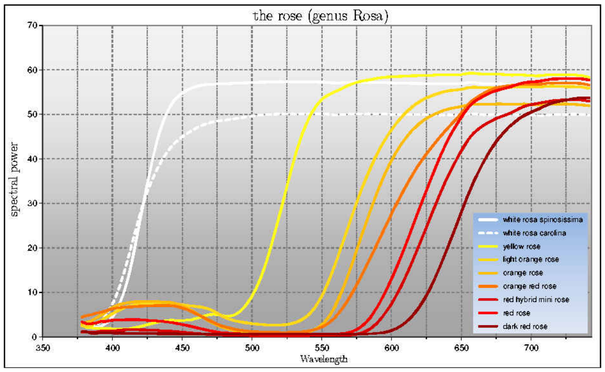

Figure 11 – Colours from red to yellow in the rose flower petal (genus *Rosa*).

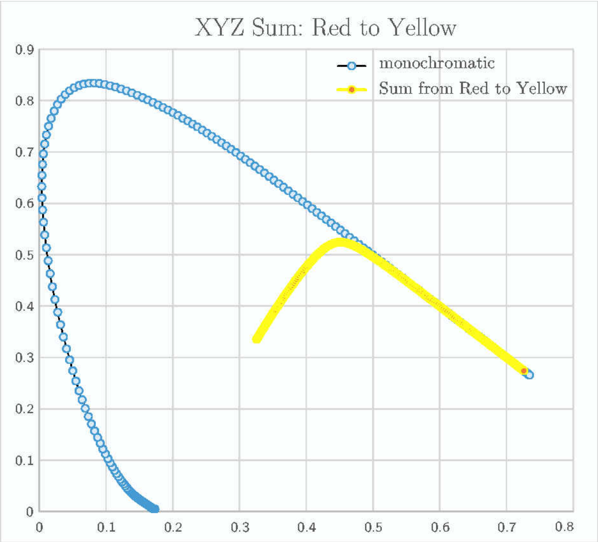

Figure 12 – The chromaticity coordinates when summing from red to yellow for the CIE XYZ colour matching functions. This is the idealized form for all the possible natural colours shown by Figures 9 to 11.

ing principles of these sensor systems are directly related to what is known about the sensor systems of natural visual systems. The four-sensor dual opponent system preserves the *colour* (that is, the relation to the nearest primary sensor) when coding spectral difference information in a way that is comparable to the primary colours of human perception. The four sensors of the opponent system also have the property of being equidistantly distributed on the visual spectrum and being spectrally restricted, which is similar to the spectral response of the visual sensors found in many species of birds and reptiles. The mammalian visual system is known to have been derived from a common distant ancestor that had similar colour vision to that of birds and reptiles, and indeed it has been shown that although mammalian colour vision is almost universally dichromatic with only a single channel for colour vision, that the two-channel colour architecture continues to remain active [6]. A single channel of colour information is insufficient for two dimensions of chromaticity, and therefore it may be assumed that colour information in dichromats is simply a degraded form of tetrachromacy. The addition of a third sensor, however, makes two-dimensional colour vision possible without having a full complement of two pairs of opponent sensors. This can be achieved by the using ratios of the three colour sensors, something which is very sensitive to variation and requires a precise response in the sensor functions. If the sensor functions are subject to natural variation then calibration would become an important factor.

With a four-sensor opponent system colour is essentially determined by the sensors themselves. A three-sensor system with the ability to calibrate its colour response has in effect virtualized the colour response that in simpler more direct visual systems is a product of the underlying hardware. The natural extension of a colour system that is able to calibrate dynamically is a colour system that is able to calibrate its primaries according to the perceptual conditions of the visual environment. Colour constancy would therefore not only be a function of the spectral composition of the light source (that is of a reference white) but also of an estimate of the primaries inherent to the specific visual environment.

A regular chromaticity space as shown in Figure 7 requires the sensor functions to be precisely equal to that shown in Figure 6. Even small differences can severely distort the colour space. If this solution is the only possible three-sensor solution to chromaticity that is regular both with respect to wavelength and with respect to white level, then a similarity to what is known about the sensors functions of the human visual system is significant. The response functions of the human visual sensors are known from the study of pigment absorption on the one hand or from a perceptual response to monochromatic light on the other. These are approximations which lack the mathematical precision with which the visual sensors may operate when in their natural environment. If it is assumed that the underlying mechanism of colour vision is to discern the wavelength of monochromatic light, then the evolutionary path from a four-sensor dual opponent system may be clearly demonstrated. If a sensor system such as the one shown in Figure 8 is degraded into a two-sensor single opponent system that has lost the ability to spectrally filter light and this system has to develop step-by-step into a two-channel opponent colour system with a regular chromaticity space then the three-sensor system shown in Figure 6 would be a solution to this problem. Such a solution would be suitable for an organism which has insufficient time to fully re-develop tetrachromacy with its sophisticated colour filters, but which had a sufficiently flexible visual information processing architecture that allows simple opponent subtraction to be re-purposed for the calculation of sensor sum ratios.

Conclusion

We have demonstrated that a visual sensor system that uses ratios of three sensors can be equivalent to a four-sensor system that uses opponency. Both the dual opponent sensor system and the three-sensor system used to demonstrate equivalency are artificial systems rather than precise models of natural visual systems. They are designed specifically to code the wavelength of monochromatic light rather than coding human colour perception, and they are fully regular through the range of wavelengths visible to their sensors. They are also fully regular for arbitrary mixtures of white light and monochromatic light, assuming the white light is equal-energy throughout the visual spectrum. Some of these sensor systems can also be usefully applied to a broad class of natural colours, specifically to the colour of flowers which range between yellow and red. Many flowers produce very saturated colours, and the underlying reason for this is that they are able to shift the point between reflection and absorption while at the same time being able to precisely absorb light and precisely reflecting light in those respective regions of the visible spectrum. When light is reflected precisely equally over a broad region of the spectrum, this is equivalent to the reflected light being monochromatic for an opponent system. It is possible, therefore, that a sensor system designed for monochromatic light could (possibly with a few *tweaks*) be applied more generally to the colours of the natural world. If it is possible to code the colours of nature as if they were monochromatic then this would suggest a basic template for how natural visual systems inherently code colour.

References

[1] Eleanor M. Caves et al, Categorical perception of colour signals in a songbird, *Nature*, 560, pg. 365–367 (2018).

[2] Timothy Goldsmith and Byron Butler, Color vision of the budgerigar (*Melopsittacus undulatus*): hue matches, tetrachromacy, and intensity discrimination, *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.*, 191, 10, pg. 933–951 (2005).

[3] Timothy H. Goldsmith, What birds see, *SciAm*, 295, 1, pg. 68–75 (2006).

[4] Leo M. Hurvich and Dorothea Jameson, An opponent-process theory of color vision, *Psychological Review*, 64, 6, Part 1, pg. 384–404 (1957).

[5] Peter Lennie, Color vision: Putting it together, *Current Biology*, 10, 16, pg. R589–R591 (2000).

[6] Jay Neitz, Joseph Carroll, and Maureen Neitz, Color vision: Almost reason enough for having eyes, *Optics Photonics News*, 12, 1, pg. 26–33 (2001).

[7] Samuel G. Solomon and Peter Lennie, The machinery of colour vision, *Nat. Rev. Neurosci.*, 8, 4, pg. 276–286 (2007).

[8] Misha Vorobyev, Evolution of colour vision: The story of lost visual pigments, *Perception*, 35, 168 (2006).

Author Biography

Thomas Bangert received his BSc in computer science from the University of Toronto (1988) and his MSc in Electrical Engineering from Queen Mary University of London (2012). He has been doing basic research in colour theory for about a decade, and before this he worked on developing the fundamentals of simplified but fully functional artificial visual systems and developing an understanding of neurons as units of computation. He is a member of IS&T, ICVS and OSA.