Article

Demands of Geometry on Color Vision

Abbreviated title: Geometry and Color Vision

Dale Purves 1,2,3,* and Chidambaram Yegappan 1

- Neuroscience and Behavioral Disorders Program, Duke-NUS Graduate Medical School Singapore, 8 College Road, Singapore 169857, Singapore; chidam.yegappan@gmail.com
- Department of Neurobiology, Research Drive, Duke University Medical center, Durham, NC 27710, USA
- ³ Duke Institute for Brain Sciences, Duke University, Durham, NC 27708, USA
- * Correspondence: purves@neuro.duke.edu; Tel.: +1-919-681-7414 or +1-919-260-4734

Abstract: The reasons for the circular sense of human color perception generated by two sorts of color opponent neurons and three cone types are not well understood. Here we use geometrical analysis to examine the hypothesis that opponency, the recursive nature of color perception, and trichromacy arise as the most efficient ways of distinguishing spectrally different points on a plane using a minimum of color classes and receptor types.

Keywords: trichromacy; opponency; color circularity; spectral images; unique colors; four-color map problem; perception

Introduction

Distinguishing retinal image regions to promote apt behavior is the broad purpose of animal vision. Whereas achromatic vision distinguishes regions on the basis of light intensity (luminance), color vision further distinguishes equiluminant regions on the basis of spectral differences, giving animals with this ability an additional behavioral advantage¹⁻³. Despite a wealth of anatomical, electrophysiological and psychophysical evidence about color vision, a puzzle since Newton's pioneering studies is why we arrange objects with minimal spectral differences among them (e.g. Munsell chips) in a closed continuum (Figure 1)^{4,5}. Equally perplexing is why we see a color gamut based on four color classes—reds, greens, blues and yellows—each defined by a particular hue that has no apparent admixture of the other three color classes⁶⁻⁹. It has long been known that these perceptual phenomena are initiated by the different spectral sensitivities of short, medium-, and long-wavelength cone types, and that the output of these receptors is processed by opponent neurons that pit color class pairs—mostly red-green and blue-yellow—against each other^{7,8,10-12} (see also ref. 31). Although trichromacy and opponency are well documented, their rationale is also uncertain^{7,11,13}.

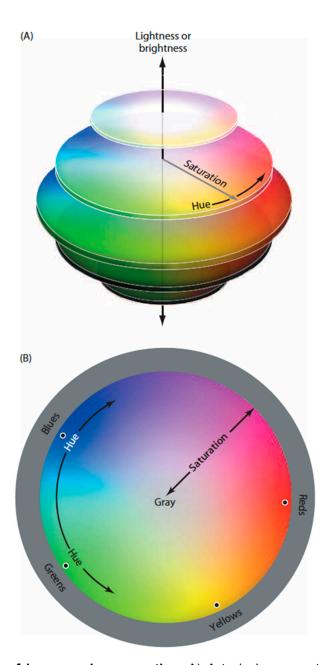


Figure 1. Diagram of human color perception. A) A typical representation of human "color space". B) A single cross-sectional plane from the diagram in (A). When asked to arrange a large number of equiluminant surfaces that vary in hue such that the color appearances among them are minimal, subjects arrange them in a closed loop that comprises four basic color categories (red, green, blue and yellow), each defined by a hue (black dots) that has no apparent admixture of the neighboring color classes (e.g., a red surface seen as having no appreciable yellowness or blueness). Although there are many other named color groups (oranges, aquamarines etc.), these are seen as mixtures of the two of the four primaries. Note that the locations of these unmixed hues (black dots) are only approximate, since their actual position in geometrical or psychophysical terms depends on the number of tiles used in such testing and cannot be specified precisely; nor is the closed loop determined psychophysically a circle.

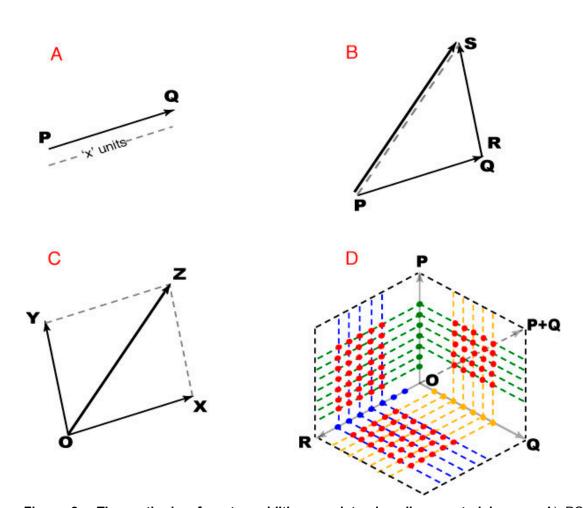
Although the retina is a curved surface, the information it captures (the retinal image) is generally thought of as two-dimensional (2D). The challenge of regional differentiation in 2D is called the "four-color map problem," and refers to the empirical minimum of four colors needed to ensure that adjoining regions of a plane or the surface of the sphere are never the same if they are geographically or otherwise different¹⁴⁻¹⁶. The relevance of this problem to color vision is that at least four perceived color classes are needed to resolve the same concern in the retinal image plane¹⁷.

Here we examine whether the color opponency, perceptual recursion ("color circularity") and trichromacy are related consequences of the need to distinguish all possible equiluminant regions in the retinal image plane by different color sensations, at the same time ensuring that no two regions of an image appear the same if their spectra illumination is different.

Methods

As illustrated on Figure 2, we use vectors to describe the geometry of spectral images. A plane considered in this way is a vector space, and any point in that space can be specified the position vector of that point with respect to an origin (Figure 2A). The position vector of the origin is a null vector whose initial and terminal points are coincident, with zero magnitude and no direction. The argument also depends on vector addition. If PQ and RS are any two non-zero vectors, their addition leads to a third vector PS that is different from both PQ and RS (Figure 2B). The magnitude and direction of PS depends on the magnitude and direction of vectors PQ and RS. The addition of vectors with PQ and/or RS as null vectors is still valid, but the resultant vector PS may be equal to both or one of the addends.





The methods of vector addition used to describe spectral images. A) PQ represents a vector directed from point P to point Q. The points P and Q are referred to as the initial and terminal points of the vector. Here, magnitude of PQ = 'x' units acting in the indicated direction. B) The tip-to-tail method vector addition, showing how PS is generated from vectors PQ and RS. C) Addition of vectors OX and OY using the parallelogram law of vector addition. D) A 2D vector space with point O as origin. The green, yellow, and blue dots represent the end points (position vectors) of the direction vectors P, Q and, respectively. The red dots end points of position vectors determined as vector sums of the position vectors corresponding dotted lines.

When adding vectors, the initial point of the second vector was placed at the terminal point of the first, resulting in a vector co-initial with the first vector and co-terminus with the second vector. Since vector addition is commutative, either of the addends can be treated as the first vector. For addition involving more than two vectors, stepwise addition was performed by taking two vectors at a time, with the resultant vector of a pair being one of the addends for the subsequent addition. Finally, when two vectors are adjacent sides of a parallelogram, their vector sum is given by the diagonal of the parallelogram (Figure 2C). This parallelogram rule gives the same result as the tip-to-tail method, except that it operates on co-initial addends, and was used to define position vectors based on other position vectors (Figure 2D). When two direction vectors are represented by adjacent sides of a parallelogram, the included

position vectors defined all the position vectors within that parallelogram (i.e., the black dotted lines connecting all the direction vectors in Figure 2D).

Results

Regional distinctions in one dimension

Rationale. To appreciate the geometrical requirements that must be met to distinguish all possible regions in planar images by color sensations, consider the simpler challenge of distinguishing all possible regions in a hypothetical onedimensional image.

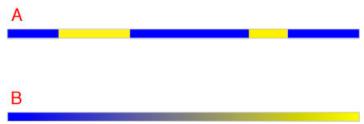


Figure 3. Distinguishing neighboring regions in one dimension. A) Two different qualities (blue and yellow in this example) suffice to distinguish any number of neighboring regions from one another along a line. B) The continuum needed when points must distinguished according to relative spectral differences at equiluminance. Notice that a gray-scale continuum would differ in that its balance point is an average of the two extremes, whereas the balance point in a color continuum such as this is neither blue nor yellow.

From a geometrical perspective, two different qualities are all that is needed to distinguish any number of regions from their neighbors in a linear image (Figure 3A). In color vision, however, the goal in this hypothetical one-dimensional (1D) scenario would be to distinguish equiluminant regions by assigning different color sensations to different spectra. Since the number of different spectra at any given level of luminance is limited only by the ability of the visual system to distinguish spectral variations, color vision would evolve as many equiluminant sensations as this constraint allowed (see Figure 1B). A corollary is that perceived differences among colors at equiluminance should accord with the relative differences among the activation of photoreceptors. If points (i.e., the smallest resolvable regions) were arranged according to minimal differences among spectra, the result would be a continuum of equiluminant color qualities (Figure

The perceived difference between the extremes of this 1D space would be maximal, with a colorless intermediate sensation that was equally different from the sensations elicited by the extremes. The continuum thus represents gradual spectral deviation in two opposite directions from a colorless balance point, with the maximum deviation eliciting the sensations of unique hues, e.g. a blue with no yellow and vice versa. Thus two color classes defined by two unique hues is geometrically sufficient to distinguish all the points comprising a one dimensional image.

Quantification. Just as a line that extends from a null point in two directions can be specified by a pair of opposing direction vectors (e.g., the x and -x vectors of

a graphical axis), the spectral continuum in Figure 3B can be described by vectors (see Methods). The direction of each position vector from the null vector indicates the dominant color class at that location, and the distance from null vector indicates the relative contributions of the two unique hues. As the extremes of opposing direction vectors, the unique hues are necessarily opponents. The corresponding direction vectors indicate spectral deviation from the uniform distribution of energy at the color neutral null vector, and the distance between any two position vectors indicate the perceived spectral difference between them.

Biological significance. Considered in terms of photoreceptors, two opposing cone types with overlapping spectral sensitivities could generate these distinctions (Figure 4), thus creating the opposing direction vectors in Figure 3B. The result would lead to the perceptual experience of human dichromats, and other mammals with only two types of cones¹⁸. Thus individuals who lack L cones (protanopes) see spectral energy distributions that would have generated red or green for trichromats as shifts in brightness and/or saturation based on blue and yellow color classes, with a reduced color gamut. Similar findings have been reported for deuteranopes and tritanopes¹⁹.

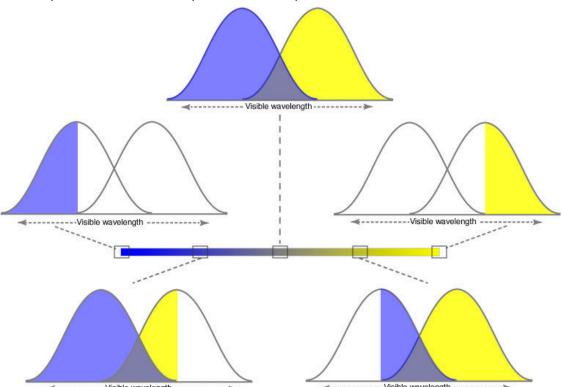


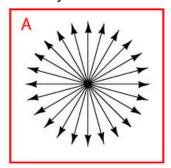
Figure 4. Color specification in 1D achieved by two opposing photoreceptor types. Top panel. Hypothetical cone sensitivity curves. When stimulated equally, the result would be perceived color neutrality. Middle panels. When the response of one of the photoreceptor types is maximal and the other minimal, the perception would be that of unique blue (left) or unique yellow (right). Bottom panel. Stimulation that would give rise to predominantly blue (left) or yellow (right) perceptions along color continuum in Figure 3B.

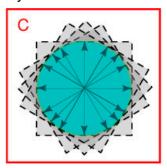
Preprints (www.preprints.org)

7 of 14

Making regional distinctions in two dimensions

As in a hypothetical 1D image, equiluminant color sensations arising from a plane can be represented by direction and position vectors. However, whereas 1D space extends in only two directions, 2D space extends in all possible directions from a null point (Figure 5A). Accordingly, the position vectors would be expressed by two pairs of opposing direction vectors, as in the (-x, x) and (-y, y) axes that define any Cartesian co-ordinate system.





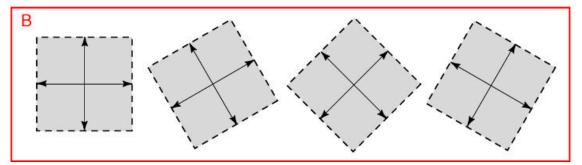


Figure 5. Distinguishing regions in a two dimensional image. A) A two-dimensional vector space extends in all directions from its null point. B) Vector addition based on any two axes in the plane defines position vectors in all directions (four of the many possible arrangements are shown as examples). C) The position vectors held in common by the full set of direction vectors are bounded by a closed loop, as indicated by the area in green.

In vector addition (see Methods), two pairs of opposing direction vectors define the position vectors in all possible directions (Figure 5B). Of the many possible combinations, any two pairs of opposing direction vectors can be used to define the position vectors in all directions. The vector space defined by all possible sets of opposing direction vector pairs thus forms a closed loop that bounds the portion of the plane held in common (Figure 5C). Although not shown, the argument is the same whether or not the two pairs of opposing direction vectors are orthogonal.

Rationale for color opponency. Given that the space held in common by all sets of opposing direction vector pairs in Figure 5C is bounded by a recursive perimeter, it follows that distinguishing points on a plane according to relative spectral differences obeys the same rules. In a two-dimensional (2D) space. however, the evolution of color vision must also contend with the four-color map problem as it pertains to spectral returns (Figure 6A). At the same time, color

vision must be able to identify all spectrally different equiluminant points in a 2D image. These dual goals can be achieved by two pairs of direction vectors that give rise to four color classes defined by four unique hues that are pair-wise opponents (Figure 6B).

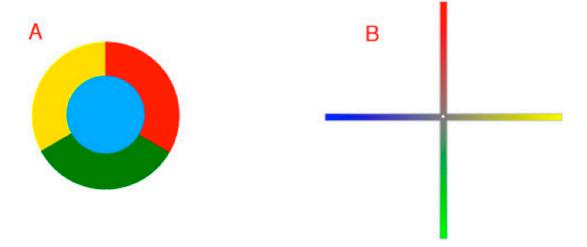


Figure 6. The four-color map problem and the color opponent axes needed to contend with it. A) The four-color map problem in 2D geometry refers to the fact that abutting regions on a plane cannot be distinguished using fewer than four colors (or other designators) of surface differences. The diagram shows an example in which the indicated regions could not be differentiated using fewer than four colors. If this patterns was an equiluminant retinal image, the four regions would have to elicit color sensations that could distinguish any possible hue within these four color categories (reds, blues, greens and yellows). B) Two opposing color quality pairs (direction vectors) extending from a null point (white dot) would be sufficient to address this issue. at the same time as differentiating all points (position vectors) on a plane by a gamut of color sensations.

Rationale for color perception as a closed loop. The rules of 2D geometry also explain how ordering spectrally distinct points on an equiluminant plane according to minimal spectral differences leads to a the "circle" of hue sensations illustrated in Figure 1. Since the perimeter of a 2D space that extends equally in all directions is recursive (see Figure 5), the equiluminant sensations associated with the position vectors that lie between the opposing direction vectors in Figure 6B will define a closed loop, with spectrally distinct points defined by vector addition based on the four representative spectral directions. Thus only four color classes are needed, although the number of perceived and named hues within each class can be very large.

Rationale for retinal trichromacy. This geometrical evidence does not, however, explain how trichromacy fits contributes to addressing the geometrical demands of color vision. On the contrary, from the argument so far one would expect that four cone types representing the two opponent axis pairs would be needed to generate the four color classes illustrated in Figure 6B. How, then, could the three cone types in humans and Old World monkeys generate the four color classes needed to distinguish all the equiluminant points on a plane and contend

with the four-color map problem (see Figure 6A)?

As shown in Figure 7A, the addition of any two non-opposing direction vectors defines all position vectors in all the directions between them (see Figure 1D). Accordingly, three direction vectors can, in principle, define all the position vectors on a plane. To do so, however, two of the three direction vectors would have be non-opposing, and must not define the third direction vector by vector addition. As shown in Figure 7B, to satisfy these constraints the third vector must lie within the area between the dashed blue lines.

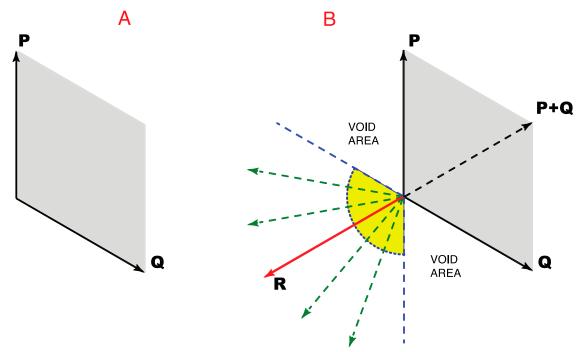
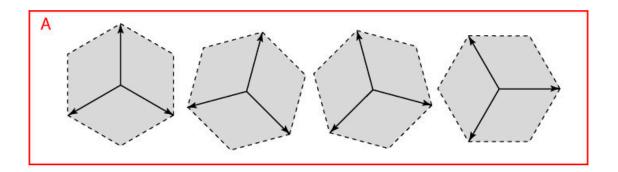


Figure 7. How three direction vectors can specify all possible locations (position vectors) on a plane. A) Two non-opposing direction vectors (P and Q) can, by vector addition, specify all the position vectors that lie between them (the gray parallelogram). B) A third direction vector, if appropriately positioned, allows the resulting triad of vectors (P, Q and R) to specify locations in all directions. To deal with the four-color map problem, however, the third direction vector must oppose the combined influence of other two direction vectors, i.e. vector P+Q indicated by the dashed black line.

Moreover, unless the third direction vector (vector R in Figure 7B) has an specific direction with respect to the two other vectors, only three color classes would arise, an arrangement that would be unable to contend with the four-color map problem illustrated in Figure 6A. Thus the third direction vector R must oppose the combined influence of the other two direction vectors (direction vector P+Q). An exception is when the influence of vectors P and Q are equal. In this case, they would generate a null vector when vector R equals the combined influence of P and Q, much as the opposing direction vectors in Figure 6B. Thus P and Q are opponents only when direction vector R opposes the vector sum P+Q. In this way the second opponent pair needed to resolve the four-color map problem is provided.



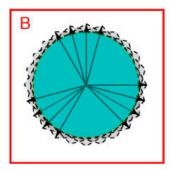


Figure 8. A closed 2D space defined by three direction vectors. A) Vector addition based on three direction vectors placed appropriately (see Figure 7) defines position vectors in all directions on a plane (only four of many possible arrangements are shown as examples). B) The position vectors held in common by the full set of such direction vectors form a circular space in this example (green area).

Each unique hue corresponds to the greatest possible deviation of spectral energy from the null vector of these directional extremes, i.e. the extremes of P, Q, R and P+Q in Figure 7. Intermediate colors would arise from the graded distribution of spectral energy between these extremes, as indicated by the gradients in Figure 6B. As shown in Figure 8, the position vectors that lie between the three direction vectors in Figure 7B define a common closed space (cf. Figures 5C and 1). Note further that the interaction of the two pairs of opposing direction vectors in Figure 7 (P vs. Q, and R vs. P+Q) is analogous to interaction of L vs. M cones, and S vs. L+M cones.

In sum, the three cone types in humans and Old World monkeys are the minimum number needed to generate the four color classes capable of distinguishing all resolvable equiluminant regions on a plane, at the same time resolving the four-color map problem.

Discussion

The arguments presented suggest that color opponency, the closed loop nature of color perception, and the evolution three retinal cones types efficiently meet these related geometrical challenges, providing a unified rationale for perceptual and physiological phenomena that are otherwise difficult to explain.

Why red, green, blue and yellow are perceptual primaries

The results leave unexplained, however, why the four primary color directions are

red, green, yellow and blue, or, more accurately, why visible spectra elicit these hue sensations. Although the geometrical framework we outline does not address this question, it does suggest an explanation. Just as the bias of magnetic north fixes the cardinal axes in geography, there are presumably biases in human spectral experience that determine the three direction vectors P, Q and R in Figure 7. The most obvious candidate for such biases is the distribution of light coming from natural surfaces. Indeed, the reflectance efficiency functions of natural surfaces give rise to three spectral groups: foliage in the yellow-green region (557-574 nm); "earths" in the yellow-orange region (576-589 nm); and water, sky and distant objects in the blue region (459-486 nm)²⁰. These biases accord with the peak sensitivities of the three cone types, which in turn accord with the argument in Figure 7. This agreement, however, still leaves the question of why the combined influence of P and Q opposes R, rather than the other options (i.e., Q and R opposing P, or P and R opposing Q). One possibility is a bias arising from the greater overlap of the spectral ranges of foliage and earth than the overlap of either of these ranges with the blue range. It may also worth pointing out while that some primitive cultures lack names for these four color categories²¹, this deficiency in vocabulary is unlikely to arise from differences in visual physiology or perception.

Why the black-white axis cannot replace an opponent color axis

Vision depends primarily on the distinguishing light intensities (luminance); indeed, many animals lack color vision^{2,18}. Since the perception of grays ranging from unique black to unique white forms an opponent axis, the question arises why the gray scale axis could not interact with a single color axis to distinguish all possible spectral points on a plane, at the same time resolving the four-color map problem.

The reason is that the color and gray scale axes concern different categories of information—light intensity versus the distribution of spectral power—which are independent and lead to different sensory qualities (lightness versus hue) and different biological advantages. These qualities can influence each other^{7,22}, but like sensory qualities in other systems (e.g., pitch and loudness) are not combined in perception.

Other rationales

Several plausible rationales for color opponency and trichromacy have been proposed. Color opponency, which has been validated both psychophysically¹² and electrophysiologically²³⁻²⁴at several levels of the primary visual pathway in experimental animals has been suggested to enhance the encoding of natural scene spectra²⁵, and/or to optimize the transfer of trichomatic color information^{26,27}. By analogy with spatial sinusoids, trichromacy has been interpreted in terms of comb-filtered spectra²⁸.

A geometrical basis of color phenomenology, in contrast, is predicated on the behavioral advantages of distinguishing equiluminant image points. From this perspective, whatever else they may accomplish opponency, circularity and

trichromacy are all manifestations of resolving the geometrical problems that must be addressed in the evolution of effective color vision.

The perceptual experience of human dichromats

Another concern is the partial conflation of trichromatic and dichromatic color experience in psychophysical testing: human dichromats retain some ability to perceive and name colors that would be expected only in trichromats. Protanopes and deuteranopes show adaption to long wavelength light²⁹, and use the terms 'red' and 'green' as well as 'blue' and 'yellow' to describe what they see³⁰. Montag and Boynton³¹ have shown further that rods are likely contributors to these abilities, as well as to cues that might be derived from light intensities. These further facts about color perception are not explained by the arguments here.

Color metamers

A final concern is metameric stimuli, which refer to different distributions of spectral power that elicit the same perceived colors in psychophysical matching tests. Such stimuli arise because the overlapping sensitivities the three human cone types allow their similar relative activation by spectrally different stimuli. Since metameric image regions would elicit the same color percepts, metamers belie the idea that all spectrally different regions in an arbitrarily complex image can be distinguished. Although metameric matches are evident in colorimetry experiments, such stimuli are rare in natural images³². Thus the evolution of color vision may simply have ignored the problem posed by metamers.

Conclusion

To be optimally successful, vision must assign distinct color percepts to all possible arrangements of spectrally different regions on a plane while contending with the four-color map problem as it pertains to color vision. Here we argue that the "circularity" of color perception, color opponency, and trichromacy evolved as related ways of meeting these geometrical challenges, providing a unified basis for perceptual and physiological phenomena that are otherwise difficult to explain.

Acknowledgments: We are grateful to Cherlyn Ng, Yaniv Morgenstern, Brian Monson, Janani Sundararajan and Rukmini Dhara for helpful criticisms.

References

- 1. Land, M.F. and Nilsson, D-E., Animal Eyes, Oxford University Press, Oxford (2002).
- 2. Mollon, J. D. (1991). Uses and evolutionary origins of primate color vision. In J. R. Cronly-Dillon & R. L. Gregory (Eds.),

Evolution of the eye and visual system (Vision and visual dysfunction, Vol. 2). Boca Raton, FL: CRC Press.

- 3. Frome, F. S., Buck, S. L., & Boynton, R. M. (1981). Visibility of borders: Separate and combined effects of color differences, luminance contrast, and luminance level. Journal of the Optical Society of America, 71, 145-150.
- 4. Newton I (1704/1952). Opticks: or, a treatise of the reflexions, refractions, inflexions and colors of light. (Based on the 4th London edition). New York: Dover.
- 5. Munsell, A.H., A Color Notation, G.H. Ellis Co., Boston (1905).
- 6. Evans, R., An introduction to color, Wiley New York, NY (1948).
- 7. Hurvich, L. (1981). *Color vision*, Sunderland, MA: Sinauer Associates.
- 8. Wandell, B. (1995). Foundations of vision. Sunderland, MA: Sinauer Associates.
- 9. Kuehni, R. G., & Schwarz, A. (2008). Color Ordered: A Survey of Color Systems from Antiquity to the Present. Oxford University Press, USA.
- 10. Macaluso C, Lamedica A, Baratta G, Cordella M. (1996) Color discrimination along the cardinal chromatic axes with VECPs as an index of function of the parvocellular pathway. Correspondence of intersubject and axis variations to psychophysics. Electroencephalogr Clin Neurophysiol. 100:12-17.
- 11. Kaiser, P., & Boynton, R. (1996). Human color vision, 2nd
- edn. Washington, DC: Optical Society of America.
- 12. Hurvich, L. M., & Jameson, D. (1957). An opponent-process theory of color vision. Psychological Review, 64, 384–108.
- 13. Hubel, D. H. (1988). Eye, brain, and vision (Scientific American Library). New York.
- 14. Cayley, A. (1878). On the colouring of maps. Proceedings of the London Mathematical Society, 9, 148.
- 15. Ore, O. (1967). The four-color problem (Vol. 27). Academic Press.
- 16. Appel, K., & Haken, W. (1977). The solution of the four-color- map problem. Scientific American, 237(4), 108–121.
- 17. Purves, D., Lotto, B., & Polger, T. (2000). Color vision and the four-color-map problem. Journal of Cognitive Neuroscience, 12(2), 233-237.
- 18. Neumeyer, C. (1991). Evolution of color vision. In: Uses and evolutionary origins of primate color vision. In J. R. Cronly- Dillon & R. L. Gregory (Eds.), Evolution of the eye and visual system (Vision and visual dysfunction, Vol. 2) (pp. 284-305). Boca Raton, FL: CRC Press.
- 19. Brettel, H., Viénot, F., & Mollon, J. D. (1997). Computerized simulation of color appearance for dichromats. JOSA A, 14(10), 2647-2655.
- 20. Hendley, C. D., & Hecht, S. (1949). The colors of natural objects and terrains, and their relation to visual color deficiency. JOSA, 39(10), 870-872.

- 21. Berlin, Brent & Kay, Paul (1969). Basic Color Terms: Their Universality and Evolution. Berkeley: University of California Press.
- 22. Helmholtz, H. L. F. v.(1924) Helmholtz's Treatise on Physiological Optics. Opt. Soc. Am., Rochester, NY.
- 23. Chatterjee, S., & Callaway, E. M. (2003). Parallel colour-opponent pathways to primary visual cortex. Nature, 426(6967), 668-671.
- 24. Shapley, R., & Hawken, M. (2002). Neural mechanisms for color perception in the primary visual cortex. Current Opinion in Neurobiology, 12(4), 426-432.
- 25. Schluppeck, D., & Engel, S. A. (2002). Color opponent neurons in V1: a review and model reconciling results from imaging and single-unit recording. Journal of Vision, 2(6).
- 26. Lee, T. W., Wachtler, T., & Sejnowski, T. J. (2002). Color opponency is an efficient representation of spectral properties in natural scenes. Vision Research, *42*(17), 2095.
- 27. Buchsbaum, G., & Gottschalk, A. (1983). Trichromacy, opponent colours coding and optimum colour information transmission in the retina. Proceedings of the Royal society of London. Series B. Biological sciences, 220(1218), 89-113.
- 28. Barlow, H. B. (1982). What causes trichromacy? A theoretical analysis using comb-filtered spectra. Vision Research, 22(6), 635-643.
- 29. Scheibner HMO and Boynton RM (1968) Resildual red-green discrimination in dichromats. J Opt Soc Amer 58:1151-1158
- 30. Wachtler T, Dohrmnn U and Ertel R (2004). Modeling color percepts of dichromats. Vision Res 44: 2843-2855.
- 31. Montag ED and Boynton RM (1987) Rod influence in dichromatic surface color perception. Vision Res 27:2153-2162.
- 32. Foster DH, Amano K, Nascimento SMC and Foster MJ (2006) Frequency of metamerism in natural scenes. J Opt Soc Am A 23: 2359-2372.



© 2016 by the authors; licensee *Preprints*, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).