

THE EVOLUTION OF TRICHROMACY

An essay to mark the bicentennial of Thomas Young's graduation in
Göttingen

J. D. Mollon

Department of Experimental Psychology, Downing St.,
Cambridge CB2 3EB, United Kingdom

1996 marks the bicentennial of Thomas Young's graduation as a medical student in Göttingen and it was probably in Göttingen that Young first gave detailed thought to the nature of colour. For we know, from his own account (Peacock, 1855; see also Figure 1), that he attended the lectures on physics given by G. C. Lichtenberg and we know from a later-published transcript of those lectures (Gamauf, 1811) that Lichtenberg discussed in detail the colour triangle of Göttingen's Tobias Mayer. Young found most of the Göttingen professors formal and distant, but he remarks that "*Arneman, in whose house I live, and Lichtenberg the lecturer on Natural Philosophy, are the most sociable*" (Peacock, 1895).

Thus from Lichtenberg's lectures of 1795-6, Thomas Young would have necessarily been acquainted with the fact of trichromacy, the fact that all colours can be produced by quantitative colour mixing with three variables. His own contribution was to grasp that trichromacy has its basis not in physics but in the physiology of the human retina (Young, 1802, 1807). It is instructive that Young did few experiments himself. In a biographical memoir, his old friend Hudson Gurney wrote: "*But he was afterwards accustomed to say, that at no period of his life was he particularly fond of repeating experiments, or even of very frequently attempting to originate new ones; considering that, however necessary to the advancement of science, they demanded a great sacrifice of time, and that when the fact was once established, that time was better employed in considering the purposes to which it might be applied, or the principles which it might tend to elucidate*" (Gurney, 1831); and on one occasion, opposing government expenditure on experimental science, Thomas Young wrote: "*...it is my pride and pleasure, as*

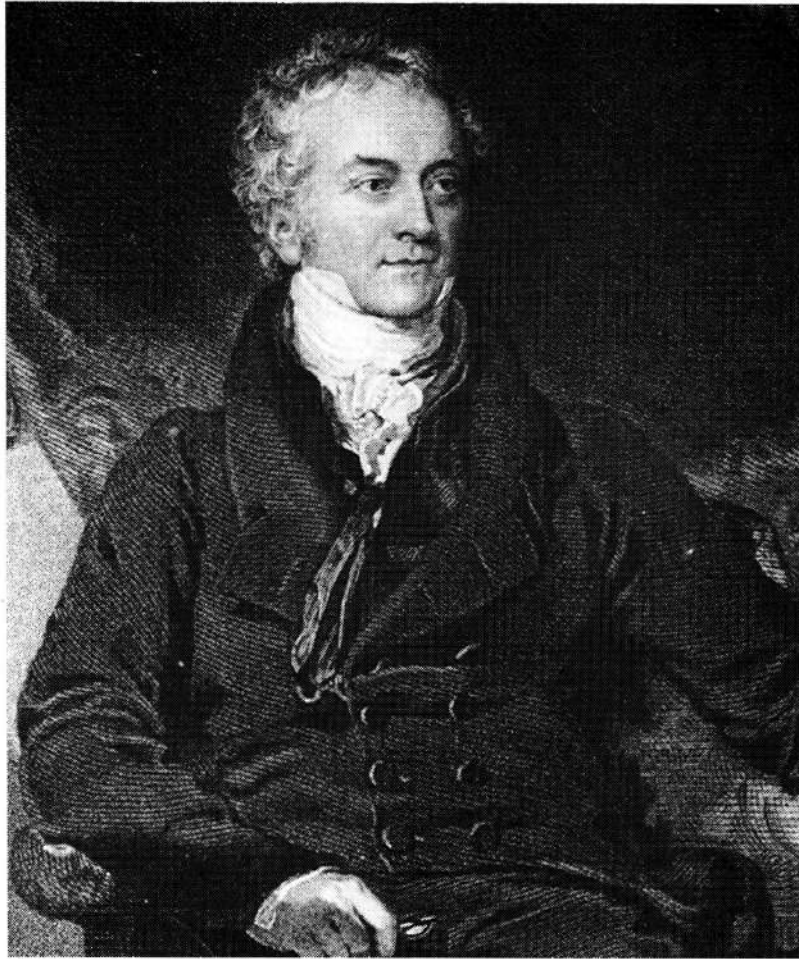
far as I am able, to supercede the necessity of experiments, and more especially of expensive ones". Young's role in the history of colour science was to resolve the contradiction apparent in the eighteenth-century literature: on the one hand there was evidence that the physical variable underlying hue was a continuous one, and on the other, there was evidence that all, or nearly all, colours could be produced by mixing three primaries. Young's insight was to realize that the trichromatic limitation was imposed by the human retina.

We now know that Young's three physiological resonators are different types of retinal cone, which can be shown by microspectrophotometry have their peak sensitivities in the violet (420 nm), the green (530 nm) and the yellow-green (560 nm) regions of the spectrum (Dartnall, Bowmaker and Mollon, 1983). The rich gamut of our hue sensations depends ultimately on a neural comparison of the rates of quantum catch in the three classes of cone.

In Thomas Young's account, and in most modern textbooks, the three cones are treated as equal components of a trichromatic system. In fact, I argue here that the different cones evolved at different times for different purposes. To understand the present state of human and primate colour vision we must consider how it came to its present state. For an earlier account of the argument in German, see Mollon and Jordan (1988).

The opsins and their genes

To the right in Figure 2 is depicted the outer segment of a cone, its multiply enfolded membranes packed with molecules of photopigment. Photopigments consist of proteins, some 356 amino acids long, bound to 11-cis-retinal, a derivative of Vitamin A1. The protein components of the pigments are called opsins and they each consist of seven helices that cross the cell membrane and are linked by loops within and without the membrane (Hargrave et al., 1983). The seven helices form a palisade that surrounds the 11-cis-retinal; and the latter is bound to a lysine in the seventh helix. The palisade is splayed, so that helices are closer together at the cytoplasmic surface, the side within the cell (Baldwin, 1993).



Thomas Young (1773–1829)

- “ At 8, I attend Spittler’s course on the History of the Principal States of Europe, exclusive of Germany.
- “ At 9, Arnemann on *Materia Medica*.
- “ At 10, Richter on Acute Diseases.
- “ At 11, Twice a week, private lessons from Blessman, the academical dancing-master.
- “ At 12, I dine at Ruhlander’s table d’hôte.
- “ At 1, Twice a week, lessons on the Clavichord from Forkel ; and twice a week at home, from Fiorillo on Drawing.
- “ At 2, Lichtenberg on Physics.
- “ At 3, I ride in the academical manège, under the instructions of Ayrer, four times a week.
- “ At 4, Stromeyer on Diseases.
- “ At 5, Blumenbach on Natural History.
- “ At 6, Twice Blessmen with other pupils, and twice Forkel.
- “ Spittler, Arnemann and Blumenbach, follow, in lecturing, their own compendiums, and Lichtenberg makes use of Erxleben’s. I mean to study regularly beforehand.”

Fig. 1 Thomas Young’s own account of his working day as a student in Göttingen in the academical year 1795-96. He graduated Doctor of Physic on 16 July, 1796.

The opsins gain a wider interest from the fact that they are members of the superfamily of G-protein coupled receptors or *heptahelicals* (Mollon, 1991a). This family has more than 200 members and includes many of the receptors that mediate the intercourse between our brain cells and between hormones and neurons. The several forms of the serotonergic, dopaminergic, adrenergic and muscarinic acetylcholine molecules, as well as the large subfamily of olfactory receptors, are all heptahelicals, distant cousins of the opsins (Watson and Arkininstall, 1994; Firestein, 1991).

Our modern knowledge of the opsins, and of their evolution, depends very much on the work of Jeremy Nathans and his collaborators, who sequenced the opsin genes a decade ago (Nathans et al., 1986 a,b). The genes that code for rhodopsin and for the short-wave cone pigment lie on chromosomes 3 and 7 respectively. As was predicted from the classical evidence for X-linkage of red-green colour deficiencies, the genes for the long- (L) and middle-wave (M) cone pigments lie close together on the X-chromosome. Typically the X-chromosome carries more than one M gene, although the total number is controversial.

Figure 2 identifies seven amino-acid sites that are known to control the spectral sensitivity of the primate L and M pigments. Our knowledge of these critical sites has come from comparisons of genotype and phenotype in polymorphic species of New World monkey (Neitz, Neitz and Jacobs, 1981; Williams et al., 1992) and from the *in vitro* expression of natural and synthetic opsin genes (Merbs and Nathans, 1992; Asenjo, Rim and Oprian, 1994). Of the seven sites, the two most important are numbers 277 and 285 in the sixth helix of the molecule.

The ancient mammalian sub-system of colour vision

The high homology of the L and M opsin genes, and their close juxtaposition, suggest that they evolved relatively recently by duplication of a common ancestral gene, probably after the divergence of the Old World and New World monkeys 30-40 million years ago. In contrast, the sequence of the short-wave (S) pigment is almost as different from that of the L and M pigments as it is from the sequence of the rod pigment, rhodopsin; and it is

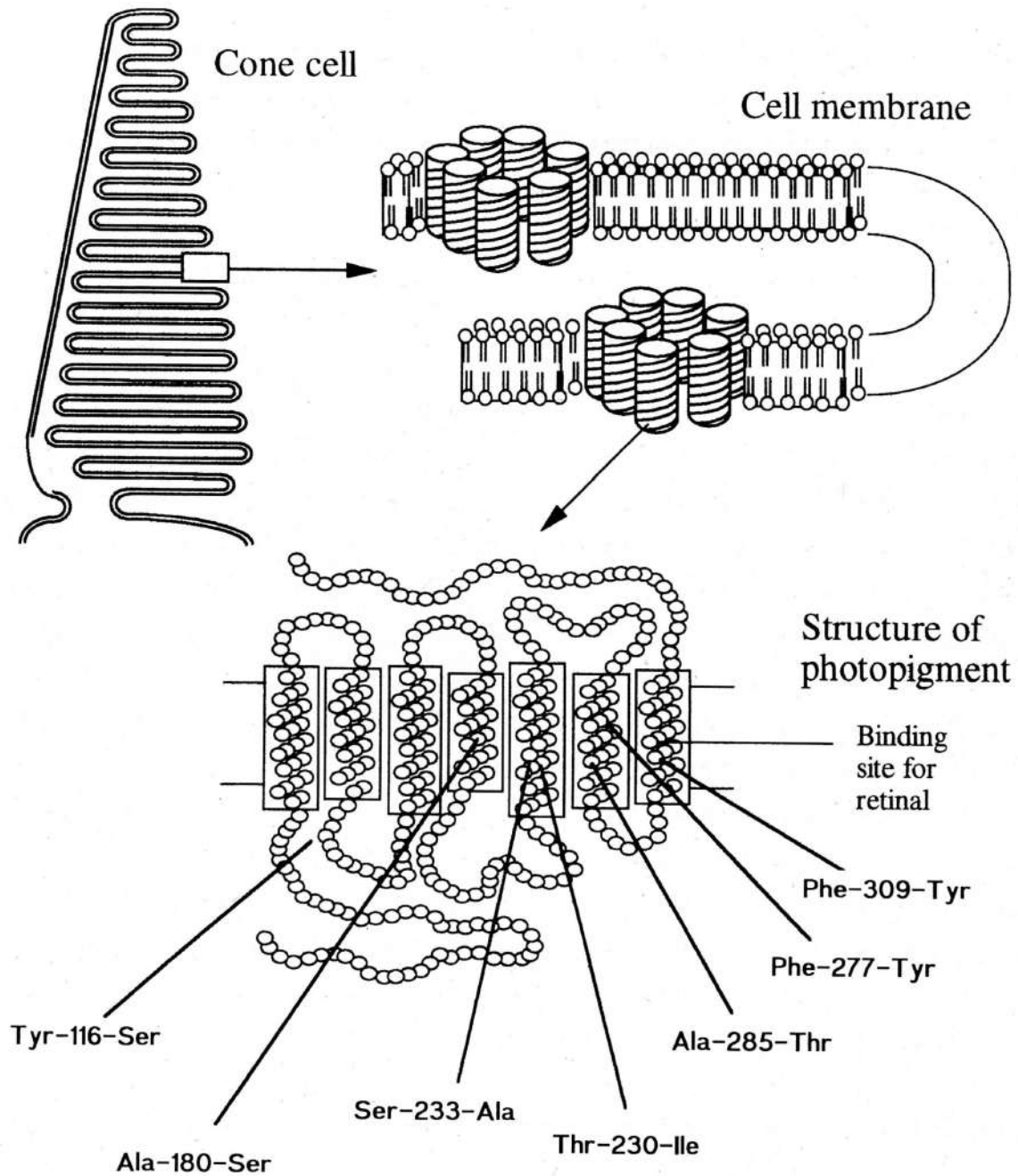


Fig. 2 To the upper left is shown the outer segment of a cone cell, with its membrane packed with photopigment molecules. The latter have a heptahelical structure, consisting of seven transmembrane helices, joined by loops alternately inside and outside of the membrane. The sequence of amino acids is depicted to the bottom right. The seven sites identified are ones where a substitution of an amino acid shifts the spectral sensitivity of the visual pigment: in each case the code to the left indicates the substitution that would shift the spectral sensitivity to shorter wavelengths and the code to the right represents the substitution that would shift the sensitivity to longer wavelengths. Indicated in the seventh helix is the lysine residue that binds to the chromophore, 11-cis-retinal.

likely that the gene for the S opsin had its origins long before the emergence of the mammals (Nathans et al., 1986; Yokoyama, 1994). It is primarily these observations that suggest that our present colour vision depends on two sub-systems, an ancient one that compares the quantum catch of the S cones with that of the L and M cones, and a second sub-system, overlaid on the first, that compares the quantum catches of the L and M cones (Mollon, 1991b).

The early mammals were nocturnal and may well have given up a richer colour vision enjoyed by their ancestors. Even now, the colour vision of diurnal mammals (even that of the trichromatic primates) is impoverished by comparison with that of many surface-dwelling fish, diurnal birds and reptiles (Bowmaker, 1991). The two cone pigments that survived the nocturnal period of the early mammals were probably (i) a pigment with a peak sensitivity in the range 530-570 nm (the ancestor of the present L and M pigments) and (ii) a short-wave pigment that was more closely related to the violet or ultraviolet pigments of non-mammalian species than it was to the 'blue' pigments of fishes (Yokoyama, 1994). It is conceivable that one or both of the corresponding genes was inactivated during the period when the mammals were nocturnal, but today pigments of the two types are found in most mammals (Jacobs, 1993). By comparing the quantum catches in the two types of cone, a basic form of colour vision is achieved, the type of colour vision that survives in most human dichromats.

The short-wave cones are rare almost wherever they are found in the mammals: in man and primates they typically constitute 3% -5% of all cones (Bowmaker et al., 1991; Mollon and Bowmaker, 1992) and they are completely absent from a 20-minute area in the very centre of the foveola. Probably their chief role has always been to provide colour vision and they are not used in the resolution of high temporal or spatial frequencies. Of course, the colour signal must carry some spatial sign, since the organism exploits chromaticity to identify particular objects and to segment the visual field (Mollon, 1992); but the information available from the short-wave cones is confined to low spatial frequencies and does not support the perception of sharp edges (Tansley and Boynton, 1976; Thoma and Schiebner, 1980). In her Berlin dissertation, Susanne Liebmann identified pairs of colours that melted into each other at equiluminance (Liebmann, 1927): we now know

that these are pairs of colours that differ only in the excitation of the short-wave cones.

This ancient sub-system of colour vision measures the relative balance of short- and long-wavelengths. The spectroradiometric measurements of Hendley and Hecht (1949) suggest that this sub-system is good for discriminating different forms of natural vegetation: when the chromaticities of foliage are plotted in the CIE diagram, they fall along a tritan confusion line, that is, on the axis along which only the short-wave signal varies. It is significant that the most celebrated of human dichromats, John Dalton, remarked: "*I can distinguish the different vegetable greens one from another as well as most people; and those which are nearly alike or very unlike to others are so to me.*" (Dalton, 1798).

One of the interesting features of the phylogenetically older sub-system of colour vision is that it appears to enjoy its own morphological substrate. Mariani (1984) described a distinct class of retinal bipolar cells that resemble midget bipolars but contact two or more separated cones, thought to be short-wave cones. More recently Dacey and Lee (1994) have identified a minority type of retinal ganglion cell, the small bistratified type, which carries the signal of the short-wave cones. There is some evidence that the signals of the ancient colour system are confined to laminae 3 and 4 of the lateral geniculate nucleus (Schiller and Malpelli, 1978; Michael, 1988) and have their own system of cytochrome oxidase blobs in Area 17 (Ts'o and Gilbert, 1988).

The phylogenetically newer sub-system of colour vision.

The second sub-system of colour vision depends on a comparison of the relative quantum catches in the long- and middle-wave cones. It increasingly seems that this second sub-system co-evolved with a class of tropical trees characterized by fruits that weigh between 5 and 50 g, that are too large to be taken by birds, and that are yellow or orange in colour when ripe. The tree offers a signal that is salient only to agents with trichromatic vision. This hypothesis of co-evolution of the tree's signal and the monkey's vision can be traced back to nineteenth century naturalists (Allen, 1892), but it

takes on new plausibility in the light of ecological evidence that there are many species of tropical trees that are dispersed exclusively by monkeys, and these are the trees with yellow or orange fruit. Such trees are found both in the Old World (Gautier-Hion et al., 1985) and in the New (Charles-Dominique, 1993). Thus in French Guiana several members of the family *Sapotaceae* appear to be specialized for dissemination by red howler monkeys (*Alouatta seniculus*) and other primates (Julliot, 1992). In some instances, the fruiting trees may be as critical to the monkeys as the monkeys are to the trees: thus the guenons of Gabon may rely on fruit for up to 85% of their diet (Sourd and Gautier-Hion, 1986). The tree offers a colour signal that is visible to the monkey against the masking foliage of the forest, and in return the monkey either spits out the undamaged seed at a distance or defecates it together with fertilizer.

It is instructive that picking fruit is a task where colour-deficient men are at a clear disadvantage. One of the first cases of colour blindness to be described was that of the shoemaker Harris. Reporting the case in the *Philosophical Transactions of the Royal Society* for 1777, Huddart wrote tellingly of him: "*Large objects he could see as well as other persons; and even the smaller ones if they were not enveloped in other things, as in the case of cherries among the leaves.*" We particularly depend on colour vision when the background is dappled and variegated, that is, when we cannot use form or lightness to find our target (Mollon, 1991b). In our modern urban environments, colour vision may be a luxury; but for frugivorous primates, it may be a necessity. In recent work in French Guiana, Ben Regan and I have been measuring quantitatively the reflection spectra of fruit and foliage, to establish how well matched is platyrrhine colour vision to the signal/noise discriminations it faces.

The morphological substrate for the second sub-system of colour vision is conventionally taken to be the system of midget bipolar cells and midget retinal ganglion cells. In the foveal region, a midget ganglion cell draws its centre input from a single cone, either long-wave or middle-wave. The antagonistic surround input is either drawn from cones of the opposite type, as held by Reid and Shapley (1992), or from a combination of L and M cones, as suggested by Lennie, Haake and Williams (1991). A different view has been advanced by Rodieck (1991), who suggests that the midget ganglion

cells subserve only the resolution of spatial detail and that both sub-systems of colour vision depend on bistratified types of ganglion cells. He believes that the latter project to the functionally defined "Type II" LGN cells of Wiesel and Hubel (1966), which are chromatically but not spatially opponent and which exhibit neutral points either near 500 nm or near 600 nm. But are there enough Type II cells? Rodieck suggests there should be about 11% as many Type II as ordinary P cells in order to account for the ratio of chromatic to achromatic spatial resolution. Empirical studies give values of 9% or 16% for the actual proportion of Type II cells. However, Rodieck's argument is misleading. We need to account separately for the spatial resolution of the two sub-systems of colour vision. We might expect the resolution of the second sub-system to be the better of the two, since L and M cones are of similar frequency in the primate retina and much more common than S cones - and since the sub-system probably evolved to detect fruit signals that subtend a small angle when viewed at a distance. Empirical evidence that the newer sub-system has better spatial resolution for colour as such is given by Regan and Mollon (in press). Yet the Type II cells with long-wave neutral points (i.e. those that draw signals of opposite sign from L and M cones) are the rarer in all the studies that Rodieck cites in his Table I. They amount to only 4-5% of all parvocellular units.

The two sub-systems of colour vision and the Urfarben of Hering.

The reader might suppose that the two sub-systems I have been describing correspond to the traditional blue-yellow and red-green axes of colour space. There are four hues that most observers judge to be pure or unmixed - red, yellow, green and blue - whereas other hues appear phenomenally to be mixtures: we feel we can see the red and the yellow in orange. Moreover, we never experience bluish-yellows or reddish-greens. On the basis of such phenomenological observations, Ewald Hering was led to postulate two corresponding pairs of antagonistic processes in the visual system, a yellow-blue process and a red-green one (Hering, 1878).

In fact, despite the statements still found in textbooks, there is no physiological evidence for a stage in the visual system where individual cells secrete the sensations of yellowness, blueness, redness or greenness. The

two sub-systems found in the retina and visual pathway simply do not correspond to Hering's two axes. A light that varied between pure yellow and pure blue, for example, would strongly modulate both sub-systems, whereas a variation between violet and lime-yellow is what is needed to stimulate exclusively the ancient colour system. The status of the *Urfarben* of Hering remains one of the unsolved mysteries of colour vision.

Conclusions and postscript.

The human brain has necessarily been bodged together: it has never had, as far as we know, a Designer and one must suppose that new systems have been tacked on in the course of evolution as our ancestors occupied new niches. Catarrhine and human colour vision particularly well illustrates this principle. I have argued that our colour vision system depends on two sub-systems, one antedating the mammals, one emerging only in frugivorous primates. The two sub-systems have different morphological substrates. They have different spatial and temporal properties. And in some psychophysical experiments (Krauskopf, Williams and Heeley, 1982) it is possible to adapt them independently.

The account of colour vision given here is similar to that advanced by Mrs. Christine Ladd-Franklin (1847-1930) who came to learn visual science in Göttingen a century after Thomas Young. At that time, in 1891, women were not admitted to lectures in Göttingen, but it is said (*New York Times*, March 6, 1930) that G. E. Müller was so impressed by this American woman that he repeated his lectures to her in private. On the basis of what was then known about colour blindness and about colour vision in animals, the redoubtable Mrs. Ladd-Franklin proposed that colour vision had successively evolved from monochromacy to a form of deuteranopia and thence to full trichromacy (Ladd-Franklin, 1892).

References

- Asenjo, A. B., Rim, J. and Oprian, D. D. (1994). Molecular determinants of human red/green color discrimination. *Neuron*, 12, 1131-1138.
- Bowmaker, J. K. (1991). The evolution of vertebrate visual pigments and photoreceptors. In: J. R. Cronly-Dillon, R. L. Gregory (eds.) *Vision and Visual Dysfunction, Volume 2, Evolution of the Eye and Visual System*. London: Macmillan, pp. 63-81
- Bowmaker, J. K., Astell, S., Hunt, D. M. and Mollon, J. D. (1991). Photosensitive and photostable pigments in the retinæ of Old World monkeys. *J. exp. Biol.* 156, 1-19.
- Charles-Dominique, P. (1993). Speciation and coevolution: an interpretation of frugivory phenomena. *Vegetatio*, 107/108, 75-84.
- Dacey, D. M. and Lee, B. B. (1994). The 'blue-on' opponent pathway in primate retina originates from a distinct bistratified ganglion cell type. *Nature*, 367, 731-735.
- Dalton, J. (1798). Extraordinary facts relating to the vision of colours. *Memoirs of the Literary and Philosophical Society of Manchester*, 5, 28.
- Dartnall, H. J. A., Bowmaker, J. K. and Mollon, J. D. (1983). Human visual pigments: microspectrophotometric results from the eyes of seven persons. *Proc. R. Soc. London B*, 220, 115-130.
- Firestein, S. (1991). A noseful of odor receptors. *Trends Neurosci.* 14, 270-272.
- Gamauf, G. (1811). *Lichtenberg über Luft und Licht nach seinen Vorlesungen herausgegeben. Erinnerungen aus Lichtenbergs Vorlesungen über Erxlebens Anfangsgründe der Naturlehre*. Vienna & Trieste: Geistinger.
- Gurney, H. (1831) *Memoir of the Life of Thomas Young*. London.
- Hendley, C. D. and Hecht, S. (1949). The colors of natural objects and terrains, and their relation to visual color deficiency. *J. Opt. Soc. Am.* 39, 870-873.
- Hargrave, P. A., McDowell, J. H., Curtis, D. R., Wang, J. K., Juszczak, E., Fong, S.-L., Roa, J. K. M. and Argos, P. The structure of bovine rhodopsin. *Biophys. Struct. Mech.* 9, 235-244.
- Hering, E. (1878). *Zur Lehre vom Lichtsinne. Sechs Mittheilungen an die Kaiserliche Akademie der Wissenschaften in Wien*. Wien: Carl Gerold's Sohn.
- Huddart, J. (1777). An account of Persons who could not distinguish Colours. *Phil. Trans. R. Soc.* 67, 260-265.
- Jacobs, G. H. (1993). Mammalian colour vision. *Biol. Rev.* 67, 413-471
- Krauskopf, J., Williams, D. R. & Heeley, D. W. (1982) Cardinal directions of color space. *Vision Res.* 22, 841-856

- Ladd-Franklin, C. (1892) A new theory of light sensation. In: International Congress of Psychology, 2nd Session. Williams and Norgate, London, pp 103-108
- Lennie, P., Haake, P. W. and Williams, D. R. (1991). The design of chromatically opponent receptive fields. In: M. S. Landy, J. A. Movshon (eds.) Computational Models of Visual Processing. Cambridge, Mass: MIT Press, pp. 71-82
- Liebmann, S. (1927) Über das Verhalten farbiger Formen bei Helligkeitsgleichheit von Figur und Grund. Dissertation, Friedrich-Wilhelms-Universität, Berlin.
- Mariani, A. P. (1984). Bipolar cells in monkey retina selective for the cones likely to be blue-sensitive. *Nature*, 308, 184-186.
- Mayer, A. M. (1875) The history of Young's discovery of his theory of colours. *Am. J. Sci.* 9, 251
- Merbs, S. L. and Nathans, J. (1992) Absorption spectra of human cone pigments. *Nature*, 356, 433-435.
- Michael, C. R. (1988). Retinal afferent arborization patterns, dendritic field orientations, and the segregation of function in the lateral geniculate nucleus of the monkey. *Proc. Natl. Acad. Sci. USA*, 85, 4914-4918.
- Mollon, J. D. (1991a). Hue and the heptahelicals. *Nature*, 351, 696-697.
- Mollon, J. D. (1991b). Uses and evolutionary origins of primate colour vision. In: J. R. Cronly-Dillon, R. L. Gregory (eds.) *Vision and Visual Dysfunction*. London: MacMillan. pp. 306-319
- Mollon, J. D. and Bowmaker, J. K. (1992). The spatial arrangement of cones in the primate fovea. *Nature*, 360, 677-679.
- Mollon, J. D. and Jordan, G. (1988). Eine evolutionäre Interpretation des menschlichen Farbensehens. *Die Farbe*, 35/36, 139-170.
- Nathans, J. (1989). Die Gene für das Farbensehen. *Spektrum der Wissenschaft* (April), 68.
- Nathans, J., Thomas, D. and Hogness, D. S. (1986a). Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science*, 232, 193-202.
- Nathans, J., Piantanida, T. P., Eddy, R. L., Shows, T. B. and Hogness, D. S. (1986b). Molecular genetics of inherited variation in human color vision. *Science*, 232, 203-210.
- Neitz, M., Neitz, J. and Jacobs, G. H. (1991). Spectral tuning of pigments underlying red-green color vision. *Science*, 252, 971-974.
- Peacock, G. (1855). *Life of Thomas Young MD, FRS*. London: John Murray.
- Regan, B. C. and Mollon, J. D. (1996) The relative salience of the cardinal axes of colour space in normal and anomalous trichromats. In: C. R. Cavonius (ed.) *Colour Vision Deficiencies*. XIII Dordrecht: Kluwer. (in press)

- Reid, R. C. and Shapley, R. M. (1992). Spatial structure of cone inputs to receptive fields in primate lateral geniculate nucleus. *Nature*, 356, 716-718.
- Rodieck, R. W. (1991). Which cells code for color? In: A. Valberg, B. B. Lee (eds.) *From pigments to perception*. New York: Plenum.
- Schiller, P. H. and Malpelli, J. G. (1978). Functional specificity of lateral geniculate laminae of the rhesus monkey. *J. Neurophysiol.* 41, 788-797.
- Sourd, C. and Gautier-Hion, A. (1986). Fruit selection by a forest guenon. *J. Anim. Ecol.* 55, 235-244.
- Tansley, B. W. and Boynton, R. M. (1976). A line, not a space, represents visual distinctness of borders formed by different colors. *Science*, 191, 954-957.
- Thoma, W. & Scheibner, H. (1980) Die spektrale tritanopische Sättigungsfunktion beschreibt die spektrale Distinkibilität. *Farbe & Design*, 17, 49-52
- Ts'o, Y. and Gilbert, C. D. (1988). The organization of chromatic and spatial interactions in the primate striate cortex. *J. Neurophysiol.* 8, 1712-1727.
- Watson, S. and Arkininstall, S. (1994). *The G-protein linked receptor*. Facts Book. London: Academic Press.
- Wiesel, T. N. and Hubel, D. H. (1966). Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *J. Neurophysiol.* 29, 1115-1156.
- Williams, A. J., Hunt, D. M., Bowmaker, J. K. and Mollon, J. D. (1992). The polymorphic photopigments of the marmoset: spectral tuning and genetic basis. *EMBO Journal*, 11, 2039-2045.
- Young, T. (1802). *The Bakerian Lecture. On the Theory of Light and Colours*. *Phil. Trans. R. Soc. London*, 92, 12-48.
- Young, T. (1807) *A course of lectures on natural philosophy and the mechanical arts*. London, J. Johnson