Reflections on Abnormal Colour Vision and Theories of Colour

HANS KALMUS*

Historically, physiology is the daughter of pathology. Healthy people struggling with an overwhelming nature had no time to bother with the workings of their organs, and doctors—naturally—were always preoccupied with disease. For them the argument from malfunction to normal function has become almost second nature and the successes of this approach in such fields as neurology, endocrinology, and other special branches of physiology are indeed impressive. But in other more general ways, the argument from disease to health has been less useful. Teratology has not much contributed to studies of normal development and the psychologies based on the study of neuroses are of rather limited use. Nearer our theme the study of the deaf has only modestly contributed to the theory of hearing, and the study of the blind little to the understanding of vision. But the exploration of minor deficiencies of the sense organs, which do not entirely abolish function but only impair it, can be more rewarding and in this respect genetical disability has a special role to play. It provides the geneticist with an opportunity to repay some of the debts to the other sciences which he has incurred in the past. The purpose of this paper is broadly to survey colour vision work from Young’s (1807) time to the present, and to find out to what extent the investigation of people with abnormal colour vision can contribute to our understanding of colour perception and indeed to our ideas of what colour is.

What is Colour?

For children and simple people colour is just an attribute of objects: grass is green, fire is red, snow is white. The association of certain colours with certain natural objects is indeed so strong that the adjectives designating these colours are in many languages identical with the nouns for objects, of which they are characteristic attributes, e.g. violet, orange, rose, lemon, etc. In ordinary life everybody, however sophisticated, uses colour for the identification of objects. This is in fact the main selective advantage of colour perception—whether for a man or an animal—as can also be inferred by the universal phenomenon of ‘colour constancy’, operating in a wide range of illuminations.

The first important step away from the naive concept of colour as an attribute of things, was taken by Newton (1704–1730) who showed that white light passing through a prism is separated according to its ‘refrangibility’ (or as we now say wavelength) into light of the different colours of the rainbow (spectrum). Since then colour has been persistently defined in physical terms, that is in modern parlance by the spectral composition of the light stimulus impinging on a particular area of the human retina. This does by no means accord with Newton’s own conclusions, as the following quotation from his Opticks shows:

The homogeneal Light and Rays which appear red, or rather make Objects appear so, I call Rubrific to Red-making; those which make Objects appear yellow, green, blue, and violet, I call Yellow-making, Green-making, Blue-making, Violet-making, and so of the rest. And if at any time I speak of Light and Rays as coloured or endowed with Colours, I would be understood to speak not philosophically and properly, but grossly, and accordingly to such conceptions as vulgar People in seeing all these Experiments would be apt to frame. For the Rays to speak properly are not coloured. In them there is nothing else than a certain Power and Disposition to stir up a Sensation of this or that Colour. For as Sound in a Bell or musical String, or other sounding...
Body, is nothing but a trembling Motion, and in the Air nothing but that Motion propagated from the Object, and in the Sensorium 'tis a Sense of that Motion under the Form of Sound; so Colours in the Object are nothing but a Disposition to reflect this or that sort of Rays more copiously than the rest; in the Rays they are nothing but their Dispositions to propagate this or that Motion into the Sensorium, and in the Sensorium they are Sensations of those Motions under the Forms of Colours.’ [Quoted from Wright, 1967.]

If Goethe had considered this statement he would not have been so scornful of Newton and if the detractors of Goethe’s Farbenlehre (1810) had considered his observations under this aspect, they would have found Newton on Goethe’s side. As it is, many contemporary scientists talking of colour would be classed as ‘vulgar people’ according to Newton. Positively speaking, we can, however, state that a colour stimulus may be characterized by the physical properties of the light—but not a colour sensation.

The lack of a unique correspondence between local light stimulus—be it a radiant source or a reflecting surface—and colour sensation is now familiar and might have been noted by scientists in the distant past. But while—to judge from their paintings—painters of those epochs who attempted realistic colour rendering must have observed that shadows can be coloured, and that in a more general way the colour of an area appears to be altered by the colour and brightness of adjoining areas, no explicit description exists of such observations. Nor was the deviant colour perception of ‘colour blind’ people recorded prior to Huddart’s classical description of a protanope (1777). This is quite remarkable if one considers that among the 1,000 or more Renaissance painters one might on chance have expected about 80 colour defectives. The colour confusions of painting apprentices cannot have passed unnoticed: were they eliminated for being just ‘stupid’ or did they learn to handle and match paints by devious means as happens in today’s art schools (Pickford, 1969)? One can only guess.

But to return to the science of colour vision: successive and simultaneous colour contrast (Hering, 1878) and of the changing colours of after images are by now well-established examples of the complex relations between light stimulus and colour sensation. An even more striking but little known phenomenon is the change of colour of a spectrally constant light source with its intensity and the duration of observation. Wheeler (1962) has shown that for normal observers filtered light, which at medium intensity appears red, changes its colour to yellow at high intensity and even appears green when thus observed for some time. In conclusion one has to state that spectral composition determines colour sensations only if the total temporal and spatial situation of an observer is sufficiently specified and his colour sense is taken into consideration.

Colour differences are only perceived above a certain light intensity, which in case of reflected light roughly corresponds to a white area, illuminated by the full moon. Below this intensity everything appears monochromatic. It is now commonly accepted that the monochromatic night vision is entirely mediated by the rods and polychromatic day vision, predominantly by a variety of cones. One may speculate why that is so. Rods can be activated by very few light quanta and it is reasonable to assume that within the species of photosensitive pigments available for natural selection only one—rhodopsine—is of the required sensitivity. The various cones on the other hand operate at higher light intensities, and several pigments, differing in their absorption properties but less sensitive, may have been available for selection.

It is easy to see why combinations of retinal receptors having different absorption characteristics are much more effective in distinguishing between lights of different composition—and thus between objects of different chromatic reflectance, than a single kind of receptor. While the output of a single receptor cannot be used to distinguish between differences in illumination intensity and differences in spectral reflection, two or more receptors, each having a different spectral sensitivity, provide two or more ‘scale readings’. For every coloured object the comparison of these readings at higher neural centres provides the material basis for the colour sensations. In the normal human retina 3 kinds of receptor provide information on 3 basic scales, from which relational values are computed, which we call colour. It is interesting that in the aerial Earth survey of the U.S.A. satellite service, photographs taken through different colour filters, i.e. through different sensors, also provide the basis for judging the nature of the surfaces.

The computed colour sensations of an infant are on the one hand associated with certain objects in his environment; on the other hand they are given names by the adults, which are repeated by the child and presumably learned by reinforcement. It has been shown by Rivers (1901) and Ray (1953) that the finesse and number of colour names of individuals is determined by the civilization in which they live.
To sum up: A modern view of colour would be that it is a named computed value (output) derived from the input of a number of retinal receptors, according to their spectral absorption properties, spatial relations, and previous exposure. Accepting this view one may now discuss at what levels of this informational system the common clinical defects of colour vision occur, how they affect colour perception, how they are compensated for, and finally what they have contributed to our understanding of normal colour vision.

Diagnosis of Colour Deficiency

The ophthalmologist's interest in colour mostly stems from his contact with patients, who are deficient in colour vision, mostly in connexion with some other manifest deficiency or disease. People whose only abnormal feature is the colour deficiency are of lesser interest to the clinician—if only because he cannot—as yet—do much to correct it. This may however change if recent Japanese claims for an electrical treatment can be confirmed. In any particular case the ophthalmologist follows a general medical routine. We shall accompany him on his way and see what colour theory can learn from his consecutive steps.

Anamnesis. Children are not aware of their defective colour vision, and this will sometimes show in funny ways. In the very first record of a colour defective man, a protanope, Huddart (1777) describes the 'anamnesis' of his subject, a shoemaker, as follows:

...'he had reason to believe other persons saw something in objects which he could not see; that their language seemed to mark qualities with confidence and precision, which he could only guess at with hesitation, and frequently with error. His first suspicion of this arose when he was about four years old. Having by accident found in the street a child's stocking, he carried it to a neighbouring house to enquire for the owner; he observed the people calling it a red stocking, though he did not understand why they gave it that denomination, as he himself thought it completely described by being called a "stocking".'

That ignorance of a person's defective colour vision may persist into maturity may be illustrated by the following typical situation, which one could perhaps describe as 'folie de couleur à deux'. At the beginning of each academic year the author asks for student volunteers who are 'colour blind', and in particular for girl students. And almost every year the following scene ensues. A boy and girl appear together, each accusing the other of colour-blindness. The disagreement usually arises from the choice and description of a tie, shawl, or other garment, and usually the boy is more confident of the 'correctness' of his colour sense. Invariably however it is he who is colour deficient simply because all the sex-linked forms of colour defects are so much more frequent in men than in women.

Now such situations, which 'serious scientists' would dismiss as anecdotal are in fact much more fundamentally revealing than many experiments involving advanced techniques. While it may not help much if regarded by itself, this incident reveals a great deal about the levels of organization at which colour information is formed and utilized, when considered in the light of our present general knowledge.

In the first place it points to the social context in which colour concepts are formed and associated with words. Our boy and girl use different vocabularies. The boy has until recently remained unaware of his defect, which implies that his classification of colours must in ordinary circumstances be sufficiently conformist not to arouse suspicion. However in certain situations he uses the wrong colour names, and further questioned appears uncertain and confused. The word 'colour-confused' has indeed been repeatedly proposed as a substitute for 'colour-blind'.

Status praesens. The first step in the diagnosis of a colour-blind person is the ascertainment of the kind and extent of his colour confusion. This is usually done by using coloured 'confusion charts' (Wright, 1946) which, following Stilling (1878) who first constructed such charts, go by the name of 'pseudoisochromatic'. Nowadays, the Ishihara and the HRR charts are the most popular.

Anomaloscope. Though the reading of colour charts provides some basis for the subclassification of colour defectives, this is usually done by anomaloscopy (Nagel, 1907). And here the trouble begins: there are always a small number of individuals who do not fit the neat instrumental classification into dichromats and anomalous trichromats; neither is there a lack of people who combine features of deutans and protans. Some of these difficulties can be explained by assuming that the exceptional individuals are compound heterozygotes if female, and carriers of both a protan and a deutan allele if male, accepting the two-loci theory. But usually one has to make several ad hoc assumptions, and the observed peculiarities are not easily explained in terms of a unified colour vision theory. The reasons I think are not difficult to find; they are...
rooted in the reductionist tendencies of science during the past century. Every phenomenon must be explained as simply as possible. Psychology in physiological terms, physiology in chemical terms, chemistry in physical terms, etc. In the case of colour vision this has led to a preoccupation with the physical and chemical properties of the retinal receptors. And while the study of these receptors has led to remarkable results, it has failed signally to produce a unified theory of normal and abnormal colour vision. There is no unanimity concerning the number of receptor classes in the abnormal retina, nor is there agreement on the nature of the sex-linked deficiencies. Following Young (1807), many people incline to the idea that dichromats, and in particular protanopes, lack one class of receptors, the red receptors, while others follow Aitken (1873) believing that both deuteranopes and deutanomalous trichromats do not lack any receptor class, but have faulty neural connexions (fusion) between the various cones.

**Receptor Theories of Colour Deficiency**

The simplest explanation of any particular kind of colour deficiency is to postulate the loss of one of the fundamental processes from the total set of such processes found in normal colour vision. This idea already formulated by Young (1807) does explain some of the facts, particularly the observations on protanopes—but by no means all. In terms of receptors we might now say protanopes have no red receptors (red cones). Whether the red cones are simply missing (loss theory), or whether in their place another kind of receptor has developed (substitution theories), is still contentious. As far as deuteranopes are concerned ‘fusion’ theories have been repeatedly formulated, i.e. the idea that either the pigments of the red and green cones may be mixed up, i.e. occur jointly in the same cone, or that the neural connexions with the higher centres are not properly separated. However most of these hypotheses do not agree with the following findings.

**Matching Proposition**

Most investigators of colour deficient people refuse to use verbal statements for diagnosis and theorizing, because they feel that they cannot cope with the complexities of colour naming. At the same time they underestimate the complexity of the matching situation during anomaloscope on which they largely rely. Matching on such an instrument utilizes two distinct, though to some extent interdependent criteria: colour (hue, chroma) and brightness. And these two variables can indeed be separately manipulated on such instruments. Now a normal trichromat matches the two halves of the test area almost entirely by colour, and he can easily distinguish differences in colour from differences in brightness. A dichromat, on the other hand, is supposed not to match by colour at all but only by brightness, and indeed many dichromats accept a wide range of red–green mixtures—including the normal one—provided the brightness is suitably varied.

However, observations indicate that many dichromats do in fact have red and/or green sensations especially when the stimuli are very saturated, and that they do, outside some wider or narrower range, refuse to match the parts of the field by brightness, or indeed to match them at all. There must then exist some overlap, where the two criteria—colour and brightness—compete. This is indeed borne out by verbal statements describing the situation or by attempts to produce two matches using each criterion separately.

Anomalous trichromats as a rule also match according to colour at a relative brightness near the normal. But their Rayleigh equation (red–green mixture) is different from the normal and their matching range somewhat wider. But again there exist ‘extreme’ protans and ‘extreme’ deutanomalous people with greatly increased matching ranges, as far as colour is concerned, and there is little doubt that brightness enters the computation of some of their matches.

The whole situation can be summarized by stating that any anomaloscope match is achieved by comparing both the colour as well as the brightness of two parts of the visual field, and that for a multitude of colour defectives neither the Rayleigh equation nor any other formula enables one to decide which criterion has prevailed. By changing from the verbal (qualitative) statements of colour naming, to the non-verbal (quantitative) matches on the anomaloscope one certainly produces data on which measurements can be performed; unfortunately, one may also lose sight of what one measures. In other words, anomaloscope readings quite frequently represent a compromise between colour and brightness matches, the relative strength of which remains unexplored. Now the importance of the colour component in any over-all match is almost certainly correlated with the subject’s colour efficiency but probably not very highly correlated: from this stems a great deal of the uncertainty in diagnosing the types of defective colour vision.

**Pigment Differences**

Until recently ideas concerning the nature and
localization of hereditary colour deficiency had to be based on indirect evidence, i.e. on inference from subjective tests. However, Rushton has shown that protanopes (1958) as well as deuteranopes (1965) differ from normal people and from each other in respect of their retinal pigments. By bleaching the retina of various observers with coloured lights and analysing the reflected light before and after bleaching he could objectively show that at least two pigments with different absorption curves—in the green and the red—are functional in the human retina. There is little doubt that a similar pigment with an absorption maximum in the blue also exists in the normal retina and that all this well accords with the Young–Helmholtz trichromatic theory.

However, Rushton’s results do not show whether the different pigments occur separately on different cones or whether two or even three of them might—for instance in certain defectives—occur together. Spectrophotometric measurements on isolated chimpanzee cones under the microscope have been published by several groups of research workers, but the results appear as yet unconvincing. So one must defer judgement.

Tritanopia

The manifestations of tritanopia, both acquired (Konig, 1897) and hereditary (Kalmus, 1955), indicate that special blue receptors shown to be present in the normal are missing in the affected person’s retina. Furthermore, Willmer and Wright (1945) have shown that even in the normal eye, the fovea centralis is tritanopic. It is thus probable that the autosomal tritanopia gene does not generally interfere with the synthesis of a particular cone pigment, but rather with the frequency, differentiation, and distribution of the ‘blue’ cones.

Night-blindness

The normal fovea centralis appears to be not only free of ‘blue cones’, it is also mostly devoid of rods; and it has been proposed (Willmer, 1961) that the rods may be implicated in blue perception, an idea that is supported by the bluish appearance of light at low intensities. This proposition is strengthened by observations on men suffering from hereditary stationary sex-linked night-blindness, who in daylight show a deficient blue sensation (Riddell, 1940). An alternative explanation would be that some ‘blue’ cones (or rods) function in scotopic conditions. The non-stationary forms of night-blindness which can be corrected by the administration of vitamin A are probably the consequence of some metabolic error in the retinene cycle, but whether the stationary forms are similarly caused is uncertain.

Symptomatic Colour Deficiencies

Deficiencies in colour perception are part of a great number of eye diseases and of toxic effects. The deficiencies may be general or localized (colour scotoma) and are usually not stationary but either progressive or reversible. Often they precede more serious degenerative processes, as in retinitis pigmentosa, or occur as ‘formes frustes’ in families where the fully fledged syndrome of late degenerative retinitis is also represented. Many—or not all—of these ‘colour-blind’ people appear deficient, when screened or tested with the charts or apparatus designed for discovering and diagnosing (Kalmus, 1965) the sex-linked stationary forms, but closer inspection usually shows that these non-stationary deficiencies do not fall into the same classes. We do not know whether these acquired lesions interfere with the receptors or the ganglionic layers of the retina.

Recognition and Naming of Colours

As mentioned before, the clinical preoccupation with diagnosis and the reductionism of contemporary biology have resulted in a relative neglect of those structures in the optical system which perform the higher integrative tasks in colour evaluation. Thus, what one may call the pleiotropic effects of colour vision deficiency have been mostly studied by psychologists and people interested in practical problems. And yet closer study of the verbal and conceptual elements in the make-up of colour deficient people is particularly interesting for a comprehension of the physiological organization of colour perception.

Vocabulary of Colour Defectives

Considerations concerning the colour sensations of people having defective colour vision are frequently dismissed by stating that nobody can experience other people’s sensations; and that it is therefore useless to talk about them. Nothing could be further from the truth. First of all there exist people who are colour-blind in one eye and more or less normal in the other (von Hippel, 1880, 1881; Graham and Hsia, 1958) and thus can overcome this ‘profound’ philosophical difficulty to some extent by simply closing one eye or the other. These people cannot only match a colour stimulus, as perceived with the defective eye
with a different colour stimulus perceived with the normal eye; they can also apply their colour vocabulary which turns out to be the complete normal vocabulary of their civilization to the sensations mediated by the affected eye. Both these faculties must have been employed by the subject described by Graham and Hsia (1958) who was filmed while performing colour matches by Farnsworth (1960); it appears that, as generally predicted for deuteranopes, her perception through the defective eye of many stimuli which she classified as ‘green’ or ‘red’ with the normal eye was one of ‘yellow’.

The situation is of course different for people colour deficient in both eyes, as they have no normal eye available to provide them with the ‘correct’ sensations in accordance with the colour names which they share with the colour normal individuals of their language group. Somehow they have to fit their abnormal sensations to the linguistic scheme derived from colour normal experience. Let us consider their situation in some detail.

While acquiring its vocabulary a child learns to apply the colour adjectives of its mother and other older people to the various objects of its environment, on the whole regardless of the kind of illumination. In the child endowed with normal colour sense, differences in sensation will on the whole correspond with differences in vocabulary, so that for instance an apple called ‘red’ by the mother will look to the child distinctly different in hue from an apple called ‘green’ or even ‘yellow’. However, little of the difference is perceived by the dichromatic child, and the transition between the colours is at different hues for anomalously trichromatic children. Both must therefore look for other criteria to fit the descriptive colour names which they hear and use to the objects described by them. This they learn to do in a number of contexts (settings), mostly by using relative brightness. Thus for a protanope a ‘red’ brick among other non-red bricks will be characterized by having a ‘yellow’ hue and by being dark, a ‘yellow’ brick will appear lighter and a ‘green’ brick also light. Uncertainty and confusion may tend to reduce the utilization of colour for the recognition of objects and the use of a colour vocabulary.

The exploration of the colour vocabularies of colour defective should yield most interesting results, but has never been seriously attempted. Our difficulty is that even among normal persons a great variability is observed when one studies colour naming. I have records of the naming of spectral radiations among college students, and to my astonishment have found that no wavelength of a group of 20 colour normal students agreed to call the light from a monochromator green, though all of them described a wide band of spectral light with this word.

**Allelism and colour vision.** The sex-linked defects of colour vision have played an important part in the history of human genetics ever since Wilson (1911) pointed out that Horner’s law (1876), which had until then been purely empirical, was an example of the distribution of a sex-linked recessive in man.

But the genetics and in particular the allelic relationships between the various protan and deutan genes have not been properly considered in their bearing on the epigenesis of the colour vision apparatus and on the classifications of deviant sensations.

In the view of most contemporary geneticists two loci on the X chromosome are responsible for most colour deficiencies, one for the protans and the other for the deutans. Probably these loci are fairly close on the chromosome, but this problem is not important for our present argument. Thus we have two allelic series ranging in their different ways from normal colour vision over anomalous trichromasy to dichromasy, and also, in females, definite dominance relationships in as much as

\[
\text{cv}^1_1 > \text{cv}^1_2 > \text{cv}^1_3 \}
\]

and this has some bearing on the question whether dichromasy is developmentally different in quality from anomalous trichromasy or whether we are confronted by a difference in degree artificially made into one of kind, either by the constraints limiting neural regulation in the colour deficient or possibly just by the instrumental properties of anomaloscopes. The differences of the various defects in the matching and naming of colours could then be largely explained by the interactions of these graded peripheral cone lesions with other structures in the retina, and in the brain.

Now it is a general experience that the allelic genes of a series affect the same site of a protein synthesis and as a rule the same primary developmental processes. Hypotheses concerning the primary causes of defective colour vision should therefore not consider dichromasy as essentially different in origin from anomalous trichromasy, but try to understand both in similar terms, for instance, as more or less severe losses, substitutions, or pigment anomalies of the cones, or alternatively as failures of neural commissions. On the other hand, the two series need not be caused by analogous developmental causes; protans may be characterized as having ‘loss systems’ and deutans as having ‘fusion systems’.
**Unique green.** A bimodality in the distribution of the spectral locus for 'unique green'(that is a wavelength which appears neither blue nor yellow among colour normal people has been claimed by Rubin (1962) and Richards (1967). Linksz and Waaler (1968) have connected these as yet controversial findings with anomaloscope readings of Waaler (1968) who, using a model anomaloscope, not any longer obtainable, reports the existence of two distinct classes of males and three classes of females in respect of Rayleigh equations, and postulates the existence of sex-linked genes an explanation. Rubin’s and Richards’ results have been contested (Hurvich, Jameson, and Cohen, 1968) and need confirmation; further speculation must be deferred.

**Matching coloured materials.** The matching of coloured materials, such as paints, ceramics, or textiles is not only of industrial and commercial importance but is also more akin than most instrumental methods to the ‘natural’ tasks of colour vision. Colours of objects, as seen by different observers or under different conditions of illumination, can for instance be assessed by means of coloured slide rules, the latest model (‘color rule’) of which is being manufactured by Davidson and Hemmendiger Inc., Easton, Pa., U.S.A. Matching points and tolerances of normal as well as colour defective people should be systematically measured by ophthalmologists, as they are likely to provide new information about the computational processes and structures which produce colour sensations. They might also provide a more rational basis for evolutionary speculation (Post, 1962) than chart and anomaloscope readings.

**References**


