The perception of light and colour and the physiology of vision—Part I. The mechanism of perception

SIR C V RAMAN

(Memoir No. 125 of the Raman Research Institute, Bangalore-6)

1. Introduction

Light, colour and vision play a role of fundamental importance in human life and much has been written about them from diverse points of view. Nevertheless, we do not find in the literature of the subject an acceptable answer to such a simple question as the following: How and why is it possible for us to distinguish by sight the different radiations emitted by a light source, such as the mercury arc for example, and name the colours we perceive? It is no answer to such a question to state that these radiations differ from each other in wavelength. While wave-optics enables us to understand why it is possible for well-defined images of external objects to be formed on the retinae of our eyes, the mechanism by which the incidence of the light on the retinae is communicated to the cerebral centres and its relation to the sensations which we perceive are subjects that stand completely outside the scope of the wave-theory of light. We have therefore of necessity to seek for an answer to our questions on the basis of the alternative approach provided by the quantum theory of radiation.

If a beam of monochromatic light falls upon a matt white surface, the illuminated object as seen by our eyes exhibits two distinct subjective features, viz., its brightness or luminosity and its hue or colour. From a physical point of view, the light falling on the surface has two distinct and definable properties, firstly, the radiant energy incident on it per unit of area per unit of time, and secondly, the magnitude of the quanta or units of light energy comprised in the beam. We are therefore led to associate the two characters of subjective perception respectively with the two physical attributes of the light, in other words, luminosity with energy-flux and colour with the magnitude of the individual light-quanta, and to infer that our eyes perceive colour by reason of the fact that light energy appears in discrete quanta and that these quanta are of different magnitude for the different rays in the spectrum. That we are justified in thus associating the sensation of colour with the magnitude of the light-quanta is reinforced by the consideration that for light to be perceived it must first be

absorbed by the visual receptor and such absorption necessarily occurs in complete quanta.

We have begun by considering the case of monochromatic light for the reason that it is the simplest case: monochromatic light does not admit of any further analysis. Only then it becomes possible to indicate the correlation between the physical properties of light and our sensory perceptions of the same in an unequivocal fashion. The sensations which result from the superposition of light of different colours or of different spectral composition present problems of greater complexity. They are no doubt of scientific interest and of practical importance, but we can scarcely hope that their study would lead us to an understanding of the fundamental aspects of the perception of light and colour. On the other hand, as is well known, they have led in the past to controversies and discussions of an unprofitable nature on various questions, of which it is sufficient to mention the two following as examples: Are there only three primary colours or a larger number? Which of the colours we perceive should be regarded as fundamental and which others as derived therefrom?

2. The duplicity of human vision

Our eyes are capable of functioning and serving as usefully over an enormous range of intensities. It has been estimated that the illumination of horizontal ground by the night-sky without moon is only about a hundred-millionth part of that received in direct sunlight, and yet it suffices for finding our way and keeping to the path provided our eyes are well adapted to darkness. That the mechanisms which function and enable us to perceive light are identical at both ends of this vast range of illumination is scarcely to be believed. Indeed, as is well known, the character of the sensations evoked is not the same in the two cases. In night-vision our eyes perceive differences in luminosity but their appreciation of colour is obviously imperfect. There are also other features which distinguish day-vision from night-vision, but we shall not pause here to discuss them.

In the year 1866, the retinologist Max Schultze announced a conclusion to which he had been led by an extended series of investigations in comparative ocular histology. He noted that the retinae of nocturnal vertebrates had many rods, and few cones or even none, while diurnal species had retinae with many cones which might even lack rods entirely. Thus, arose the idea that the cones are the receptors for daylight-vision and the rods for night-vision. That two types of vision exist and must be distinguished is very generally recognized, and they have indeed been given special names, viz., photopic and scotopic vision respectively. In the literature of the subject it is also generally assumed that photopic vision is mediated by the cones and scotopic vision by the rods in the retinae. But this assignment of the two types of vision to two sets of receptors in the retina does not really take us far. We have to ask ourselves whether the physical and

physiological processes are the same in the two types of vision or whether they are different and if so, in what respects.

In night-vision, our eyes have to serve us at the very lowest levels of illumination and a supersensitive receptor is clearly necessary for their successful functioning in such circumstances. On the other hand, in daylight-vision, the levels of illumination are higher; indeed, the quantity of light available in day-vision is, judging by the standards obtaining in night-vision, enormously larger than is adequate for its effective perception. Hence, a relatively insensitive receptor is what is actually needed. A further aspect of the matter that should be taken note of is that our eyes have to serve us both at very low and at very high levels of illumination, though, of course, not simultaneously. It would evidently be desirable that the supersensitive mechanism of night-vision is put out of action when the insensitive mechanism needed for daylight-vision is functioning; it would, of course, then be necessary that the apparatus of night-vision is restored and put back into action when it is actually needed. No such switching off and on is needed in the case of daylight-vision.

The foregoing remarks make it clear that the mechanisms of photopic and of scotopic vision would differ in certain respects. The exact nature of such difference is however best discussed at a later stage after the actual facts regarding the photopic and scotopic perception of light have been reviewed.

3. The facts of colour perception

From our present point of view, the most important facts concerning colour and daylight-vision are the relationships which exist between the subjectively perceived sensations of luminosity and colour and their physical counterparts, viz., the total energy-flux and the magnitude of the energy-quanta appearing in different parts of the visible spectrum. We may first briefly recall the distribution of colour in the spectrum. The colours observed in it differ so obviously from each other that common usage finds it convenient to give them different names to aid us in recalling them to our minds, viz., red, orange, yellow, green, blue and violet. The ranges over which these designations are applied vary with the individuals concerned, and only a kind of general average can therefore be regarded as possessing any significance. Expressed in wave-numbers and therefore directly related to the magnitude of the energy-quanta, these ranges are, respectively, red (13,000–16,000), orange (16,000–17,000), yellow (17,000–17,600), green (17,600–20,200), blue (20,200–23,000) and violet (23,000–26,000). It is noteworthy that while the ranges in wave-numbers covered by the designations red, green, blue and violet are all of the same order of magnitude, the colours designated as orange and yellow cover much smaller ranges.

We shall return presently to a more detailed discussion of the colour sequence observed in the spectrum, and proceed meanwhile to consider the relation

between the subjectively perceived brightness in different parts of the spectrum and the physically measurable energy-distribution in it. The essential features of this relationship can be established in a very simple manner by holding a diffraction grating in front of the eye and viewing the straight filament of a tungsten lamp, the current flowing through which is controlled by a rheostat. Observations of the first-order spectrum into which the luminous filament appears drawn out show that the distribution of luminosity in it is quite different from the energy-distribution in thermal black-body radiation as given by the Planck formula. According to the latter formula, the wavelength of maximum energy in the spectrum lies in the remote infra-red when the source has a temperature of 500° K., and though it moves towards the red end of the visible spectrum as the temperature is raised, it remains outside the visible spectrum even at as high a temperature as 3,000° K. On the other hand, the maximum luminosity as visually observed appears well within the spectrum even when the temperature of the filament is so low that it glows only feebly with a red colour. As the temperature of the filament is raised and it becomes brilliantly incandescent, the point of maximum luminosity in the spectrum moves swiftly towards shorter wavelengths and finally reaches a point in the greenish-yellow region of the spectrum at about 5600 Å beyond which it ceases to move any further. This is also the wavelength of maximum visual intensity in the spectrum of direct sunlight. In the spectrum of the light of the blue sky, in which the energy-distribution relative to sunlight is radically altered, the position of the maximum visual intensity is nevertheless not appreciably different. It is evident from these facts that it is a characteristic of daylight-vision that the eye has a highly pronounced maximum of visual sensitivity at about 5600 Å on either side of which the visibility of radiation falls off rapidly. It is also evident from the observations that such falling off of visibility is less rapid towards the red end of the spectrum than towards the violet.

The features referred to above find quantitative expression in the "relative luminous efficiency" curves for monochromatic radiations appearing in the visible spectrum which have been determined by several investigators. The two methods most often used for this purpose are the so-called step-by-step method and the flicker method. In the former method, the difficulties of heterochromatic photometry are minimised by dividing up the spectrum into a fairly large number of strips (twenty or more) and successively comparing the intensities of adjoining strips. In the other method, the luminosity of monochromatic radiation is determined by the method of comparison at the rate of 10–20 times per second against a constant source of white light, the intensity of the monochromatic light being varied until the flicker experienced is a minimum. Differences are noticed between the determinations made by different observers, but by averaging the results obtained with a large number of observers, a luminous efficiency curve is obtained which can be regarded as a property of normal human vision. The results obtained by the two different methods are about the same, though some
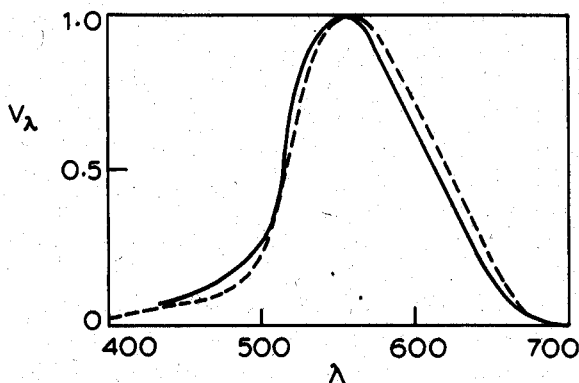
Figure 1. Relative luminous efficiency curves by the step-by-step method (———) and the flicker method (-----).

discrepancies are noticeable. This is shown by the two curves reproduced as figure 1 above exhibiting the results obtained by Gibson and Tyndall with the step-by-step method, and by Coblenz and Emerson with the flicker method.

We now return to the colours exhibited by monochromatic light. Earlier, it was indicated that the varying colours of the spectrum represent our subjective perception of the physical fact that the energy-quanta are different in its different regions. On this bais, we should expect the colours of the spectrum to form a continuous sequence, in other words that the spectrum can be divided into a great many parts or sections of which the colour can be perceived to be different. That is indeed actually the case. The number of distinct patches in the spectrum which when set side by side cannot be matched by any alteration in relative brightness has been estimated as about 250. If all the sections into which the spectrum may thus be divided were of equal width on a frequency scale, the smallest spectral shift giving rise to an observable change in colour would be about 50 wave-numbers. If this figure were accepted, it would mean that the human eye possesses a most remarkable degree of sensitivity to differences in colour and that it is capable of appreciating a difference less than one-half of 1%. In the magnitude of the light quantum received and absorbed by the retina as a difference in colour of the radiation.

It should be remarked, however, that the minimum spectral shift needed for a noticeable change of colour is not the same at all parts of the spectrum. It exhibits very significant variations, being in some parts smaller and in other parts larger than the figure stated above. Numerous investigations have been made in the past to determine the minimum in various parts of the spectrum and when the results reported are compared with each other, one finds a general agreement regarding certain features, viz., that in the region around 4900 Å and also in the region around 5900 Å, the minimum has quite small values, being respectively about 9

and 11 Å, and that in the intermediate region, viz., around 5400 Å, it is distinctly larger, being about 21 Å. Expressed in wave-numbers, these same figures are, respectively, 38 and 26 and 62 wave-numbers as against the averaged figures of 50 stated above. At wavelengths less than 4900 Å and greater than 5900 Å the determinations show that the minimum shift of wavelength increases rapidly. In these regions, however, the quantitative results of different investigators differ rather widely from each other.

The method by which the data of the kind referred are obtainable may be briefly described as follows: The observer views the two halves of a photometric field illuminated respectively by light from two adjacent regions in the spectrum of which the separation can be varied. The relative brightness of the two halves of the field is adjusted to equality by the use of a rotating sectored disk or by some other device. If, when so matched, the two parts of the field appear definitely different, the spectral shift giving rise to the difference represents the least perceptible to the eye as a difference in colour. Figure 2 below exhibits "the hue
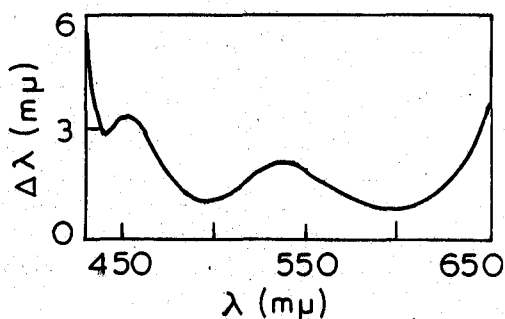

Figure 2. Hue discrimination curve in the spectrum, according to Wright and Pitt.

discrimination curve" in the spectrum obtained in this manner and reported by two observers. The abscissae in the curves are the wavelengths of monochromatic light in millimicrons, while the ordinates are the smallest shifts of wavelength (also in millimicrons) giving rise to an observable difference in colour. The figure exhibits the features already referred to above, and another not there mentioned, viz., the existence of a marked dip in the hue discrimination curve at about 4400 Å. Most of the published investigations report this feature and there is no doubt that it represents a well-established result.

Quite apart from the results of specific measurements made by the methods referred to above, it should be stated that a simple inspection of the spectrum by the unaided eye indicates that the colour sequence in the spectrum changes rapidly in certain regions, less rapidly in others and much less rapidly in certain others. An observer with normal colour vision will notice that the change of the blue to the green of the spectrum takes place within a relatively small range of

wavelengths between, say, 4900 and 5000 Å. Likewise, he will notice that the transition from the yellow to the orange of the spectrum takes place between, say, 5900 and 6000 Å. On the other hand, the transition from green to yellow is continuous, so much so that it is difficult to locate with any exactness the boundary between these two colours. The boundary between the blue and violet regions of the spectrum is placed in very different positions by different observers and indeed, it would seem that Newton was not altogether without justification when he put an additional colour "indigo" between blue and violet.

4. The mechanism of photopic vision

We now proceed to consider the significance of the facts regarding colour perception set forth above. For the reasons already explained, we associate the sensation of luminosity with the physical intensity or energy-flux, and the sensation of colour with the magnitude of the energy-quanta comprised in the radiation. Both the luminosity and colour of monochromatic radiation admit of study by appropriate physical methods which lead to quantitative results. These results do not contradict the physical interpretations proposed and they accordingly require to be further examined and explained. In other words, we have to answer various questions concerning them. In the first place, why does the luminous efficiency of monochromatic radiation vary in the manner observed and exhibit a maximum around 5600 Å? What is the factor which limits the power of the eye to discriminate differences in colour in the different parts of the spectrum? Why is it as great as it actually is in some parts, and why it is smaller in others?

The human eye has often been compared with a photographic camera. There are indeed some resemblances; in both cases we have a light-tight box covered inside with dark pigment to absorb unwanted light, a shutter to exclude light when desired and an adjustable diaphragm to regulate its admission when necessary. In both cases also, we have an optical system with front and rear components capable of adjustment so as to form a focussed image at the rear of the box. But to stress these structural similarities between a photographic camera and the human eye is liable to make us forget that the basic purposes which they are intended to serve are fundamentally dissimilar. A photographic camera serves to make a record of the light image on the sensitive film which can be developed and preserved. On the other hand, the retina of the eye serves both as a receiver and as a transmitter to the cerebrum of the energy of light falling on it. Whereas a photographic film can be used only once, the retina of the human eye is always ready for use. Indeed, vision would be useless to us if it did not carry with it the ability to view objects in different directions in quick succession one after another.

It is self-evident that light which merely passes through the retina and is absorbed by the pigmentary epithelium behind it would be ineffective in vision. In

other words, the absorption of light by pigments present within the retina itself is the first and necessary step for vision to be possible. But it is equally evident that unless such absorption goes hand in hand with a retransmission of the absorbed energy to the cerebral centres through the optic nerves, vision cannot result. It should also be remarked that since the colour perceived is determined by the magnitude of the energy-quanta comprised in the radiation, we have to assume that each quantum of energy absorbed is transferred immediately to the neurons without addition or subtraction. It would not be possible otherwise to explain the remarkable precision actually exhibited by our sense of colour discrimination. Merely as an example of such sensitivity, it may be mentioned that the two components of the sodium doublet whose wavelengths are 5890 and 5896 Å can be perceived to be of different colours in appropriate circumstances.

The facts of colour perception thus lead us to a quite simple and definite picture of the functioning of the retina in photopic or day-vision. We postulate that the light-sensitive receptors in the retina contain certain pigments which possess the power to transfer the light-energy absorbed by them to the neurons and return to their state of lowest energy. The known facts regarding the propagation of nervous disturbances and the production of electrical potentials by the infall of light on our eyes suggest that the energy transferred from the pigments assumes the form of an electrical disturbance in the receptors which is conducted through the optic nerves and on reaching the cerebral centres there excites the sensations we perceive. But we shall not pursue this aspect of the matter any further here.

5. The characters of scotopic vision

We now proceed to consider the actual facts regarding the perception of light at low levels of illumination, also referred to as scotopic vision. Earlier in this memoir it was suggested that the duplicity of human vision calls for a mechanism which could switch off the apparatus of scotopic vision when there is ample illumination and photopic vision can therefore function, and would switch it on again when the illumination is very feeble and photopic vision ceases to be possible. That such a mechanism is actually provided is evident from our own sensations when we move from sunshine into a dark chamber and *vice-versa*. The switching-on process is the slower of the two, since it is necessary to remain in the dark for at least 30 min for our eyes to develop the full sensitivity characteristic of scotopic vision. The switching-off process is somewhat quicker, but its duration is found to depend on the intensity of the light to which the retina is exposed.

Earlier, it was mentioned that in scotopic vision, the colour sense is imperfect. Such imperfection is due largely to the circumstance that the great sensitivity to light characteristic of scotopic vision does not extend to the parts of the spectrum which are both luminous and colourful, viz., the red and orange regions. Indeed, when any object of a scarlet-red colour is viewed in a darkened room, it appears

quite black. The characteristic features of scotopic vision are however best appreciated by studying the appearance of a continuous spectrum in scotopic conditions. A simple and convenient way of doing this is for the observer in a dark room to examine sky-light entering the room through a long vertical slit through a glass diffraction grating held in front of his eye. If the observations are made in day-time, the first-order diffraction spectrum of the slit then seen exhibits the familiar features characteristic of photopic vision. If, on the other hand, the observation is made at night when the sky is lit by moonlight (or in the absence of the moon by the lights of the city), the spectrum appears greatly shortened by the absence of all wavelengths greater than about 6000 Å. By making the observations during the twilight hours (evening or morning), the progressive change from the photopic to the scotopic spectrum or *vice-versa* can be readily followed. The changeover can be hastened or slowed down by the simple method of narrowing or widening the slit through which sky-light enters the dark room, or by the observer moving away from or towards the slit. In either case, large changes in the visual intensity of the spectrum result and the entire series of changes from the photopic to the scotopic conditions can be quickly and conveniently followed. Indeed, one can go down to the stage in which the intensity of the spectrum approaches the threshold of vision.

A characteristic feature of the spectrum seen under scotopic conditions is that the wavelength of maximum visual intensity shifts towards the blue as compared with photopic vision. Quantitative determinations of the relative luminous efficiency curve at low levels of illumination have been made by several observers, the method used being that of direct photometric comparison against a white field of low luminosity. A typical set of results is shown graphically in figure 3. It will be seen that the relative luminous efficiency becomes negligible at wave-
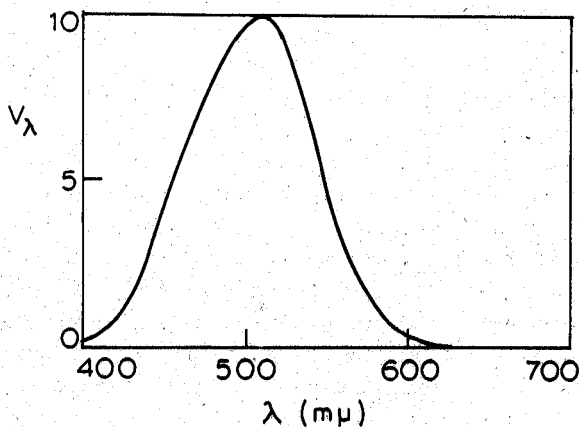

Figure 3. Relative luminous efficiency in scotopic vision.

lengths greater than about 6000 Å, while its maximum appears at about 5100 as against 5600 Å in photopic vision.

A question of special interest is whether the spectrum seen under scotopic conditions exhibits the colour sequence seen in photopic vision (apart from the disappearance of the longer wavelengths) or whether the scotopic spectrum is achromatic as is often stated. Observations made by the methods described above gave a clear answer to the issue here raised. Scotopic vision covers a wide range of intensities differing at its two ends by a large factor; whether in any particular case, we are observing the spectrum under scotopic conditions is indicated by the relative luminosity of the regions around 5000 and 6000 Å, the latter being negligible in scotopic conditions, while in photopic conditions, the latter is much more intense than the former. While the scotopic spectrum exhibits no visible colour anywhere at the lowest levels of intensity, this ceases to be the case at greater intensities. A greenish tinge becomes visible around the region of maximum luminosity; this is more obvious at greater intensities, and the differences in colour between the centre of the spectrum and its two ends are also then perceptible. As we approach the upper limits of scotopic vision, the colours observed become more vivid, but remain distinctly weaker and less saturated than in photopic vision.

6. The mechanism of scotopic vision

In their fundamental aspects, scotopic and photopic vision resemble each other closely. It is just as easy to view different objects one after another in quick succession at low levels of illumination as at higher levels and indeed one is then less troubled by after-images, glare and other disturbances. Visual photometry is possible in scotopic as in photopic vision and indeed in some respects easier, as for instance in heterochromatic photometry. If scotopic vision had been entirely achromatic at all levels of illumination, this feature would differentiate it fundamentally from photopic vision. But, as we have seen, such is not the case. The colour sense is actually present at the higher levels of illumination falling within the scotopic range and the colours then perceived, though less vivid, do not otherwise differ from those seen in photopic conditions.

In view of what has been stated, one would be justified in assuming that in their basic aspects, the mechanisms of scotopic and photopic vision are similar, in other words that in either case, the retinal pigments function by a simple transference of the light energy absorbed by them to the receptors, followed by a return to their normal energy state. The differences between the characters of photopic and scotopic vision described earlier do not appear to be inconsistent with such an assumption. That the relative luminous efficiency curve for scotopic vision differs from that determined for photopic vision is readily understood on the basis that the retinal pigments which function in the two cases are different.

That scotopic vision goes down to lower levels of illumination also becomes intelligible on the basis that the retinal pigment which functions is present in substantial quantities and can therefore trap a larger proportion of the incident energy quanta and pass it on to the receptors.

It now remains to explain how the switching-off and switching-on of scotopic vision take place respectively when the observer moves from low levels of illumination to high levels and *vice-versa*. It may justifiably be inferred from the actual facts of observation that when the observer has been long enough in bright light, the pigment which mediates scotopic vision has been transformed by the incidence of such light on his retina into a different substance which cannot so function; *vice-versa*, when the observer remains in darkness long enough, the pigment is regenerated in some fashion. The switch-off would proceed at a rate proportional to the intensity of light to which the retina is exposed and would be the quicker, the more intense such illumination is. The regeneration of the pigment, on the other hand, would occur in the absence of light or at very low levels of illumination and its rate would be determined by purely chemical or biochemical factors. The destruction of the pigment by bright light and its regeneration in the dark here contemplated are merely auxiliary processes which make it possible for human vision to function efficiently both at high and at low levels of illumination. There is no need and indeed no justification for assuming that they form an essential part of the actual mechanism of scotopic vision. Neither would photopic vision demand the destruction and regeneration of the visual pigments which function in it as an essential part of the visual mechanism.