THE IMPORTANCE OF CONTRAST FOR THE ACTIVITY OF SINGLE NEURONS, THE VEP AND PERCEPTION

ROBERT SHAPLEY
Rockefeller University, New York, NY 10021, U.S.A.

Abstract—The brightness of a visually perceived object is mainly determined by the average local contrast around the border between object and background. This fact is demonstrated here with several examples of equiluminant objects on nonuniformly luminant backgrounds. Even in Mondrian-like patterns resembling those used by Land and McCann (1971), equiluminant objects may appear to be of unequal brightness. This result does not agree with predictions of the Retinex Theory. The importance of contrast in vision is also suggested by neurophysiological findings, both classical and recent, that reveal the dependence of visual responses on contrast over most of the visual operating range of mean illumination. The dependence on contrast appears to be the result of retinal gain control mechanisms and is not due to center-surround interaction in the receptive field. We have discovered parallel neural channels with high and low contrast gain in the monkey's visual pathway by means of single unit techniques. Visual evoked potential measurements suggest that similar visual pathways, and with high and low contrast sensitivity, exist in man and monkey.

INTRODUCTION

Human observers rapidly and accurately estimate the brightnesses of an array of reflecting objects in a complex visual scene. This behavioral performance is relatively invariant with changes in illumination on a slow space scale or a slow time scale. It is not known how this visual problem is solved by the human brain. The most well known computational theory for the brain’s solution to this problem is Land’s elegant Retinex Theory. My recent experimental work implies that the Retinex Theory needs some revision to account for brightness vision.

There are four main properties of visual stimuli that have been found to affect the apparent brightness of objects. These are: luminance, contrast, spatial frequency content and assimilation. I will discuss these in sequence.

Other things being equal, the subjective impression of an object’s brightness is monotonically related to its luminance, the amount of physiologically effective light per unit area of the object (see Wyszecki and Stiles, 1967).

When other things are not equal, especially the luminance distribution in the surroundings of the object, an object’s brightness is not simply related to its luminance. The apparent brightness depends to a large extent on the physical contrast between the object and its background. Contrast may be defined as \((L_{\text{OBJ}} - L_{\text{BG}})/L_{\text{BG}}\), where \(L_{\text{OBJ}}\) is the luminance of the object and \(L_{\text{BG}}\) is the luminance of the background. Contrast is a measure of the relative variation of luminance. Probably the reason why brightness depends on contrast is that the visual system associates brightness with reflectance not luminance (Land and McCann, 1971; Stockham, 1972). While an object’s luminance may change with lighting conditions, its reflectance is an invariant property of the surface of the object. The physical contrast as defined here only depends on the reflectance of object and background and not on the level of illumination (the argument is given in Shapley and Enroth-Cugell, 1984). Thus, it is plausible that contrast should be the major cue for brightness.

Powerful psychophysical evidence that contrast was most significant for brightness vision was obtained by Heinemann (1955, 1972). Neurophysiological evidence indicates that the response amplitude of retinal ganglion cells is proportional to contrast up to some saturating value (Shapley and Enroth-Cugell, 1984). It has been known for a long time that neurons in the visual cortex respond to contrast rather than simply to luminance (Hubel and Wiesel, 1962). It is reasonable to presume that the cortical response-dependence on contrast is caused by the retinal response-dependence. The mechanisms that produce the contrast dependence
have been identified as the retinal automatic gain controls for light adaptation (Shapley and Enroth-Cugell, 1984).

Contrast is not the sole factor that influences brightness. Another crucial process is spatial filtering which removes low spatial frequencies from the "neural image" of the visual scene (Robson, 1975). This has the effect of suppressing neural responses to spatially gradual changes in illumination which are, under usual natural circumstances, due to variations in the illuminant and not due to gradual changes in reflectance. This low spatial-frequency suppression has been observed at all levels of the visual pathway (Robson, 1975; DeValois and DeValois, 1980). It is understood to be due in part to lateral inhibition: neural inhibition of a neuron from other neurons which have spatially displaced or larger visual receptive fields (Ratliff, 1965).

One of the consequences of spatial filtering is the enhancement of the visual salience of reflecting objects at the expense of (the usually slow) spatial variations in illumination (cf. Ratliff, 1965). The spatial frequency content of the sharp-edged borders of natural objects is broad-band. Thus, neural filtering-out of low spatial frequency information does not degrade the response to objects. Such filtering does prevent slow variations in the illuminant from being perceived (Stockham, 1972). Thus, spatial frequency filtering, at the retinal and cortical level, working together with the retinal automatic gain controls that produce contrast-dependence, allow the visual system to produce an illumination-invariant response to objects illuminated by light sources that vary slowly in space and time.

Another factor comes into play when the visual scene is not simply an object on background but becomes only as complicated as object (1) on object (2) on background. Then there is the possibility that the visual system might compute the brightness of object (1) by adding the brightness of object (2) to that of object (1). This possibility has been observed and has been called assimilation (Helson, 1963; Hurvich, 1981). Assimilation is a perceptual phenomenon that tends to antagonize the dominant dependence of perception on contrast; assimilation makes the effect of a background additive while contrast makes it subtractive. The concept of assimilation arose in studies of color vision where this phenomenon is observed as the tendency for an object on a colored background to take on the color of the background (Hurvich, 1981). Elegant experiments by Helson (1963) and Arend et al. (1971) demonstrated that assimilation also played a role in the perception of brightness. There was also evidence for an assimilation effect in Heinemann's experiments (Heinemann, 1955). Implicit in the Retinex Theory is the assumption of completely global assimilation. Assimilation is not local, but is not completely global according to our recent measurements, discussed below.

The emphasis in this paper is on the importance of contrast, which I believe to be the most significant stimulus property in the visual estimation of brightness. I will offer the reader several perceptual demonstrations of the effect of contrast on brightness. Then electrophysiological results will be presented concerning the neural machinery for contrast perception.

**METHODS**

**Visual stimulator**

The basic tool for many of our experiments is an electronic visual display, designed in our laboratory by Norman Milkman, Michelangelo Rossetto and Gary Schick (cf. Milkman et al., 1980). The instrument used in the present experiments produces two-dimensional patterns on a Tektronix 608 cathode ray tube (CRT) monitor. The patterns are homochromatic; in most of our experiments they were white (P4 phosphor). The instrument produces a raster display of 256 picture elements (pixels) per line, 256 lines per frame and 270 frames/sec. Thus there are 65,536 pixels per frame. The instrument contains control registers, four RAM memory modules (256 bytes each) which contain the spatial profiles of four independent spatiotemporal functions and one RAM memory (65,536 by 2 bits) that contains the bit map which controls which one of the spatiotemporal functions is displayed at each pixel location. The two bits for each pixel are led to a fast analog switch which chooses between the four spatiotemporal waveforms. The spatiotemporal waveforms are created from the spatial profiles in the four RAMs by digital/analog multiplication in a multiplying D/A converter. The spatial profile values are multiplied by temporal modulation and depth of modulation signals which are fed to the instrument by a PDP 11/23 computer. Besides the great flexibility the instrument has in the kinds of
patterns it produces, it has the great advantages of digital accuracy of positioning (to within 0.04 deg of visual angle at our usual viewing distance) and high precision of intensity control (nominally more than 20 bits of precision, though this is probably an overestimate because of noise in the instrument and noise picked up on the cables). A linearization circuit added to the Tektronix 608 by Mr Rossetto allows us to modulate contrast in a linear manner up to 60% at a mean luminance of 100 cd/m². We measure the light output from the CRT displays before and during each experiment, to make sure the nominal contrasts and mean luminance are correct. For this purpose we employ a Pritchard Model 1980b Photometer (Photo Research, Inc.).

Electrophysiological single-unit experiments

Our techniques for single unit recording are described in detail in previous publications (Hochstein and Shapley, 1976; Kaplan and Shapley, 1982). Briefly, we use glass micropipette electrodes to record extracellularly in the optic tract or lateral geniculate nucleus (LGN). The animal, cat or monkey, is anesthetized lightly with urethane, or is decerebrated by midcollicular transection. Eye movements are minimized by means of intravenous infusion of a muscle relaxant, gallamine triethiodide. The physiological condition of the animal is monitored through measurement of blood pressure, expired carbon dioxide, core body temperature, and electrocardiogram and electroencephalogram and appropriate steps are taken to keep these in the physiological range. The eyes of the animal are protected with plastic contact lenses. In the cat experiments, image quality is improved by using contact lenses with a small (3 mm dia.) artificial pupil. Image quality is checked with an ophthalmoscope throughout each experiment, and data are taken only when the retinal image is sharp.

Neural responses are analyzed using the same PDP 11/23 computer that controls the visual stimulator. Stimuli are presented repetitively and periodically in time. The neural impulse rate is averaged and responses are quantified as the amplitude of modulation of the impulse rate in synchrony with the modulated stimulus.

RESULTS

Brightness perception and contrast

A standard demonstration of the effect of contrast on brightness is Fig. 1. Here two equiluminant circles are placed on a bipartite field. The luminance of the circles is equal to the average luminance of the two halves of the bipartite field. The two circles are not equally bright to most observers; the circle on the darker background appears brighter than the circle on the lighter background.

Figure 2 is a picture of another two equiluminant circles but this time they have been placed on a linear gradient of luminance. That is, the luminance across the background is \( L(x) = L_0 (1 + mx) \). The slope \( m \) is the rate of change of background luminance with position \( x \), where \( x = 0 \) is the middle of the picture, and \( x \) varies from \(-5\) to \(+5\) cm on the cathode ray tube (CRT) screen which was the original image photographed to make Fig. 2. \( L_0 \) is the mean luminance of the background. The luminance of the circles is also \( L_0 \). I believe that there is a simple explanation for the brightness differences of the circles in Fig. 1 and 2 in terms of the sign of the contrast around their borders. In each of these pictures, the average local contrast around the circle on the left is positive and so it appears bright. The average contrast around the circle on the right is negative and it appears dark.

While the appearance of Figs 1 and 2 can be accounted for by the local border contrast, it cannot be explained adequately by the Retinex Theory for brightness (Land and McCann, 1971; Land, 1983). Without going into all the details, the Retinex Theory is constructed to give a constant visual response to a constant reflectance. Therefore, in a picture like Fig. 1 in which the two circles are equiluminant and there is no gradient of illumination, the Retinex Theory must, as well, explain the simple facts that pictures like Fig. 1 were accounted for by the Retinex Theory in their Footnote 1:

"Although sensations of lightness show a strong correlation with reflectances in most real-life situations, there are many important departures from this strong correlation. . . . And, of course, any general theory must, as well, explain the simple situations in which the surround comprises the entire environment . . . ."
An implication of the above quotations is that departures from the Retinex Theory occur mainly in nonreal-life situations, especially in "simple situations in which the surround comprises the entire environment." Figure 1 is such a "simple situation." While this argument may appear plausible, in fact the departures from Retinex Theory are widespread, even in complex pictures. In all the cases I will present, Retinex Theory makes qualitatively incorrect predictions while prediction of apparent brightness based on the average local contrast is correct, at least qualitatively. I will concentrate on generalizations of Fig. 1, that is pictures in which objects of equal luminance are perceived to be of unequal brightness, in contradiction to Retinex Theory.

The first illustration of the general importance of local contrast is Fig. 3. This is a complex picture which shows twelve circles of equal luminance on a complex background that is composed of a linear gradient of luminance in the lower half of the picture, and a staircase of luminance in the upper half. To make the picture more comprehensible, refer to the luminance profiles of the lower and upper halves of the figure as diagrammed in Figs 4 and 5. In Fig. 4 the locations of the circles along the horizontal axis of the picture are marked by arrows, and their luminance, which is the mean luminance of the gradient, is indicated by the dashed horizontal line. The position of the circles in the upper half of Fig. 3 and the mean luminance of the staircase, are indicated in a similar way in Fig. 5. Inspection of Fig. 3 reveals that the circles of positive local contrast always appear light and the circles with negative local contrast always appear dark, regardless of whether they lie on gradient or staircase backgrounds. The brightness of the circles is unaffected by covering or uncovering the upper half or lower half or any other fraction of the picture, as long as the border of the circle and its immediate background are visible. These observations are in contradiction to the Retinex Theory, which predicts: (1) that the circles on the staircase should look approximately the same brightness as each other, and (2) only the circles on the gradient should appear to be of different brightness, and (3) the circles on staircase or gradient should change in apparent brightness as parts of the picture are covered. In the visual system's response to Fig. 3, local contrast seems to be much more significant than the long-distance interactions postulated in the Retinex.

For those readers who are not familiar with the implications of the Retinex Theory, it is worth pointing out why the circles on the gradient in Fig. 3 should look different in brightness but the circles on the staircase should look the same in brightness, according to that theory. In the most modern version of the Retinex (Land, 1983), the brightness at a discrete area is determined from the average of the relative

---

Fig. 4. This is the luminance profile of the lower half of Fig. 3, the equiluminant circles on a gradient. Zero luminance is indicated by the horizontal axis. The positions of the equiluminant circles are marked by arrows. The mean luminance is drawn as a dashed line.
Fig. 1. Two equiluminant circles on a bipartite field of luminance. This is a photograph of a picture created on a Tektronix 508 CRT monitor with an electronic visual display (Milkman et al., 1980). Under our standard viewing conditions, the display on the CRT screen was 10 by 10 cm in area and had a mean luminance of 100 cd/m².
Fig. 2. Two equiluminant circles on a linear ramp of luminance. In this picture, the background luminance is described by the equation $L(x) = L_0^*(1 + mx)$, where $m$ is the slope of the luminance gradient, $L_0$ is the mean luminance of the gradient and $x$ is the spatial position with respect to the midline of the picture: minus to the left and plus to the right. The luminance of the two circles is the mean luminance, $L_0$. 
Fig. 3. Many equiluminant circles on a staircase of luminance, in the upper half of the picture, and on a linear gradient of luminance, in the lower half. The gradient and the staircase were produced so that the luminance along the left and right borders of the picture would be uniform along the vertical axis. The mean luminance of the staircase and the gradient were also set to be identical and also equal to the luminance of the circles. The average local contrast around the circles on the gradient is lower than for the circles on the staircase, as can be seen by comparing Fig. 4 and 5, which show, respectively, the luminance profiles of the lower and upper halves of this figure.
Fig. 9. A photograph of two equiluminant and equal contrast, circles that look unequally bright. This figure, from unpublished work by Reid and Shapley, illustrates the phenomenon of assimilation. The two small circles are equiluminant. The annuli that surround them are equal in luminance to each other. Therefore, the border contrast around the two circles is equal. But because the annuli are unequally bright, because of induction by a luminance gradient in the outer surround, the two inner circles appear of unequal brightness to most observers. However, it is a small effect and the picture requires scrutiny. The luminance profile of this picture is given in Fig. 10.
reflectances of that area compared to every other areal element in the scene. The relative reflectance of areas within one object compared to those areas in another object is the sum, along the path between the two objects, of the logarithms of the ratios of the luminances falling on adjacent retinal elements. This sum-of-the-logs-of-the-ratios formula is only true, however, if the luminance ratios exceed a criterion threshold amount. When the logarithm of the luminance ratio between two adjacent areas is below the threshold amount, the Retinex sets the logarithm to zero. Therefore, it sets the reflectance ratio between those two elements to one.

Now, let us consider Retinex Theory predictions for Fig. 3. In the case of the gradient background, the luminance ratio of adjacent retinal elements could be assumed to be below threshold. Then contributions to the average relative reflectance of one of the circles would come from areal elements near its border and the borders of the other circles, and the estimated relative reflectances of different circles would be unequal. However, in the case of the staircase, because luminance is constant except at the abrupt steps which must be assumed to be above the Retinex’s threshold, the estimated reflectances of all the circles are equal. meaning that they all must be equally bright were Retinex Theory right. For example, the Retinex computation of the relative reflectance between circle(1) of luminance $L_0$ on background $L_{B1}$ and circle(2) of luminance $L_0$ on background $L_{B2}$ is:

$$\{\log\left(L_{B1}/L_0\right) + \log\left(L_{B2}/L_0\right)\} + \log\left(L_0/L_{B2}\right)$$

$$= \log\left((L_{B1}/L_0)\right)^{\ast}(L_{B2}/L_0)^{\ast}(L_0/L_{B2})$$

$$= \log[1] = 0.$$  

This is the logarithm of the apparent brightness ratios between any two equiluminant circles on the staircase. zero. Therefore, the apparent brightness ratios predicted by Retinex Theory are all $\exp(0) = 1$. That is, all the circles should look equally bright. As you can see, it does not predict the appearance of Fig. 3.

It is possible that Fig. 3 is not complicated enough for the Retinex effects to become predominant. Therefore, I manufactured Fig. 6 which is composed of two staircases of luminance: an ascending staircase in the lower half of the picture, and a descending staircase in the upper half. In this picture, as in Figs 1-3, the circular objects are all equiluminant, at the mean luminance of the staircases. By equation (1), Retinex Theory predicts that all the circles should look equally bright. In this picture, the circles do not all appear to be the same brightness and therefore it is another qualitative counterexample to the original Retinex Theory.

It might be argued that all the pictures shown up to now are not good tests of the Retinex Theory because they are not Mondrian-like.

---

Fig. 5. Luminance profile of the circles on the staircase: the upper part of Fig. 3. Conventions are as in Fig 4.
patterns, like those favored by Land and McCann (1971). The crucial aspect of the Mondrian-like patterns is that the border of one object in the Mondrian is not completely surrounded by another object of uniform luminance. Therefore, Fig. 7 was produced. This is a Mondrian-like pattern with two test areas of equal luminance: the circle in the lower right quadrant of the picture, and the square in the upper left (the brightnesses of these two test objects were not dependent on their shapes). The measured relative luminances of the areas in the picture, when it was illuminated diffusely and evenly, are given in Fig. 8. Note that the luminance of each patch in this picture is constant across the patch; there are no gradients of illumination which might allow the Retinex Theory to predict that equiluminant areas could look different. The pattern was constructed to provide another strong test of the Retinex Theory, which must predict in such a pattern that equiluminant areas look equally bright, by the argument given above in equation (1). It is immediately evident that even in a uniformly illuminated Mondrian pattern, equiluminant objects may appear to be of unequal brightness, presumably because of unequal average local contrast. If one calculates the average physical contrast around the borders of the circle and the square equiluminant objects, one obtains $-0.11$ for the circle and $0.38$ for the square. The bright object is always associated with positive contrast, and the dark with negative contrast. The reader may note that there are other pairs of equiluminant patches in Fig. 7 which do not appear equally bright, also contrary to Retinex Theory. Equiluminant patches in more complex Mondrian-like patterns exhibit the same phenomenon of nonequality of brightness if the average local contrasts around their borders are unequal.

I believe the implication of these perceptual demonstrations is that the computation of brightness is much more localized in fact than in Retinex Theory. It cannot be completely local, because border contrast some distance away determines the brightness at a point. Nevertheless, the computation cannot be as global as in the Retinex Theory or else Figs 1–3, 6, 7 would look different.

Besides local contrast, another component of the brightness computation is assimilation, the effect on the brightness of an object of the brightness of its background. In general it has been hard to separate assimilation from contrast because the way backgrounds have been made brighter or darker is by varying their luminances and this affects border contrast of an object on the background. R. Clay Reid and I have done some experiments recently on assimilation where we avoided this difficulty by varying the brightness of a background by brightness induction from an outer background (Shapley and Reid, 1985). In this way we could change brightness while keeping contrast fixed. An example of the kind of pattern we used is shown in Fig. 9. The luminance distribution in Fig. 9 is diagrammed in Fig. 10. The outer background is a gradient of luminance. The small inner backgrounds are equiluminant annuli. The test spots are equiluminant circles. These circles not only are equally luminous, they also have equal local contrast around their borders. Nevertheless, they appear to be of unequal brightness; the circle on the brighter inner background appears to be a little brighter. This must be due to assimilation. Reid and I measured the magnitude of this assimilation effect and compared it to the strength of the effect of local contrast on brightness. Assimilation is about half as strong as contrast, and it wanes with distance between the borders of the test spot and inner background. These results to some extent explain why the Retinex Theory cannot account for the earlier figures. In the Retinex Theory, a summation of brightness from one point to another in the visual field is postulated to occur without any decrement with distance (see Land and McCann, 1971 and Land, 1983). This process
would be equivalent to total assimilation with no spatial decrement. The fact that assimilation does get weaker with distance implies that the calculation of brightness is more localized than is proposed in the Retinex Theory. The computation of brightness contains more complications: Reid and I have observed nonlinear effects of brightness induction and assimilation. However, this topic is beyond the scope of the present paper.

Electrophysiological results

The importance of contrast is also revealed by electrophysiological results that suggest that contrast and color are signaled to the brain by different parallel pathways. This is the conclusion of work done by Ehud Kaplan and I on single cells in the LGN of macaque monkeys. The LGN in monkeys as in man is a highly layered neural organ. There are six main cell layers: four dorsal layers composed of small cells (parvocellular layers) and two ventral layers composed of larger neurons (magnocellular layers). We recorded the response amplitudes of cells in the LGN to drifting gratings with a spatial frequency high enough that they would stimulate the center of the receptive field only (cf. Shapley and Lennie, 1985). We were surprised to find that the cells in the parvocellular layers had a poor contrast sensitivity by comparison with cat geniculate and retinal ganglion cells. The magnocellular cells were about 10 times more sensitive than the parvocellular neurons and were comparable in sensitivity to cat visual neurons (Shapley et al., 1981; Kaplan and Shapley, 1982). The situation in the monkey's LGN is demonstrated in Fig. 11 which is a sketch of one histological section through a monkey's LGN, and shows the course of an electrode track through it. The first unit with high contrast sensitivity recorded on this track was the one marked with a dot at the top of the magnocellular layers. We repeated this kind of experiment many times, always marking with dye spots the places where highly sensitive cells were recorded and they were always located in the magnocellular layers.

The function of the organization diagrammed in Fig. 11 has been interpreted in terms of separate signal processing for contrast and color. The parvocellular neurons are narrowly tuned in the dimension of wavelength, while the magnocellular neurons are more broadly tuned. Therefore, the function of the parvocellular neurons may be more connected with wavelength discrimination and colour vision, while the magnocellular neurons may be the thalamic relay stage for a separate contrast-sensing pathway. In recent work we have found that the retinal input to the LGN cell determines its contrast sensitivity (Kaplan and Shapley, 1984a). We have done this by recording pairs of LGN cell impulses and the unitary S potentials which are the locally, extracellularly recorded EPSP's from the retinal input (Levick et al., 1972: Kaplan and Shapley, 1984b). Thus, we believe that parallel processing begins in the retina.

Fig. 10. Luminance profile of the assimilation demonstration. The small circles are the two inner increments on top of the annular backgrounds which were at the mean luminance indicated by the dashed horizontal line.
Visual Evoked Potentials (VEPs). The VEP can be used to investigate the importance of contrast in human vision with an electro-physiological as opposed to a psychophysical technique. While this has been the focus of much study in other laboratories, as discussed below, I would like to illustrate a basic point with experimental results obtained by colleagues at Rockefeller University: Conte et al. (1983). The experiment was to measure the occipital VEP to the following stimulus: a 7.3 c/deg sine grating was modulated in contrast above and below a mean level of contrast by 100% of the mean contrast, so that at the trough of the modulation waveform, the sine grating disappeared and the screen became blank at the mean luminance, while at the peak of the modulation waveform, the contrast of the grating was twice the mean contrast. The stimulus is referred to as “Appearance-Disappearance” for obvious reasons. The “Appearance-Disappearance” VEP was averaged and analyzed into two Fourier

Fig. 11. Drawing of a histological section of the monkey's LGN and an electrode track through it. This is a tracing of an approximately coronal section through the LGN. It shows the four dorsal parvocellular layers and the two more ventral magnocellular layers. An electrode track is indicated by the dashed line. The filled circular spots were locations where current was passed through the dye-filled micropipette to mark electrode location during a cell recording. The upper dot was at a place where impulses from a typical color-opponent parvocellular neuron of low contrast sensitivity was recorded. The lower spot, at the top of the magnocellular layers, was the locus of recording of the first cell on the track that had high contrast sensitivity. From Shapley et al. (1981).
Contrast and activity of single neurons. VEP and perception

Fig. 12. Human VEP to contrast modulation of a sine grating. This is the response of a human subject recorded with gold cup electrodes placed on the occiput and the forehead. The stimulus was contrast modulation of a 7.3° deg sine grating. The contrast was modulated at 6.3 Hz in the "appearance-disappearance" mode, that is, the stimulus was \[ s(x, t) = \epsilon \cdot L_\phi \cdot [1 + \sin(2\pi f t) \cdot \sin(2\pi k x)] \]
where \( \epsilon \) was the average contrast, \( L_\phi \) was the average luminance, \( f \) was the temporal frequency, and \( k \) was the spatial frequency. The 1st and 2nd harmonic amplitudes of 1 min of VEP data are plotted against average contrast.

Conte et al. (1983).

DISCUSSION

The need for a new theory of brightness

There have been few attempts to provide a comprehensive theory of brightness perception in terms of an experimentally testable model. The Retinex Theory of Land and his colleagues is such a theory and has justifiably attracted the interest and the attention of vision scientists. However, models exist to be disproved and improved. The perceptual results presented in this paper demand the formulation of a new theory of brightness vision to improve on the Retinex Theory. The crucial experimental results are that the retina sends a signal to the brain coded in terms of local border contrast, and the brain seems to assign brightnesses to areas in complex scenes that correlate highly with their average border contrasts. However, contrast is not everything; the amount of assimilation must be taken into account. At present, we are just beginning to find out the magnitude and rules of interaction for assimilation. When these are known, a comprehensive theory for brightness may be formulated in terms of the combined effect of contrast and assimilation. However, even in as primitive a visual behavior as perception of brightness, higher level (for example, form dependent) information may complicate life (cf. Kanizsa, 1979).

It is worth remembering that the Retinex Theory was initially proposed as an explanation for color vision. It was originally thought that the extension of the concept of the Retinex from brightness to color only required repeating the Retinex calculations of brightness (or "lightness") for three independent cone photoreceptor systems (Land and McCann, 1971; Land, 1983). The brightness problem was supposed to be the easy case where all three cone systems provided the same answer in the brightness calculation. It is now clear that the Retinex Theory computation does not account for brightness vision. It is still possible that Retinex Theory helps to explain color vision better than it can account for the phenomena of brightness. However, because color constancy with change of illuminant is not as robust as brightness constancy (Hurvich, 1981), there is some reason to doubt the general validity of the Retinex Theory in color.

Contrast and retinal gain controls

The idea that the visual dependence on contrast was a result of light adaptation rather than
Therefore, the neural mechanisms that lead to lateral inhibition was (as far as I know) first explicitly stated by Whittle and Challands (1969). Helmholtz (1909) probably had a somewhat similar idea when he suggested that simultaneous contrast might really be successive contrast, because of eye movements. That retinal responses depend on the contrast of the stimulus rather than the luminance, and that this is a consequence of the gain control acting on the center of the receptive field, is implicit in the results of many investigators of cat ganglion cell physiology (Enroth-Cugell and Robson, 1966; Sakmann and Creutzfeldt, 1969; Barlow and Levick, 1969; Cleland and Enroth-Cugell, 1970; Enroth-Cugell and Shapley, 1973; among others reviewed in Shapley and Enroth-Cugell, 1984). It is also known that the center’s gain control is not affected by light that falls on the receptive field surround (Enroth-Cugell et al., 1975). Therefore, the neural mechanisms that lead to retinal response dependence on contrast seem to be associated with the neural machinery of the receptive field center of retinal ganglion cells.

The gain controls of light adaptation have sometimes been supposed to be too slow to account for the phenomena of brightness perception, some of which can be perceived in brief flashes. However, it is now known that under the conditions of most experiments, light adaptation is rapid, being 90% complete in under 200 msec for a 1 log unit jump in mean level. There are slow components as well but they only influence precise quantitative measurements. Light adaptation is also very localized in space, a requirement for computing the correct value of contrast on shallow gradients (as in Figs 2 and 3). These properties of retinal gain controls are discussed in detail in our recent review article (Shapley and Enroth-Cugell, 1984) and I will spare the reader the details here.

Parallel pathways for color and contrast in the primate

Our finding of two different parallel systems for color and contrast in the primate visual pathway (Shapley et al., 1981; Kaplan and Shapley, 1982) has been reinforced by the subsequent work of Hicks et al. (1983), Derrington and Lennie (1984), Hawken and Parker (1984) and Blasdel and Fitzpatrick (1984). It also is consistent with the earlier findings of Sperling et al. (1978) and Lee, Creutzfeldt and Elefandt (1979). Schiller and Colby (1983) have recently presented results on responses to luminance and color contrast in the monkey LGN which are also consistent with the idea that the magnocellular pathway is the high gain path and the parvocellular neurons possess low contrast sensitivity.

The single unit studies of contrast sensitivity pathways have been complemented by investigations of the VEP. In addition to the work presented above, Spekreijse et al. (1977) have presented evidence that the “contrast” component of the human VEP possesses very high contrast sensitivity and saturates at very low contrast. Nakayama and Macek (1982) have shown that there are two components of the VEP to contrast reversal in the macaque monkey. One of these components has high sensitivity and saturates at low contrast. The other is a much lower sensitivity component and does not saturate at high contrast. These authors interpret their results as indicating two cortical response mechanisms which get input from the magnocellular and parvocellular parallel pathways separately.

In the literature on visual evoked potentials, “contrast” often is used in a different way from the way I have used it in this paper. For example, see Spekreijse et al. (1977). The Amsterdam group has consistently meant by contrast what I would call the magnitude or absolute value of (what I call) contrast. My definition of contrast has a signature, plus or minus, while what they call contrast is by definition positive. I hope this difference in nomenclature will not obscure the fundamental similarity of results obtained in Amsterdam and New York.

Acknowledgements—I thank my colleagues Ehud Kaplan, R. Clay Reid, James Gordon, and Norman Milkman for their help with work reported here. Our research has been supported by grants from the National Eye Institute of the U.S. National Institutes of Health: grants R01 EY1477 and R01 EY188. For making the Workshop on Systems Approach to Vision possible, I thank the U.S. Air Force Office of Scientific Research and the Netherlands Royal Academy of Arts and Science.

REFERENCES


