MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ARTIFICIAL INTELLIGENCE LABORATORY

AIM-296   January 1974

AN ESSAY ON THE PRIMATE RETINA

David Marr

ABSTRACT

This essay is considerably longer than the published version of the same theory, and is designed for readers who have only elementary knowledge of the retina. It is organized into four parts. The first is a review that consists of four sections: retinal anatomy, physiology, psychophysics, and the retinex theory. The main exposition starts with Part II, which deals with the operation of the retina in conditions of moderate ambient illumination. The account is limited to an analysis of a single cone channel -- like the red or the green one -- the rod channel being referred to frequently during the account. Part III considers various interesting properties of retinal signals, including those from the fully dark-adapted retina; and finally the thorny problem of bleaching adaptation is dealt with in Park IV. The general flow of the account will be from the receptors to the ganglion cells, and an analysis of each of the retinal cells and synapses is given in the appropriate place.

Summary

0. A brief review of the facts known about the retina is presented.

1. It is proposed that one function of the primate retina is to compute lightness by a method derived from the two-dimensional parallel algorithm of Horn (1974).

2. The computation consists of three stages: (1) A centre-surround difference operation, computed in approximately logarithmic units, the result being carried by the bipolar cells. (2) An approximately constant threshold applied to this signal. (3) The inverse transform of (1), performed in the amacrine layer, whose output is lightness. Lightness probably appears at X-cells, which should therefore provide the information for subsequent colour naming.

3. The operation of the midget bipolar-midget ganglion cell channel is analysed in detail. It is shown that the small, stratified amacrine cells are well placed to carry the necessary additive lateral connexions between nearby midget bipolar terminals; and the diffuse amacrine cells, for supplying the necessary subtractive coupling between the two lateral systems in the inner and outer thirds of the inner plexiform layer.

4. In particular it is necessary that:-

   (a) A large proportion of the midget bipolar dyad synapses should be with stratified amacrine cells. All synapses in such a dyad complex, including the amacrine/bipolar synapse, must have a computationally positive sign.

   (b) Diffuse amacrine cells must receive excitation from one layer (from midget bipolar, and possibly from stratified amacrine cells), and must send inhibitory synapses to the other layer, to midget ganglion and to stratified amacrine cells, but not to the midget bipolar axon terminals. The synapses from midget bipolar to diffuse amacrine cells need not be accompanied by a reciprocal amacrine/bipolar synapse, whereas those to a stratified amacrine cell should be.

5. Midget ganglion cells, and perhaps all X-cells, should behave like detectors of lightness. Their centre-surround receptive field organization arises from suitable setting of the DC level of the retinal output.

6. When the illumination falls below a certain minimum level, the lightness computation must be abandoned.

7. If receptors are desensitized over a region, the difference in signal size at its boundary should cause a pseudo-signal to arise from ganglion cells all over that region. This may be important for various properties of bleaching adaptation.

Introduction

It has long been thought that the assignment of subjective "lightness" and "colour" to visible surfaces depends mainly upon the use of comparative, not absolute, measurements of luminance made by the visual system (Helmholtz 1962 (1867)). An anecdotal expression of this opinion may be found in the lecture by Rushton (1972 p27P-31P). No precise questions about how these comparative measurements might be combined to produce global colour or lightness assignments were asked until Land formulated what he called his retinex theory (see Land & McCann 1972). Land's work has however been largely ignored by the main community of retinal physiologists, (e.g. Brindley 1970 makes no reference to his work). This is apparently because it seemed to consist of observations, about simultaneous colour contrast, that have been quite well known for a long time.

What then has Land contributed? His demonstrations and the effect that bears his name are very well-known, but I believe that his most important contribution was to try to quantify the phenomena that he so ably demonstrated. He noticed that the solution of problems in simultaneous contrast is an important issue, and deserves more than hand-waving and neglect. He attempted to approach the problem in a quantitative manner, (Land & McCann 1971), and although his method is unsatisfactory, his contribution is nevertheless valuable. A formal solution to the

two-dimensional problem was obtained by Horn (1974): this paper enquires about its relevance to the primate retina.

This essay is considerably longer than the published version of the same theory, and is designed for readers who have only elementary knowledge of the retina. It is organised into four parts. The first is a review that consists of four sections: retinal anatomy, physiology, psychophysics, and the retinex theory. The main exposition starts with Part II, which deals with the operation of the retina in conditions of moderate ambient illumination. The account is limited to an analysis of a single cone channel - like the red or the green one - the rod channel being referred to frequently during the account. Part III considers various interesting properties of retinal signals, including those from the fully dark-adapted retina; and finally the thorny problem of bleaching adaptation is dealt with in Part IV. The general flow of the account will be from the receptors to the ganglion cells, and an analysis of each of the retinal cells and synapses is given in the appropriate place.

PART I: Review

1 The cells and cell contacts of the primate retina

Most qualitative, and some quantitative aspects of the
structure of the primate retina are well understood, thanks to
the early work of Cajal (1911), and the recent thorough studies
by Missotten (1965), Dowling and Boycott (1966), Boycott and
Dowling (with Kolb) (1969), and Kolb (1970). The primate retina
contains five types of functional elements: the receptors
themselves, the horizontal cells, the bipolar cells, the amacrine
cells and the retinal ganglion cells, whose axons constitute the
optic nerve (see figure 1).

### 1.1 The anatomy of the receptor and outer plexiform layers

At the outermost extent of the outer plexiform layer lie the
primary receptors, the rods and cones, (see Brindley (1970)
chapters 1 & 2). The outermost parts of the receptors contain
the photosensitive visual pigments, and the innermost parts make
contact with other retinal neurones. The rods are smaller and
more cylindrical than the cones, and their bases are the so-
called rod "spherules", that contain a single group of
invaginating processes from lower cells. The bases of the cones
widen out into the cone "pedicles", in each of which there are
between 12 and 25 groups of invaginating processes (Boycott and
Dowling 1969, Kolb 1970) (see figure 1a for the general
arrangement).

The cells whose processes contact the receptors are the so-

Legend to Figure 1

Figure 1a shows the general structure of the primate retina,
redrawn from Boycott & Dowling (1969 figure 98) and Kolb (1970
figure 56). The cones contact the horizontal cell dendrites (HD),
and three kinds of bipolar cell: midget (MB), flat midget (FMB),
and flat bipolar cells (FB). The arrangement of these processes
in the cone pedicles is shown in figure 1c (from Kolb 1970 figure
60). The rods synapse with the horizontal cell axons (HA), and
with the rod bipolar cells (RB) in the manner shown in figure 1b
(from Kolb 1970 figure 59). The bipolar cell axons synapse with
the amacrine cells (A1-A5), and with the ganglion cells (G1-G4):
the different kinds of cells are described in the text. Figure 1d
(from Dowling & Boycott 1966 figure 14) illustrates a dyad
synaptic complex. This is composed of synapses from a bipolar
cell (B) to an amacrine (A) and to a ganglion cell (G), together
with a synapse back from the amacrine process to the bipolar
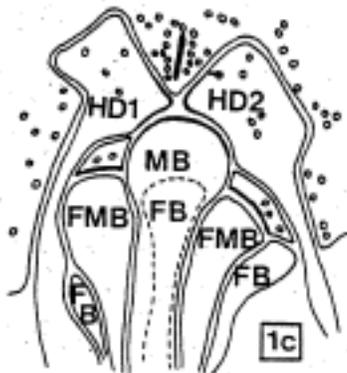axon. In addition, amacrine-amacrine and amacrine-ganglion
synapses are seen.
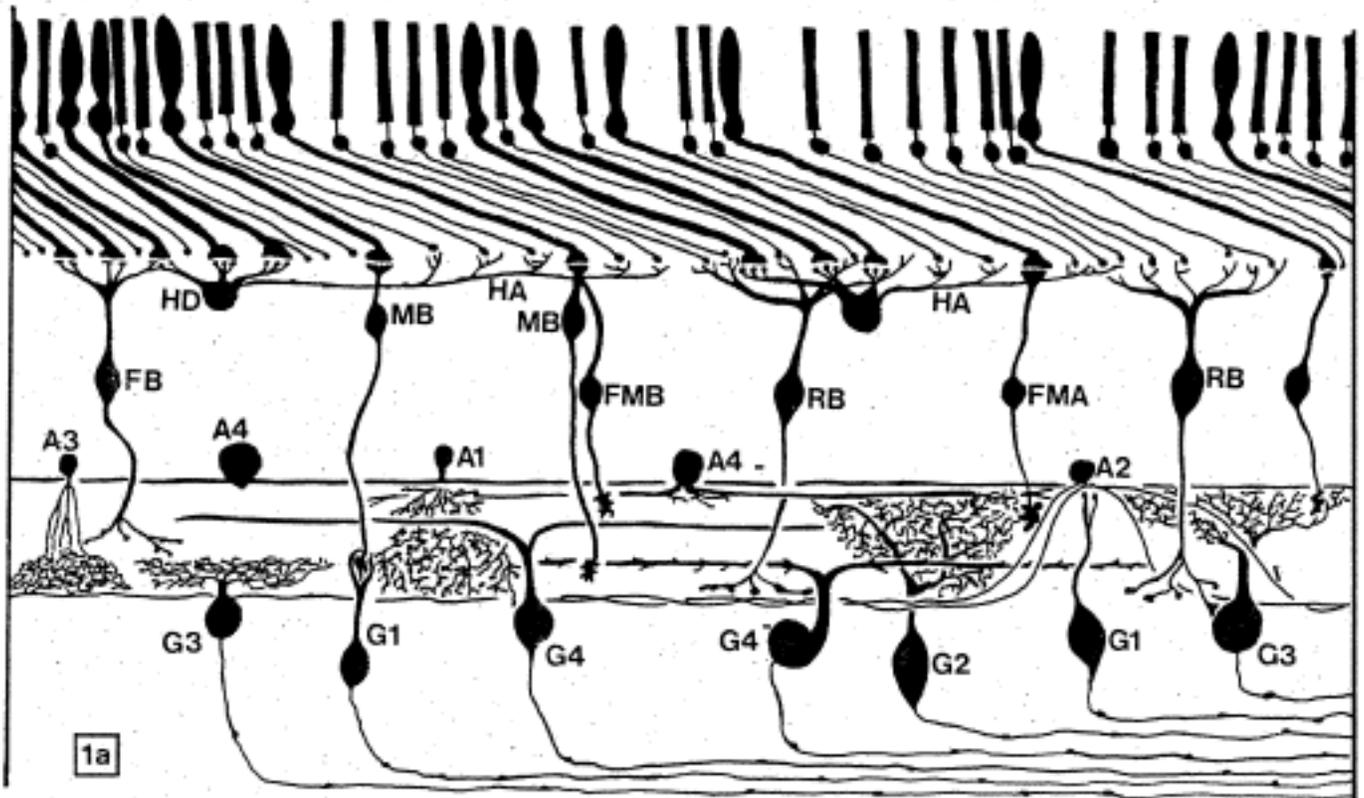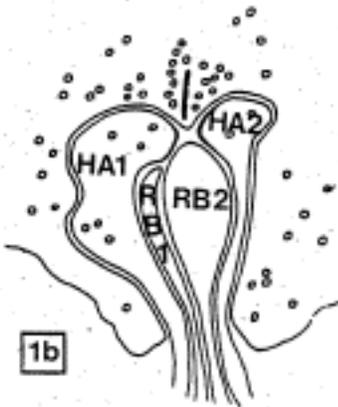
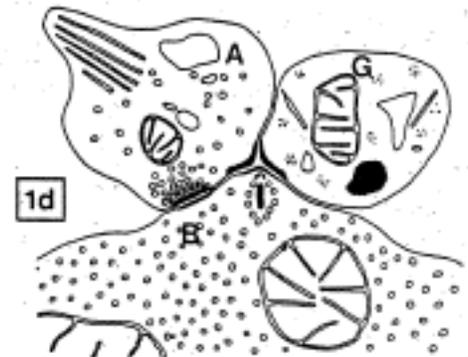FIGURE 1

called "horizontal" and "bipolar" cells. There are probably two
kinds of horizontal cell, and there are four kinds of bipolar
cell. The horizontal cells are distinguished by the size of their
dendritic processes, which contact only cones. The "large"
horizontal cells contact about 12 cones, and the "small" ones,
about 7. The horizontal cell axons contact the rod spherules
(Boycott and Dowling 1969, Kolb 1970): the number of such
contacts made by a single horizontal cell axon is too large to
assess. The horizontal cell axon is extremely thin along most of
its length, and is difficult to stain in the primate. In both
the rod spherules and the cone pedicles, the horizontal cell
processes form the large lateral elements of the invaginations,
and there are always two of them for each invagination. In the
rod spherules, these two always come from different horizontal
cells, and in the cone pedicles, this is almost always true. The
same horizontal cell may however provide one of the lateral
elements to many of the invaginations on the base of a single
cone. The number of different horizontal cells that contact a
single cone is about 6: the number of contacts with nearby
horizontal cells is greater than the number with those lying
further away. The number of horizontal cells that contact a
single rod is always two.

The other contacts with the receptors are made by the
bipolar cells. A firm distinction can be made between rod and
cone bipolar cells. Rod bipolar cells never contact cones, and

vice versa. There are usually two rod bipolar processes
contacting a given rod spherule, and these processes always end
as the central elements of the invaginations (see figure 1b). The
number of rods that have been seen to contact a given rod bipolar
lies between 14 and 45. The only other contacts made by the rods
are with other rods, and with cones: these are discussed below.

The bipolar cells that contact cones can be divided into
three classes: midget bipolars, flat midget bipolars (recently
discovered by Kolb) and flat bipolars. There is one midget, and
one flat midget bipolar to every cone, and each kind of midget
bipolar cell contacts its cone at each of its 20 or so
invaginations (Kolb 1970). Each so-called flat cone bipolar cell
contacts about 6 cones.

The central invaginating process in the cone pedicle is
always formed by the midget bipolar cell. The lateral processes
are horizontal cell dendrites, and adjacent to the midget bipolar
process are the two processes from the flat midget bipolar cell
(see figure 1c); there are therefore twice as many contacts to a
cone from its flat midget bipolar cell as there are from its
midget bipolar. On either side of these, and occasionally also
outside the umbrella of horizontal cell processes, appear the
terminals of flat cone bipolar cells.

The invaginating structure in the rod spherules is similar
(see figure 1).

The ultrastructure of the synapses that we have been discussing is intriguing, and has attracted much attention (see Gray & Pease 1971). In figure 2, there appears a diagram of this synapse obtained from electron microscope studies that were made using a stain that is specific for various parts of the synaptic apparatus. Perhaps the most conspicuous feature of the synapse is the synaptic ribbon (sr in figure 2), which is a half moon structure about 1mu by 0.5mu, and about 0.03mu thick (Gray & Pease 1971), containing a protein (or polypeptide) that is rich in aromatic amino acids (Bunt 1971). A similar structure is present at other retinal synapses where it is also known that there are no action potentials. It is thought that this ribbon is the analogue of the dense projections in conventional synapses, which may be concerned with guiding synaptic vesicles to their appropriate position on the synaptic membrane. Post-synaptic thickenings are visible on the parts of the horizontal cell processes closest to the synaptic ribbon, and there is no reason to suppose that this synapse is fundamentally different from any other chemical synapse, except for the peculiar fact that the presynaptic trigger appears to be a hyperpolarisation, rather than a depolarisation. (This and other evidence has provoked Toyoda, Nosaki & Tomita 1969 to suggest that the vertebrate photoreceptor is actively depolarised by darkness, and passively repolarised in light). The distance between the synaptic gutter (sg in figure 2), and the top of the midget bipolar process is
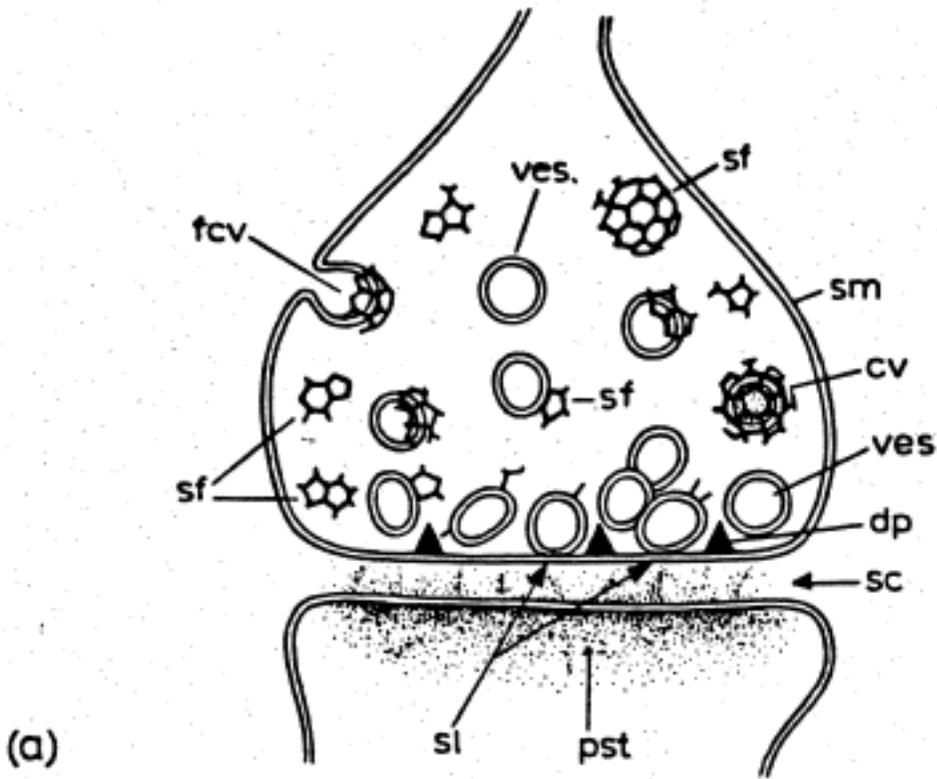
FIGURE 2
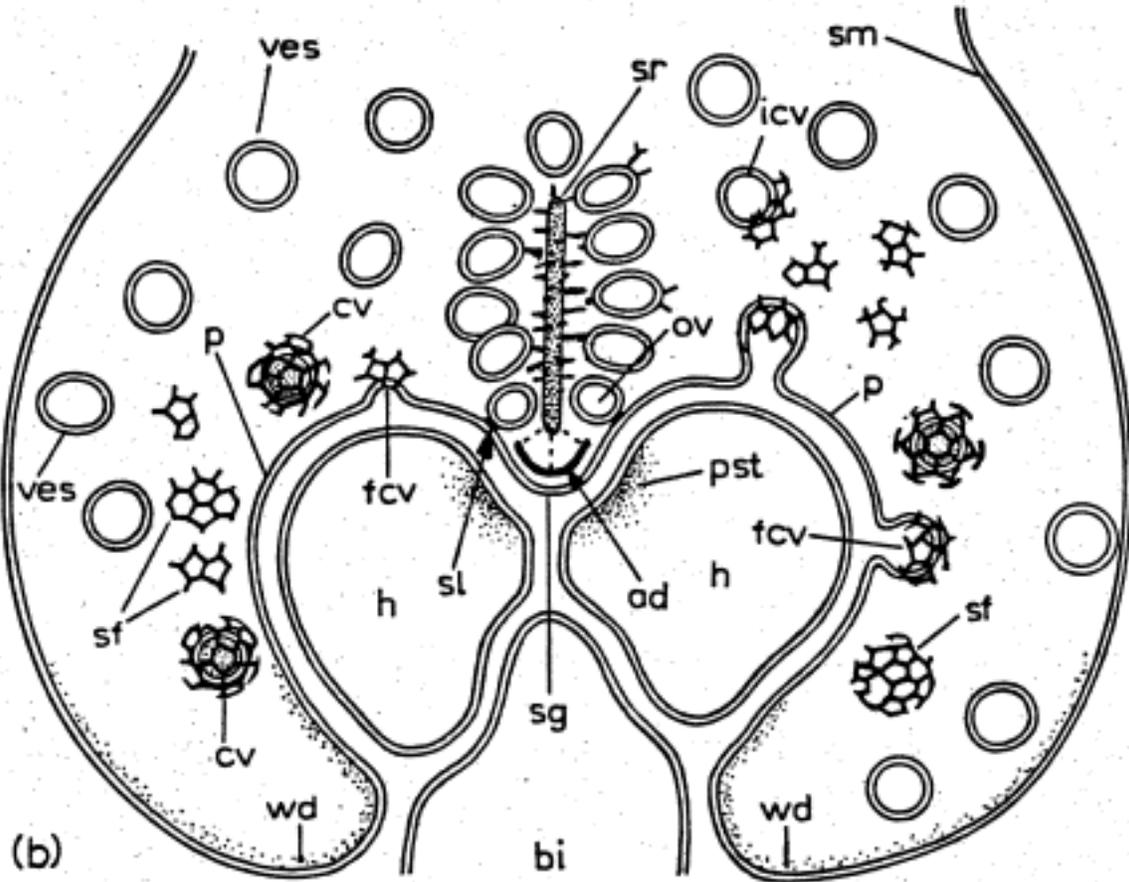
Legend to Figure 2

Figure 2a shows a diagram of a section of a synapse from the
mammalian cerebral cortex, and 2b, of a retinal receptor synapse.
The ribbon has been cut transversely. The abreviations used are
as follows: ad, arciform density; b or bi, bipolar cell process;
cv, complex (coated) vesicle; dp, dense projection; fcv, forming
complex vesicle; h, horizontal cell process; icv, vesicle with
shell portion still atached (incomplete complex vesicle); ov,
outer row or ribbon vesicles; p, pouch or pouch surface membrane;
pst, postsynaptic thickening; sc, synaptic cleft; sf, shell
fragments; sg, synaptic gutter; sl, specific localities; sm,
surface membrane; sr, synaptic ribbon; ves or sv, synaptic
vesicle; wd, dense undercoating wall of presynaptic bag. (From
Gray & Pease 1971, figure 1).

about 0.3mu, and the diffusion time for transmitter across such a
distance is small compared with the time for the response of a
cone to light.

Some receptor-receptor contacts are visible in most
preparations, but their number is very sensitive to the
particular staining or embedding techniques used (Dowling &
Boycott 1966, Missotten 1965 p.58). Such contacts may not be
functional in primates.

1.2 The inner retinal layers

The bipolar cells join the outer plexiform layer to the
inner. The two types of retinal cell that reside here are the
amacrine cells, and the retinal ganglion cells, which synapse
with each other, and with the bipolar cells. The fine details of
the structure of the inner plexiform layer are not as well known
as are those of the outer, but a certain amount of information is
available.

Boycott and Dowling (1969) have reviewed the previous
literature, and have described the following types of amacrine
cell in the primate retina. The word "diffuse", in this context,
refers to processes that are distributed perpendicular to the
sclera, and the word "stratified" is used to mean layered
parallel to the sclera. All the cells that have been seen in the
primate retina have roughly circular symmetry. The diameters in
what follows refer to the diameter of the spread of the amacrine

cell processes, or of the ganglion cell dendrites.

1.2.1 The amacrine cells

(A1) Narrow field diffuse amacrine cells, having a diameter of
10-50mu., average about 25mu., found all over the retina.

(A2) Wide-field diffuse amacrine cells, having processes that
spread out gradually as they descend to the level near the
ganglion cell bodies, and spread out there to attain a diameter
of up to 600mu. These cells are particularly likely to synapse
with rod bipolar terminals, and are unlikely to contact the
ganglion cell bodies.

(A3) Stratified diffuse amacrine cells, having a diameter of 20-
50mu., are restricted to the top, middle, or to the lower third
of the inner plexiform layer, but are diffusely distributed
within one of them. A given stratified diffuse amacrine cell
probably makes frequent, but not exclusive contact with a
particular ganglion cell that has its dendrites similarly
distributed.

(A4) Unistratified amacrine cells, whose diameter lies between
100 and 1000mu., extend their processes in the plane immediately
corneal to the inner plexiform layer.

(A5) Bistratified amacrine cells, with a diameter of about
100mu., send horizontally distributed processes to the planes
corneal and scleral to the inner plexiform layer.

1.2.2 The ganglion cells

(G1) The midget ganglion cells are of two kinds, one with terminals in the outer third of the inner plexiform layer, and the other with terminals in the inner third. The middle third seems to be free of midget ganglion cell terminals. This fits well with the known distribution of the midget bipolar terminals (see above). There is probably a one-to-one correspondence between midget bipolar and midget ganglion cells.

(G2) Diffuse ganglion cells, dendritic diameters ranging from 30-75mu, the smaller diameters occuring nearer the fovea.

(G3) Stratified diffuse ganglion cells, like the stratified diffuse amacrine cells, are diffuse within the outer, middle, or inner third of the plexiform layer. There may be more in the outer third than in either of the others. Diameters range from 40mu near the fovea, to 80mu in the periphery.

(G4) Unistratified ganglion cells, occurring at all levels, have a diameter of about 200mu.

1.2.3 The synapses of the inner plexiform layer

The most common synaptic complex found in this region of the retina is the so-called dyad synapse, (see figure 1). At a dyad synapse, a bipolar cell contacts both a ganglion and an amacrine cell, and close by there is (probably) a further synapse from the amacrine cell back onto the bipolar terminal (Dowling and Boycott (1966)). In addition to the dyad synaptic complex, amacrine to

amacrine, and amacrine to ganglion dendrite synapses are seen.

The proportions in which the various types of synapse occur.
in the human retina are roughly as follow:

the complex of dyad + amacrine to bipolar: 3

amacrine to amacrine : 1

amacrine to ganglion : 1

bipolar to amacrine soma : 1/12

(from Dowling and Boycott 1966 table 1).

2 Aspects of Retinal Physiology

It would be absurd to try to summarise the whole of the
physiology of the retina when such an excellent monograph as
Brindley's (1970) already exists: but it is equally impossible to
omit any summary altogether, since the emphasis of the knowledge
that is required differs somewhat from that of Brindley's
account. Furthermore, a number of important results have been
published since 1969, and it is convenient to collect them
together for ease of reference.

The photochemistry of the retina is fairly well understood,
(see Brindley 1970 chapter 1), but the means by which the
chemical events provoke a signal in the retinal neurons, and
indeed the nature of the important part of that signal, remain
somewhat obscure. The most likely candidate, supported by recent

observations of Penn & Hagins 1969 (rods), of Tomita, Kaneko,
Murakami & Pautler 1967 (cones), and reviewed by Brindley (1970
pp50-54), is that photochemical events affect the membrane of the
outer segment of the receptor: this causes a change in
permeability to potassium and/or to sodium; this causes a
hyperpolarisation of the outer segment, which affects the inner
segment by passive conduction, and the resulting signal causes
synaptic transmission to the underlying retinal neurones. The
difficulties with this theory are firstly, that the size of the
hyperpolarisation at the inner segment of a rod in response to
one quantum would be about 100 microvolts: this is a tiny signal.
And secondly, it is even then unclear how a single quantum can
produce the required changes in the membrane of the outer
segment.

It is fortunate that the retinal cells of the mud-puppy
Necturus maculosus are unusually large, for this has enabled
Werblin and Dowling (1969) to obtain intracellular records of
many types of retinal cell (see also Brindley 1970 chapters 2 and
3). These authors confirm and establish a number of important
facts, and these are summarised next.

2.1 The receptors

The receptor response is a hyperpolarisation, as was already
known, that consists of an initial transient which decays to a
steady level; see (Kaneko & Hashimoto 1967, Tomita 1968, Naka

1969, Toyoda et al. 1969, Werblin and Dowling 1969). When
illumination is removed, there is a small "off" transient, and
the potential returns very slowly to the base line. The latency
of the response is long - about 50 msecs - and its magnitude is
about 5 mV, a typical resting potential being 30 mV.   There are
two important points to note about the receptor response:
firstly, it seems to obey the relation

$$V/Vmax = I/(I + K)$$

(see Naka & Rushton (1966, 1967), Naka (1969), and papers cited
in section 3). K is about 800 quanta/rod/sec for humans). This is
consistent with Werblin & Dowling's report that the response is
almost linear over about 2 log units after which it begins to
saturate, because the stimuli have to be relatively strong in
order to produce an effect on the receptor that can be detected
with an intracellular micro-electrode.   Secondly, and very
importantly, it is independent of the illumination of
neighbouring receptors (see also Tomita (1968) in fish).

2.2 The horizontal cells

     The horizontal cell response, like that of the receptors, is
a sustained hyperpolarisation that is graded with intensity over
about 3 log units. It has a latency of about 100msecs in the
mudpuppy. From the fact that the response of the horizontal cell
is greater than that of the receptors, and saturates at higher
intensity when annular rather than spot stimulation is used,

Werblin and Dowling concluded that the response is probably formed through the weighted summation of many sites, each of which can be saturated.

2.3 The bipolar cell response

The bipolar cell receptive field is organised into two concentric, antagonistic zones. Werblin and Dowling report that about half of the units hyperpolarise to central illumination, and about half depolarise; but that illumination of the periphery is capable only of removing part or all of the signal due to the centre, not of producing a signal in the opposite direction. According to Werblin and Dowling, the magnitude of the bipolar response can be held constant for a fixed ratio of centre-to-surround illumination over a wide range of absolute intensities. Since the latency of the surround effect is about 100 msecs greater than that of the centre, however, the bipolar response consists of an initial transient to any change in illumination even if there is no change in contrast.

2.4 The amacrine cell

Hitherto, all responses have consisted of graded slow potentials: the amacrine and the ganglion cells are however capable of sustaining action potentials. The threshold for an amacrine cell spike is at least as low as that of the earlier cells, but only if the intensity change is sudden. By varying it

slowly, one can increase the intensity many times without
provoking a regenerative response. It is particularly interesting
that Werblin and Dowling were able to find two kinds of amacrine
cell response: some units had very broad, uniformly sensitive
receptive fields, and responded at "on" and at "off" to
illumination of any part of the field. Others had narrow centres
(100-200mu) and larger surrounds; these responded at "on" to
central illumination and at "off" to peripheral illumination. To
diffuse illumination, they apparently responded at both "on" and
"off".

2.5 The ganglion cells

The traditional picture of the response of a cat's retinal
ganglion cell is that presented by Kuffler (1953): (see Brindley
(1970 pp81-89) for a review of earlier work). According to this
picture, the receptive field of a ganglion cell is circular, and
is divided into two concentric, antagonistic regions. The size of
the centre ranges between about 0.1 and 2degs of visual angle,
and that of the surround is somewhat larger. The size of the
dendritic trees of the retinal ganglion cells agrees closely with
that of the centres of the receptive fields (Brown & Major 1966,
Boycott & Dowling 1969). Barlow, Fitzhugh & Kuffler (1957) found
that when dark-adaptation is nearly complete, the surround effect
disappears quite suddenly, and all lights that have any effect
upon a cell have the same effect. This change is not linked to

the change from cone to rod vision that gives rise to a number of other phenomena.

The true picture is probably rather more elaborate than this. In the light-adapted cat's retina, three substantially different kinds of ganglion cell response can be recorded. Their discoverers, exercising perhaps undue academic restraint upon the nominative process, christened them X, Y (Enroth-Cugell & Robson 1966) and W cells (Rodieck 1967, Stone & Hoffmann 1972). These categories are distinguished in a variety of interesting ways. A rough grouping is by conduction velocity of their axons in the optic nerve: Y cell axons are the fastest, (35m/sec in the periphery, 22m/sec centrally); the X cells are next (18m/sec in the periphery, 10m/sec centrally); and the W cells are the slowest (usually <10m/sec) (Stone & Freeman 1971).

2.5.1 Properties of W cells

W-cells, noticed first by Rodieck (1967) and seen again by Fukada (1971, his two "unclassified" cells), were studied more thoroughly by Stone & Hoffmann (1972). There are two kinds of W-cell; the less common one, first described by Rodieck, shows spontaneous activity that is suppressed by almost any change in contrast. It appears that such cells can be stimulated by a fast object leaving the visual field at about 200 deg/sec, but no other stimulus has been found that raises their firing rate. The more common kind of W-cell is excited by any change of contrast.

A typical receptive field is 0.9 deg in diameter, and the cell

responds to a 0.2 deg spot flashing anywhere in the receptive

field.   W cells may constitute a large fraction of the retinal

output, and their axons probably project to the superior

colliculus.   (Some unpublished work by J.Stone and collaborators

shows that at least some W cells are specific for stimuli moving

in a certain direction in the cat. Thus the distinction between

the cat and the rabbit retina may be one of degree, rather than

kind.)

2.5.2 X-cells and Y-cells

The original distinction between X and Y cells was reported

by Enroth-Cugell & Robson (1966).   It has been confirmed by a

number of workers, who have added to the list of discriminating

properties (Fukada 1971, whose type I and type II cells are

respectively Y and X cells, Fukada & Saito 1971, Cleland, Dubin &

Levick 1971). X and Y cells may be distinguished in the following

seven ways.

(XY1) Response to standing contrast

If a stationary sinusoidal grating pattern is introduced and

withdrawn at various phase angles, the responses of X and of Y

cells are quite different. (See figure 1 of Enroth-Cugell &

Robson 1966).   The X-cells give a short transient response when

the grating is introduced, and the transient decays over about

200msecs to the sustained level, which is maintained.   The

response of the X-cell is easily understood if it is regarded as
the sum of a central and a larger antagonistic surround region,
both with approximately gaussian weighting functions. The
sensitivity of an X-cell, but not its response magnitude, behaves
like a linear function of the appropriately weighted energies
incident on these regions. Consequently, its behaviour at a
phase angle of 0 deg is the reverse of its behaviour at 180 deg,
and the responses at 90 and at 270 deg are both null.

The response of the Y-cell, on the other hand, consists
entirely of a non-linear transient. The mean discharge frequency
for Y-cells is greatly increased when grating patterns drift
across their receptive fields: this is not the case for X-cells.
(XY2) Response to fine, moving gratings

In agreement with their essentially linear behaviour, as a
grating pattern moves across the receptive field, the X-cell
response is modulated about a mean, which gradually tends to zero
as the spatial frequency of the grating is increased. For a Y-
cell, on the other hand, the moving grating provokes an
unmodulated increase in discharge level that persists while the
movement continues.
(XY3) After effects of stimulation

After stimulation by a moving grating, as described in
(XY2), the Y-cell response returns to its resting level. That of
the X-cell, on the other hand, is depressed for as much as 30
secs.

(XY4) Size of receptive field

The optimal size of an X-cell stimulus is usually 0.5 to 1.0 degs in diameter. That of a Y-cell is 1 deg or more: no Y-cell has been seen with an optimal size of 0.5 deg or less.

(XY5) Speed of moving stimulus

If a contrasty target is moved very fast (200deg/sec) across the receptive field, X-cells fail to respond, but Y-cells do respond.

(XY6) The effect of flicker

The average impulse frequency of X-cells is insensitive to flicker over a large range of frequencies. That of a given Y-cell is unimodal, centred on some maximum frequency value. The critical fusion frequency for a Y-cell is positively correlated with conduction velocity, and with the maximum impulse frequency for that cell: neither of these is true for an X-cell.

(XY7) The periphery effect

McIlwain's (1964, 1966) periphery effect, whereby retinal ganglion cells may be excited by moving stimuli that are far outside their receptive fields as conventionally defined, was present for all Y-cells, but weak or absent for X-cells (Cleland et al.1971). The form that this response takes is that any continuous movement of a large object far (say 15 deg) from the cell's receptive field, raises the cell's maintained discharge rate.

2.5.3 A note on the fate of X and of Y cell axons

The distinctions between X and Y cells are apparently
preserved by the lateral geniculate nucleus (Stone & Hoffmann
(1971), Cleland et al. (1971)). (In particular, the conduction
velocities of optic nerve axons that converge on a given
geniculate cell are astonishingly close (0.1 m/sec) - a degree of
precision that must have implications for how this particular
developmental problem is solved.) Perhaps more surprising is the
way these cells terminate in visual cortex. According to
Hoffmann & Stone (1971), at least 40% of, and perhaps all,
complex field cells in those parts of the cat's visual cortex
from which they recorded (area 17 and the 17-18 boundary), are
monosynaptically driven by fast afferents from the geniculate.
These are geniculate cells that are driven by retinal ganglion
cells of class Y. They also found that cells with simple, or with
hypercomplex, receptive fields are not discharged by fast
afferents, and a proportion are discharged monosynaptically by
slow afferents. Hoffman & Stone's finding contradicts that of
Denney, Baumgartner & Adorjani (1968), but it is probably
correct, because of collateral evidence from later studies. In
particular, in an investigation of areas 17 and 18 using
antidromic stimulation, Stone & Dreher (1973) found that many
geniculate Y cells project to both 17 and 18, whereas geniculate
X cells project only to area 17. It seems that there are two,
parallel pathways to complex cells: Stone (1972) has discussed

the relationship of his and Hoffmann's finding to the work of
other investigators.

2.6 The independence of the rod and of the three cone channels

There is some evidence that the rod and the three cone
channels are processed independently (see articles by Alpern
1965, Gouras & Link 1966, Gouras 1966, 1967, Alpern Rushton &
Torii 1970a and d, Westheimer 1970, Westheimer & Wiley 1970, and
McKee & Westheimer 1970). Recent papers (Lennie & MacLeod 1973,
Barlow & Sakitt 1973, see also Brindley 1970 pp75-86) cast doubt
on a number of these findings, however; and although there is
little information available about chromatic interaction in the
primate retina (Hubel & Wiesel 1960), it is probable that some
interaction is visible in primate retinal ganglion cells, and
certain that such interaction takes place in the lateral
geniculate (De Valois 1965, Hubel & Wiesel 1966).

2.7 Gain and latency studies on ganglion cells

In two papers, Cleland & Enroth-Cugell (1968, 1970) have
started an accurate study of the characteristics of the ganglion
cell response. They found that the weighting function for the
central excitatory area of a ganglion cell had the form of a
plateau (the "uniform centre") with exponentially decreasing
sensitivity over a surrounding annulus. They also showed that the
threshold for sinusoidal stimuli at various diameters depends

only on the characteristics of the centre, and that adaptation

effects for such stimuli depend on summation effects acting over

the same areas as the centre.

The finding of the weighting curve for the way incident flux

summates over the central region enabled Cleland & Enroth-Cugell

(1970) to define the concept of "effective flux". The effective

flux of an incident light is that flux which would have the same

effect if it all fell in the uniform centre - that part of the

central area where summation is linear. They found that, when a

cell responded with a pure central response to an incremental

flux, applied against a steady background, the gain and the

latency were a function of the total effective flux of the

increment and the background. In other words, the field

adaptation level is set very quickly for such stimuli - it is set

within the response time.

It is interesting in this connection to recall Naka &

Rushton's (1968) experiments on the effect of adaptation upon s-

potentials. Apart from their main finding, that the effects on s-

potentials of after-images and of real backgrounds differ, they

found that the effect of adaptation is to attenuate the receptor

signals in some fixed manner before they ever reach the s-units

horizontal or bipolar cells, see Brindley (1970 p.78). Thus an

important part of the retinal gain-setting mechanisms lie in the

outer layers.

Enroth-Cugell and Pinto (1972a,b) continued the study of

ganglion cells by studying pure surround responses. Pure surround
responses could be recorded from only half of the tested cells,
for reasons of varying certainty; but on 6 or 7 occasions, pure
centre, and pure surround, and the combined responses were
obtained for the same cell (a long and exacting experimental
process). They found that on these occasions, the centre and the
surround responses interacted in a way that was probably
algebraic, (Enroth-Cugell & Pinto 1972a).

3 Dark Adaptation

The literature on dark adaptation is extensive, and has
become particularly controversial in the last few years. This
summary can start by referring the reader to a paper by Barlow
(1964), in which an excellent review of previous work is given,
and one of the two rival modern hypotheses is proposed.

Very briefly, the history of the subject up to that time is
as follows. The oldest quantitative relation, known since Weber
(1834), is the so-called Weber-Fechner law (see figure 3). This
relates the increment threshold to the intensity of the
background illumination in the following way:

$$t/t_o = (I + I_o)/I_o \qquad (1)$$

where t_o is the absolute threshold, t the increment threshold,
I is the intensity of the background illumination, and I_o is a

constant, associated with the background noise levels in the
retina. The conditions under which the Weber-Fechner law is true
are somewhat complex (Barlow 1957), but roughly speaking, it is
true for stimuli of large area (5 degrees) or of long duration (1
sec) presented on a background that is at least 1000 times
absolute threshold. For stimuli that are both short and small,
on backgrounds that are less than about 10,000 times threshold,
the measurements are well fitted by the relation

$$t/to = ((I + Io)/Io)**1/2 \qquad (2)$$

which Barlow interpreted as the optimal criterion for extracting
a signal from a noisy random variable with mean (I+Io)/Io.   The
relation (2) for small signals is pleasing, because it can be
understood: relation (1) is however altogether surprising,
because almost all other sensory modalities have threshold
relations which involve the log of the background signal. As we
shall see, this reflects the fact that over an enormous range,
the output from the visual receptors is approximately linear with
the incident energy.

The second quantitative relation that was known since the
work of Crawford (1937, 1947) was that varying degrees of
bleaching of receptors, and varying levels of background
illumination, have very similar effects on the measurement of
increment threshold for a flash (see figure 3).

The theory of the effect on increment threshold of an
immediately preceding bright light had already passed through a

Legend to figure 3

Increment threshold plotted against time since a bleaching
stimulus, and against the intensity of the background
illumination. The straight portion of the latter curve is the
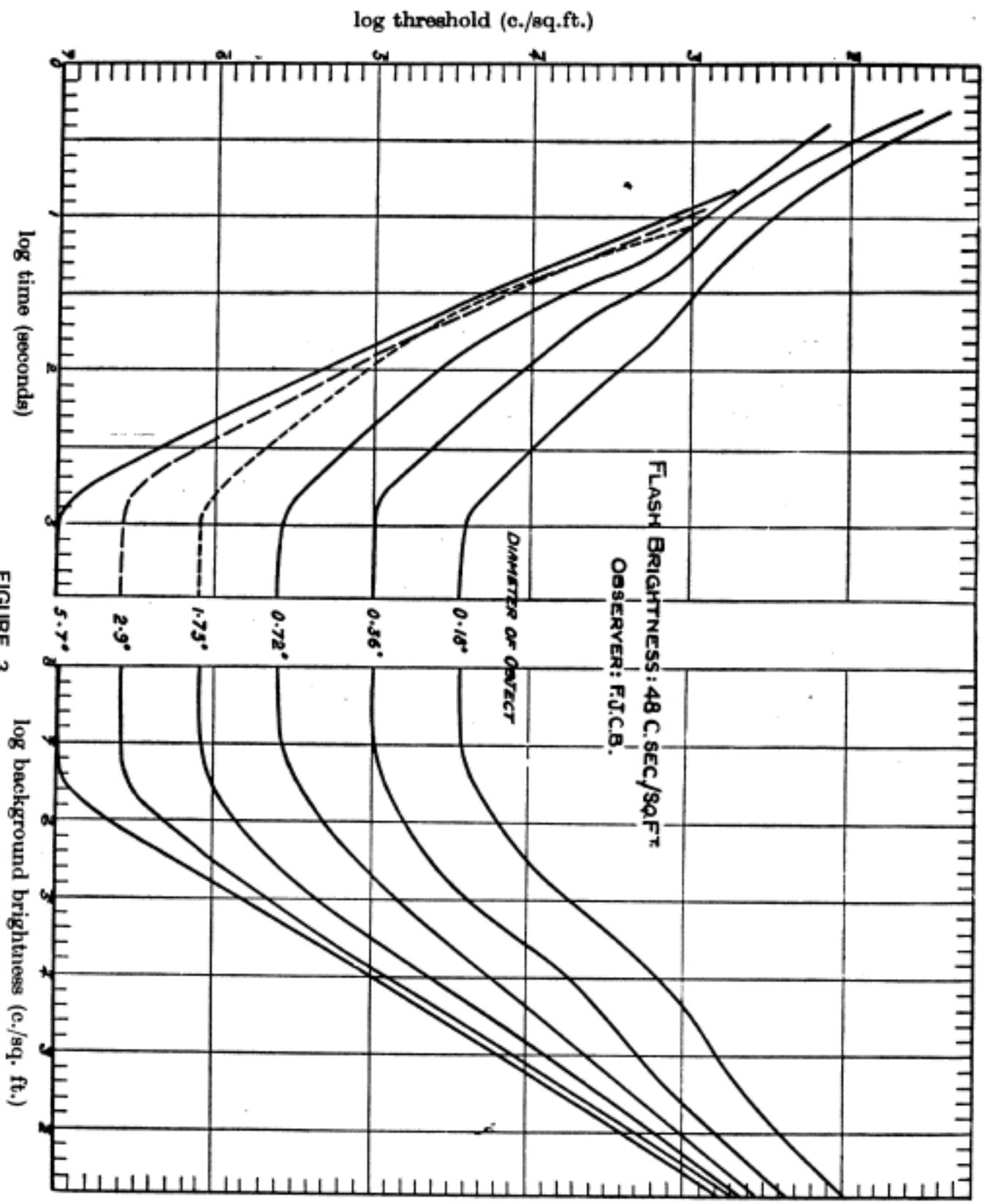Weber-Fechner relation. (From Crawford 1947, figure 7.)

FIGURE 3

log threshold (c./sq.ft.)

log time (seconds)

log background brightness (c./sq. ft.)

FLASH BRIGHTNESS: 48 C. SEC./SQ.FT.
OBSERVER: F.J.C.B.

DIAMETER OF OBJECT

5·7°
2·9°
1·75°
0·72°
0·36°
0·18°

number of stages by this time. The earliest serious attempt to
think about it was probably the simple photochemical theory,
whereby the raised increment threshold was attributed to the
reduced sensitivity of receptors after bleaching by light.
According to this theory, if t is the threshold when a
proportion B of the rhodopsin is bleached, and to is the
absolute threshold, then $t/to = 1/(1-B)$. This theory was
already sufficiently discredited in 1938 for alternative
explanations to be sought, and was finally most directly
disproved by Rushton (1961), using a technique whereby the
fraction of rhodopsin that was bleached was measured by direct
observation of the living retina.

In 1940, Lythgoe observed that in order to cause a very large
change in threshold, only a small amount of pigment needs to be
bleached: in fact, when $B=1/2$, the increment threshold for cones
in humans lies between 50 and 500 times the absolute threshold,
depending on the size of the test flash. From about this time,
therefore, adaptation was held to be organised by a neural
mechanism of a suitable, but unspecified nature; and this view
had the temporary effect of releasing investigators from any
obligation to define this mechanism with any precision.

Lythgoe's observation about the disproportionate effect of a
small amount of bleaching naturally raises two questions:
firstly, does the threshold of an individual receptor (rod say)
change very much for moderate (say 20%) degrees of bleaching? And

secondly, are the outputs of receptors pooled in some way,
threshold decisions being then taken in the light of information
about the activity in the pool as a whole. The neatest (though
not the earliest) answers to these questions were provided by
Rushton & Westheimer (1962), and by Rushton (1965a). It was shown
firstly that the threshold for perception apparently changes
drastically even when the flash falls on rods that received few
or no quanta during the bleaching exposure. Hence the threshold
setting process must be the result of a summation over some kind
of pool, whose result operates in a uniform way on signals from
the receptors that contribute to that pool. Secondly, using a
grating as a bleaching and as a test stimulus, and by varying the
phase relation between the two situations, it was shown that
there is no apparent difference between the responses of rods
that have and have not been previously bleached. The accuracy of
the grating experiment of Rushton & Westheimer (1962) has however
been questioned recently (Barlow & Andrews 1973), who found that
although there exists some elevation of threshold by pooled
signals, it is nowhere near as large as previously reported.

In a welcome reaction to the imprecision of the neural
"theories" of the 1940's, Wald (1955,1957) proposed a new and
quantitative theory of adaptation. It was essentially a
'compartmental' version of the older photochemical theory.
According to this idea, there are a number of compartments in
each of which there is a chance B that a given molecule of

rhodopsin will be bleached by exposure to the light-adapting light. B is then the fraction of rhodopsin bleached. If k is the number of molecules in each compartment, then the probability p that no molecule is bleached is $(1-B)^{**}k$. It is assumed that if a molecule is bleached, the whole compartment becomes refractory: hence $p=t/t0$.

  i.e. $t/t0 = p = (1-B)^{**}k = \exp(kB)$ for small B.

   This theory is precise, attractive, and generated many fruitful experiments - evidence of inherent virtue firm enough to withstand even the fact that those experiments also effectively disproved it. Barlow (1964) lists the reasons why the theory cannot now be held, and points out that its valuable legacy was to cause the relation

$$\log (t/t0) = kB \qquad (3)$$

to be established empirically over a large range of B   (in particular for B near 1, which was not predicted by the compartmental theory) (Dowling 1960 in the rat, Rushton 1961 in man).

   Barlow then presented his own characteristically elegant appraisal of the situation. The problem is to explain the formulae (1) and (3). In effect, his hypothesis is that the receptors are noisy devices, and that when they are bleached, they become yet noisier. A receptor that is bleached emits a signal that is indistinguishable from that which it emits when receiving light of an intensity Ib, where $Ib=\exp(kB)$, for some

constant k. This explains equation (3) above; and if we make the
further hypothesis that in the case where there is real light of
intensity I, falling on a receptor that is bleached an amount
B, the two signals should add, then the increment threshold t/to
should be given by the formula

$$t/to = (I + Ib)/Io, \quad where Ib = Io.exp(kB).$$

This is found to be correct. Perhaps the most convincing evidence
for this theory is however contained in the results of Barlow and
Sparrock (1964). In these experiments, subjects had to match the
brightness of an after-image against the brightness of a real,
stabilised background illumination. It was found that the results
from this experiment matched those from increment threshold
curves obtained from real and after-image backgrounds. These
results, and those of a number of other investigators, including
the remarkable findings of Crawford (1937, 1947) on the
equivalence for increment threshold of bleaching and background
illumination, are therefore explained very neatly by Barlow, and
all are agreed that the new theory captures many facets of the
experimental results.

But not all of them. Before discussing the exceptions, it is
convenient to note three facts about Barlow's theory. Firstly,
the high rise in threshold that is provoked by relatively small
amounts of bleaching is attributed to an inherent property of the
receptors - their unfortunate increase in noisiness. Secondly,
Ib, the equivalent background light produced by this noisiness,

is imagined to be totally indistinguishable from the signal
produced by a real background light (stabilized on the retina) of
intensity Ib. There should therefore be no observable
differences between these two cases, and in many respects this is
true. Thirdly, Barlow himself comments on the odd fact that the
intensity Ib of the pseudo-signal varies with the exponent of
the fraction B of bleached pigment. Barlow mentions some
possible mechanisms for achieving the exponentiation (his p.57)
but is clearly not completely happy about the matter.

Meanwhile, Rushton had been pursuing his investigation along
somewhat different lines. Fuortes and Hodgkin (1964) analysed the
potentials recorded from cells in the ommatidia of limulus, and
defined a model that accurately describes the values that they
found. It turns out that the same model describes threshold
relations in man, and it is therefore important to sketch it here
(see Rushton 1965 FL). In this model, the attenuating mechanism
for signals from the receptors is regarded as being a leaky cable
of length s and leakiness a' (see figure 4a). The signals from
the receptors (let us consider the rods for now) pass through
this cable, and the output serves two functions: one is to form
the transmitted signal of the system, and the other is as a
feedback that determines the value of the leakiness a'. If the
output of the system is V', then the equation of the system
(assuming that the cable is open-ended and $\exp(-2a's) \ll 1$) is

$$V' = I'\exp(-sa') \qquad (4)$$

(a)

β

$V + \beta$

I

V

$V + \beta$

$\langle ---- \ S \ ----\rangle$
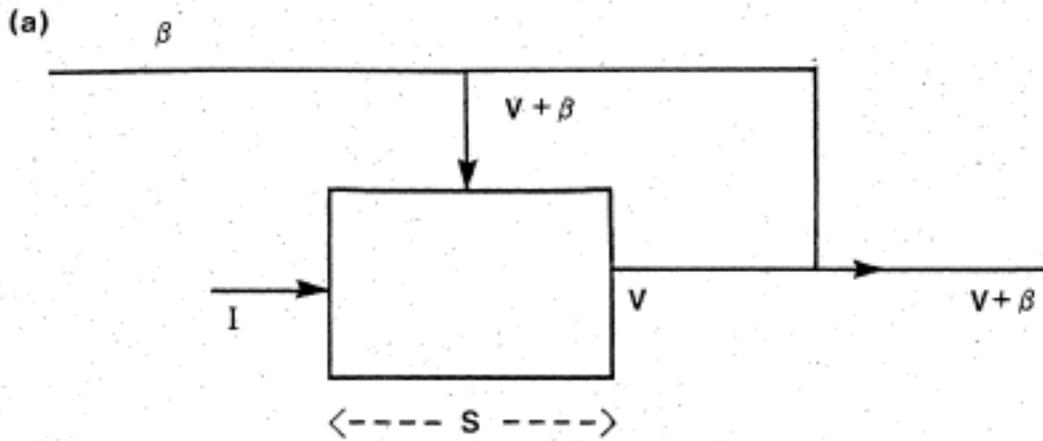
cable length s(variable) leakiness a

(b)

$8°$

$2°$

λ

$[\mu]$

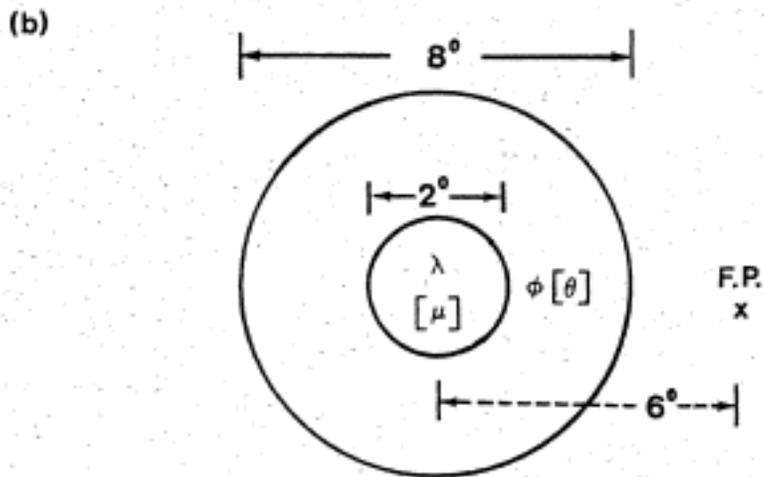$\phi[\theta]$

F.P.

x

$6°$

FIGURE 4

Legend to figure 4

Figure 4a shows the cable model drawn upon in the text: figure 4b illustrates the stimulus for Alpern's contrast-flash effect. The test flash λ on background μ, occurs with surround flash $\wedge$ on a surround background θ. F.P. is the fixation point, causing λ to fall 6degs temporal to the fovea.

I' is the input signal (which varies directly with intensity over
a considerable range) from the receptors. Inserting the feed-
back, we get a'=a0+a1V1, where a0 and a1 are constants. After
appropriate choice of units, this becomes

$$I = Vexp(V) \qquad (5)$$

Rushton then examined the predictions that follow from this
model in the various threshold experiments. He first shows that
the Weber-Fechner law can be derived (but only approximately):

$$dI/dV = exp(V) + Vexp(V) = exp(V) + I \qquad (6)$$

Assume that I attains threshold by causing dV to reach some
value dVo: then with I = 0,

$$to = dIo = dVo.exp(0) = dVo$$

Now (here is the disturbing approximation) over the
relevant range, it turns out that to within 0.1 log units,

$$log(exp(V) + I) = log(1+1.2I)$$

Hence t/to = 1+1.2I   (t is the increment threshold),
which is the Weber-Fechner law in slightly changed units.

So far then, so good. The interest of the model however
attaches to its treatment of bleaching. Rushton points out that
the effect of having bleached a fraction B of rhodopsin is
equivalent, for threshold purposes, to a background light of
intensity Ib, where (I0+Ib/I0)=exp(kB) for some constant k.
Remember that this was the observation of Crawford's that formed
the central pivot of Barlow's theory. As Rushton repeated, it is
most natural to assume that the equivalent background light Ib

enters the system through the same path as before, and with intensity $I0(\exp(kB) - 1)$. But in a conversation with Rushton, Hodgkin apparently pointed out that a more elegant formulation results if one assumes that the effects of bleaching enter the system through the _feed-back_ path, rather than through the input. The delight of this idea is that it removes the irritating necessity for raising the bleaching signal exponentially.

In this case, the expression (4) becomes

$$V' = I'\exp-(sa' + ra'') \qquad (7)$$

where r is the signal from the bleached rhodopsin entering the feedback added to V1. Repeating the derivation above, one obtains

$$I = V\exp(V + \beta) \qquad (8)$$

(where β=about 45B in humans after suitable choice of units), and

$$dI/dV = I + \exp(V + \beta) \qquad (9)$$

which reduce in the unbleached condition (B = 0) to equations (5) and (6).

One of the most attractive features of Barlow's formulation, in which the bleaching signal enters through the same path as the light signals, is that the equivalent background due to bleaching is expected simply to add to any luminous backround that is present. Rushton notes this, and attacks with vigour the question of what his new hypothesis predicts. His argument, a construction of some elegance, is the following. Consider two states, represented by subscripts 1 and 2, that give the same output (V1=V2). In state 1, there is bleaching B1, and in the

second, full regeneration has occurred (B2=0). Using (8) and (9), we get:

$$I1 = \exp(\beta1).I2$$

and $(t1/t0)=\exp(\beta1).(t2/t0)$

where $ti/t0 = dIi/dV0$, and $dV0 = t0$, the threshold value required for perception of the output signal.

In other words, if the receptors are bleached an amount B1, the increment threshold curve for a background of intensity I1 is the same as the increment threshold curve for an unbleached receptor, displaced an amount $\beta1$ along the Fechner curve (see figure 3). This is a perfectly valid interpretation of Crawford's results. Finally, to explain the results of Barlow and Sparrock, one needs to assume only that perceived brightness corresponds to the quantity $V+\beta$: the derivation involves the approximation we met above, but otherwise it looks acceptable.

The Rushton theory and the Barlow theory thus differ in visible effect only in the place at which the bleaching signal enters the system. Barlow's theory asserts that it enters along the same path as the light signals from the rods, and Rushton's theory asserts that it enters through the feedback. Any phenomenon that distinguishes experimentally between the effects of the bleached signal ("pseudo light") and its equivalent (stabilized) real light is evidence against Barlow's theory: and it turns out that there are a number of such phenomena, though they are somewhat controversial.

Firstly, Rushton (1965b) used a perforated screen through
which to deliver a bleaching flash. The increment thresholds
predicted by the two theories in this case are different: in
Barlow's case, they should vary with the log of the average
intensity of the bleaching signal (taken over the whole field of
view); and in Rushton's case, they should vary with the average
of the log. Rushton demonstrates that the latter is in fact
observed, but Brindley (1970 p.182) points out that Rushton
failed to control eye-movements, and eye movements tend to be
very important in this kind of experiment.

In the same year, Westheimer (1965) had observed that the
threshold to a small (0.1degs) stimulus p, concentric with a
brighter region (of size d say) on a 12deg diameter background
b, varied with d: increasing d beyond 0.75deg lowered the
increment threshold to the small spot. Teller, Andrews, and
Barlow (1966) confirmed this surprising finding, and showed that
it remained true for stabilized retinal images. Westheimer (1968)
then revealed that the effect is no longer present if either of
the larger stimuli consisted of bleached receptors (after images)
instead of the equivalent real light. Westheimer concluded that
the difference postulated by Rushton, between the channels used
by real and by pseudo light, needs to be maintained.

In 1968, two other papers were published that support this
view. Ernst (1968) found that a real light background permits
better resolution of superimposed flicker than the equivalent

degree of bleaching. And in an electrophysiological

investigation, Naka and Rushton (1968) showed that the effects of

bleaching and of background on non-colour-sensitive s-potentials

in fish are very different, and that this was true in both

species that they investigated. Furthermore, it appears that the

great change of sensitivity that accompanies dark-adaptation has

already occurred by the time the s-units have been reached.

(Other evidence (see Brindley p.78) suggests that the s-units are

the horizontal cells: the conclusion of Naka and Rushton was

demonstrated only for long times; the behaviour of transients was

not investigated - indeed it is very difficult to do so, owing to

the high frequency attenuation produced by the necessarily high

impedance of the tiny micro-electrodes that have to be used in

these studies).

   In his summary of this state of affairs, Brindley (1970

p.183) comes out strongly against Rushton's idea of two channels,

asserting that some combination of Barlow's and of the early

photochemical theory should suffice to rescue the situation.

Brindley's reason is the very sound principle of biological

economy. I quote: "Rushton's suggestion explains his own

observations and those of Westheimer, but on grounds of

biological economy it is as unattractive a notion as Barlow's is

beautiful. Long-lasting insensitivity after quite small degrees

of bleaching is a wholly disadvantageous property of the visual

system. Barlow attributes it to a regrettable but unavoidable

feature of the organization of rods and cones; Rushton to an

inessential feature, fairly complex and altogether harmful. One

would have expected such a feature never to have evolved, or to

have been eliminated by natural selection if it did evolve."

This is a strong case: its key point, however, lies in the

word _inessential_. If we can establish that two channels, far from

being a wasteful and useless feature of retinal organisation, are

actually a consequence of its proper functioning, the way will be

clear for us to accept at least some of Rushton's theory. Perhaps

I may be permitted to quote once more, this time from Barlow

(1964) where he makes a suggestion for a resolution to be passed

by the International Conference of Ophthalmologists. It is that

"Apparently maladaptive features are signposts to visual

mechanisms awaiting discovery". In this case, the visual

mechanisms concerned are those related to computing lightness

(see section 4). The present theory implies that there are

indeed two channels (in disagreement with Barlow), but that the

second channel cannot be used to control attenuation (in

disagreement with Rushton), because it does not appear until

after the attenuation has occurred.

At a time of life when most scientists are content to relax

on the sidelines of their subject, Rushton has continued to

pursue his chosen field with undiminished distinction and vigour.

In 1965, Alpern (1965) discovered what he called the "contrast-

flash" effect: if the centre of a visual field is illuminated

with a brief flash λ, and a concentric annulus with a flash Φ,
then the presence of Φ influences the threshold for perception
of λ in a certain way (see figure 4b). That way will concern us
later, but for now, we need to note three points about the
contrast-flash effect:

(i) Rods and cones act independently in this effect: there

is no interaction between them.

(ii) The effect is not caused by simple distraction

(iii) Most surprisingly, the disruptive effect of the surround

flash Φ is very insensitive to its exact timing

relative to λ. Φ can occur up to 50 msecs before or after λ.

In an admirable collaboration, Alpern and Rushton (1967),
Alpern, Rushton and Torii (1970a, 1970b) used the contrast flash
phenomenon to establish a number of important results. The first
task was to examine the effects of adaptation on each and on both
of the regions in the contrast-flash situation. They found
essentially that the same relation holds as Alpern found in the
unadapted state, provided that the values of λ and Φ in the
unadapted state were replaced by sensible analogue measures for
the adapted state. More precisely, if the curve for the
unadapted state is

$\lambda/\lambda_0 = f(\Phi/\Phi_0)$ (λ_0 and Φ_0 being absolute thresholds)

the situation for the adapted state is

$\lambda/(u.\lambda_0) = f(\Phi/v.\Phi_0)$,

where u and v measure the thresholds for the current

adaptational states of λ and Φ respectively. Now, by the Fechner relation,

$$u = (\mu + \mu_0)/\mu_0,$$

and $$v = (\Phi + \Phi_0)/\Phi_0$$

In other words, the results of this experiment have the following interpretation. What was happening in the unadapted state was that signals from λ and from Φ were interacting at some stage, before they left the retina on their way to the brain. The exact form of this interaction does not greatly matter, but it is described by the function f. After adaptation, the interaction between the two may be regarded as having the same form f, provided one assumes that the signals λ and Φ are first attenuated by an amount that depends upon the local adaptation state, according to the formula $th_0/(th + th_0)$.

The next pair of papers (Alpern, Rushton and Tormii 1970a and b) asked what was the form of the signal that was attenuated. In the first of these, the experiment was conducted in the unadapted condition and consisted of varying independently the area and the luminance of the surround region Φ. To their carefully simulated astonishment, they found that whatever the form of the signal from Φ that interacted with λ to create the contrast-flash phenomenon, the size of that signal varied with the total scotopic energy incident on the surround. This was true over a large range. The exact formula for the nerve signal N is in fact

$$N = \Phi/(\Phi + \sigma) \qquad (10)$$

where σ is the size of incident luminance that gives half the

saturation value of N. In fact, σ has the value of 800

quanta/rod/sec: hence over a considerable range (e.g. 1/100 to

100 quanta/rod/sec) the signal N is effectively linear in Φ. It

is pointed out by these authors that this relationship also

describes the results of many intra-cellular measurements made in

retinal cells (Naka and Rushton 1966, Naka and Rushton 1967,

Tomita 1968, Naka 1969 (a very accurate study), Werblin and

Dowling's 1969 curves).

Combining the results of these and the previous experiments,

one derives the final relation for the size of the nerve signal

for stimuli of large area and of long duration in the adapted

state to be

$$N = (\Phi/(\Phi + \sigma)) . (\theta 0/(\theta + \theta 0)) \qquad (11)$$

(In the case where the receptors are bleached, θ is the

equivalent background light).

In the accompanying paper, Alpern et al. (1970b) use all the

artful tricks available to these experienced and ingenious

investigators to verify that this relation holds over the entire

intensity range.

The next problem is to apply the same techniques to the case

where the background is replaced by bleached receptors. Alpern,

Rushton & Torii (1970c) found that their results were well

fitted by the equation

$$1/N = (1 + \theta/\theta_0) + (\sigma/\ )(b + \Phi/\Phi_0) \quad (12)$$

where b = 10**12B, and B is the fraction of pigment that is
bleached. (12) reduces to (11) when B = 0. This equation was not
tested for backgrounds consisting of a combination of bleaching
and of real light.

A note on the disappearance of stabilised images

It might be thought that the phenomenon of the disappearance
of stabilised images would have aroused a great deal of
speculation and experiment. While it has been the subject of a
number of papers, (Barlow 1963), there has been surprisingly
little comment on the necessity for such a phenomenon. For
clearly, on Barlow's theory of bleaching signals, something has
to be done about all the images captured by varying degrees of
bleached pigment on the retina, otherwise one would have
considerable trouble trying to see real images. Evidently, the
introduction of some fairly relaxed condition on the time course
of retinal signals, sufficient to ensure that entirely stable
images disappeared, would remove the irritation attendant on
their continual presence without undue burden on the cerebral
machinery that must allow for eye-movements.

As far as I am aware, however, only one paper has been
published recently that bears upon this topic. Carpenter (1972),
remarking upon an observation by E.Darwin (1794), examined the
perception of after-images on backgrounds of different kinds. His

findings are the following. For an after image, a background luminance can always be found (after 30 secs) against which the after image disappears (reversibly). He called this luminance Ie, the eclipsing luminance, and found that Ie depends not at all on properties of the bleaching signal that caused the after-image, but depends wholly on the immediate history of the background luminance. Essentially, he found that, if stared at for long enough, any background luminance I becomes the eclipsing luminance. He suggested that the signals that form the input to Rushton's devices ought properly to be viewed as the time derivatives of certain quantities that are intimately associated with processes in the receptor and elsewhere, and he proposed a model that has a number of attractive features. However, it fails to explain the case of changing from a lower to a higher background luminance (indeed from his plotted results, it is almost as if the eye fails to change under these conditions). Carpenter's model is interesting, however, because it postulates the existence of an intermediate photo-product, whose concentration is the "bleaching signal" required by the Rushton-Hodgkin model.

Finally, mention must be made of the interesting paper by Sharpe (1972), who studied the visibility and fading of the entoptic shadows of retinal blood vessels. He showed that, for perception of fine detail, the slow drift of the shadows across the retina is essential, whereas to see coarser detail, the

contrast need only be temporally modulated. He also found that the rise in contrast threshold of the pattern of shadows as they are viewed is partially binocularly transferred, which proves that some of the perceptual fading has a central origin.

4 Land's retinex method

Intensity is the product of illumination and reflectance. The visual system is really interested in reflectance, and so we are interested in studying methods that can extract reflectance from measurements of intensity. Lightness is an approximation to reflectance that is based on the usually valid assumption that changes in reflectance are discontinuous, whereas changes in illumination are gradual. The retinex is a way of extracting lightness from information about intensity.

Land has characterised this operation sufficiently for a computer program to be written that can simulate it, and the description of that program is roughly as follows. Given a picture that looks like one of the works of Mondrian, the computer moves over the field in a pseudo-random walk, measuring the luminance of each point, as viewed through a filter of an appropriate kind (red, blue, or green). At each point, the program notes any discontinuities in the luminance, and ignores smooth changes that will commonly be due to changes in lighting

rather than changes in reflectance. The system then uses the
record of discontinuities to integrate back to the original. Thus
the output from the retinex operation is the same as the input,
except that the effects of slow changes have been eliminated.
This is done independently for all three colour channels, and the
results are then combined with some suitable normalising
constants to produce the final, computed colour. The relation
between the output from the three channels, and the "correct"
colours in the scene, can be set by the viewer if he "knows the
colour of" (can assign an internal colour name to) some objects
in the scene.

Land is careful to point out that the Retinex operation
should not be viewed only as an aid to colour vision: the
variability in illumination of monochromatic scenes is such that
the retinex process is just as indispensable to seeing them, in
conditions of everyday life. This is borne out by the great
trouble that must be taken to achieve correct lighting even for
black-and-white television, because no retinex operation is (at
present) performed before transmission of the picture.

Horn's algorithm for the Retinex effect

The implementation given by Land for his Retinex theory is
unsatisfactory firstly because it is not particularly easy to see
exactly what is happening from it; and secondly because it
relies on covering the Mondrians by a sufficiently dense random

walk, which is a somewhat messy process. Horn, in a companion
paper to this one, has shown how the retinex function may be
computed by a uniform procedure. It is easy to understand, as
follows. What Land requires from his retinex process is that it
measure local gradients, and accepts only discontinuities. In a
discrete space, this corresponds to measuring local contrast and
applying a threshold to it so that amounts that are small enough
to be due to gradual changes in illumination are ignored, and
the larger ones are let through.

The beauty of Horn's algorithm is however that it has such
an easy inverse. Write x for the output from a receptor, x' for
the output from the gradient operator, x" for x' after a
threshold has been applied, and x✻ for the final output. Then if
N(x) denotes the neighbours of x ,   the equations are as follows:

$$x' = x - 1/n \sum_{N(x)} y \qquad (1)$$

$$x" = x' \quad \text{if } |x'| > \text{threshold } t \text{ (say)} \qquad (2)$$

$$= 0 \quad \text{otherwise}$$

$$x✻ = x" + 1/n \sum_{N(x^*)} y✻ \qquad (3)$$

That these are inverses may be seen as follows. Take a point
source, with 6 neighbours arranged hexagonally round x.
Then $x=1$, $y_i=0$ $(i=1,...,6)$, and

$$x'=1, \quad y'_i = -1/6 \quad (i=1,...,6)$$

and $x✻=1$, $y✻=0$ is a solution to (3), for boundary conditions
that are zero everywhere. The solution is unique, and the two
transformations are inverses for point sources. Since any

luminance pattern may be regarded as a linear combination of point sources, the transformations are inverses in general. If a threshold is interposed, we replace x' by x" as in (2) above, and the effects of slow changes are removed. Computer simulation of the effects of this are given by Horn (1974).

Thus the retinex operation may be performed by three stages: the first takes local differences; the next stage is a threshold operation, where the size of the threshold is set appropriately; and the third, reconstituting stage is easily performed by a resistive network. The threshold may be set by looking at the gradient histogram, and choosing the threshold to exclude the central peak.

This algorithm looks much more promising for implementation in biological hardware, and it represents an important step forward. As we shall see, however, it is not quite correct; but it has probably brought us to the brink of being able to relate part of the function of the retina to its structure. In the rest of this essay, I explore the consequences of assuming that the retina performs something like the retinex computation.

PART II: The operation of the light-adapted retina
in the absence of significant bleaching

1   General Considerations

In conditions of moderate illumination, the retina has to
compute the retinex function from its inputs, and present the
answer in a form that is suitable for transmission along the
optic nerve. Luminance is the product of reflectance and
illumination, and lightness is an approximation to reflectance:
hence in order to remove the effects of illumination by a linear
process, the computation must be performed on logarithmic
measures of brightness. Because there are a number of methods
that could be used for computing the retinex function, the
analysis of retinal structure must be preceded by a brief
discussion of the constraints dictated by physiological
considerations.

1.1 The need for attenuation

The receptor response is what Rushton calls an H-curve
$(I/(I+K))$, and so is nearly linear over a large log intensity
range. Cells that transmit signals using spikes probably do not
share this large dynamic range, so in addition to the retinex
process, considerable attenuation of the signal must take place
at some stage. Psychophysical evidence relevant to deciding when
the attenuation takes place is discussed in section 3.5.

1.2 Separate channels for positive and for negative signals

Nervous elements can transmit only a one-sided response.
Hence if a number that can be either positive or negative has to
be signalled, either two channels must be used, or zero must be
coded as half the maximal response. If a condition on the
absolute size of the number has to be applied by using some kind
of threshold mechanism, one probably has to use two channels.
This argument, together with the relevant physiological evidence,
means that algorithms have to be selected that use variables
which are either always positive or always negative.

1.3 Degrees of freedom available

If a system is to perform the retinex computation using a
method analogous to Horn's algorithm, the initial operation must
be some kind of difference function, and the result of this
difference must be passed to the next stage for reconstitution of
the image. There are many variants of Horn's algorithm that
compute something equivalent to the required retinex function.
For example, the difference function may be a pure subtraction;
the log of the results of a division; or even the raw result of a
division, (i.e. the geometric mean of x/y for y in the
neighbourhood $N(x)$), provided that the next step is to take its
logarithm. For each of these, there exists an appropriate form of
the reconstitution algorithm.

1.4 Relevant physiological information

There is a large amount of information available about the retina, and because of the high prevailing standards of retinal physiology, a successful theory must explain almost every finding. Conversely, since so much is known, it is possible to be specific about most of the details of the retinal computation. There are some places where the account I have given, though probably the most likely, is not the only possible one: in such cases, I have erred on the side of specificity rather than of caution, because a definite statement that can be disproved is more fruitful than a vague one that cannot.

2 Computing the Difference Function

From 1.3, it is clear that the first stage in the retinex computation must consist of some kind of difference function. The possible candidates for the output of this stage are the receptors, the horizontal cells, the bipolar cells, or perhaps some later cell. The most likely candidates are the bipolar cells, and the following arguments constitute what is virtually a proof that it must be them.

2.1 Differences must be taken early

If a retinex operation is to be performed on a visual input,

it must be the first thing that is done to the information.
Contrast detection for local operations only may take place
before reconstitution, and may perhaps proceed in parallel with
the difference operation for the retinex, but they may not
precede it. Further, since the retinex must precede any
operations that depend on global assignments (like colour
naming), and will greatly facilitate line and edge finding (by
removing the effects of slow changes), there is no reason not to
implement it at the first available opportunity in a vision
system. Hence the first place at which the transform could be
computed will be where it is actually computed.

2.2 The difference signal must be after the receptors

The first possible place at which the difference signal
could be coded is by the receptors. A possible explanation is
that the horizontal cells, or perhaps receptor-receptor contacts,
provide the necessary antagonism to compute the gradient. The
latter possibility is ruled out, because the antagonism has to be
colour-coded, and the packing of the blue cones is inconsistent
with this. Hence, for at least this channel, it would be
necessary to create the antagonistic surround via the horizontal
cells. But the horizontal cells must be driven by the receptors:
hence, if they fed back onto and affected directly the signals
from the receptors, the receptor output signal would be the
steady state solution to an expression of the form

$$x' = x - 1/n \sum_{N(x')} y'$$

which is the wrong transform. Hence the difference function

cannot be signalled by the receptors: they must carry the pure

receptor response.

2.3 The horizontal cells must provide the antagonism

Since the receptors must signal the flux of light that they

receive, some other agency must provide the antagonism necessary

for computation of the difference function. This agency could not

be the midget bipolar cell because it contacts only one cone.

Hence the midget bipolar cell must carry the difference function,

and the horizontal cell must provide the necessary antagonistic

surround for computing the local gradient. Ideally, there should

be as many horizontal cells as there are receptors: there are

not, however, and because the horizontal response is a linear,

not a logarithmic function of intensity, the compromise

represents some degree of approximation. However as we shall see,

the receptor-bipolar interface is probably a very complex

structure, and a final judgement of its performance must wait

until it is thoroughly understood.

2.4 Evidence from the mud-puppy

The above arguments were constructed from purely anatomical

evidence, because there is no physiological information about

these cells in the primate retina. Werblin & Dowling's (1969)

recordings from the mud-puppy retina are however perfectly in
accord with this view. The receptors show no centre-surround
organization: neither do the horizontal cells; yet the bipolar
cells do. Because the structure of the mud-puppy retina is so
similar to that of the primate, (Dowling & Werblin 1969), this is
strong supportive evidence that the horizontal cells are
responsible for antagonism which interacts with the receptor
signal at its transmission to the bipolar cell.

2.5 Absence of interaction between horizontal cells

     The present theory requires that there should be no direct
horizontal cell - horizontal cell interaction. The difference
function for a single cone channel, though it need not consist of
taking only the difference between a central receptor and 1/6 the
sum of its six neighbours, should have a spread that is only
slightly less local. There should be no trans-retinal
transmission that would follow from significant interaction
between horizontal cells. In particular, the peculiar long-
distance effects known to occur for retinal ganglion cells (see
part I 2.5.2 (XY7)) should not be present at the bipolar cell
dendrites. (They will be discussed in section 3 of Part III).
Dowling & Boycott searched carefully for synapses between
horizontal cells, but failed to find any.

2.6 Remarks about horizontal cell physiology

The consequences that follow from the requirement for no
interaction between horizontal cells in each retinex system, and
for no interaction between the different systems, are these:
firstly, there should be no interaction between the horizontal
processes that form the lateral elements of the receptor
invaginations; secondly, the horizontal cell axon must be without
function; and thirdly, the physiology of the horizontal cell
"dendrites" should be very similar to that of its axonal
processes. B.B.Boycott (personal communication) has suggested
that the horizontal cell axon may be vestigial.

3 Transmitting the Difference Signal

Having decided that the bipolar cells must carry the initial
difference signal, we now ask what is its form, and how is it
carried.

3.1 Positive and negative channels

Whatever the exact shape of the bipolar cell message, it
should in principle be continuously variable over a large range,
both positive and negative for a subtraction or for the log of a
division, or less than and greater than 1 for a raw division. In
the mud-puppy, this is not the case (see Part I section 2.3): the

effect of the horizontal cells is to attenuate transmission from the receptors, not to polarise the bipolar cell in the opposite direction. Further, two kinds of bipolar cells are found, with so-called "on-" and "off-centre" responses. It therefore looks as if, in the case of the mud-puppy, the difference function is signalled along two channels: one carries x'' if x'' is positive, (or >1) and the other carries $-x''$ when x'' is negative (or <1).

This is to be expected, for the reasons outlined in section 1.2. The difference carrier has to transmit a two-sided signal, and this signal has to be thresholded, whether its magnitude varies (roughly) linearly with the signal or logarithmically. Hence two channels are used. One other small point is the additional bonus of improved accuracy: this may be useful, even if the signal ranges were of the order of 100000, because the bipolar channel is carrying differences rather than absolute values.

3.2 Extension to the primate retina

In view of these arguments, the organization of the bipolar cells found in the mud-puppy makes such very good sense that one is entitled to expect a similar organization in the primate retina. If the retinex process is to be carried out at the ultimate resolution of single receptors, it will be necessary to supply each receptor with two kinds of small bipolar cell, one for the positive, and one for minus the negative component. It

cannot be coincidental that each cone contacts one midget, and
one flat midget bipolar cell (Kolb 1970). The present theory thus
explains the presence of the two kinds of midget bipolar cell,
though it cannot predict which channel is which.

In the case where the retinex process is carried out at
lower resolution, as it may be in many non-primate retinas, in
the rod channel of the primate retina, and perhaps also in
parallel from cones in the primate retina, it is not necessary to
match the two channels so exactly, though the closer they are the
better. Thus, where bipolar cells contact more than one receptor,
the condition for strict duality of bipolar cells is relaxed
slightly, though the resolution of the two channels must be the
same. It will however always be necessary to have two channels.
Hence every receptor must contact at least two bipolar cells, one
plus and one minus. This consequence is particularly interesting,
since each rod in the primate retina usually contains only two
bipolar processes: the theory therefore predicts that these must
have opposite signs.

3.3 The difference function used in the mud-puppy

Werblin & Dowling (1969) reported that the bipolar cell
response in the mudpuppy retina depends, over a large range of
absolute energies, upon the ratio of the energies in the centre
and the surround. The receptor response curve is an H-function
$((I/(I+K))$, see 2.1 of Part I), and Naka & Rushton (1967) found

that horizontal cells also exhibit an H-curve response to
illumination. Rushton (1972 p24P) believes that the H-relation is
continued in the bipolar cells, but does not cite any direct
evidence.

The question is an important one, because one is naturally
interested in the exact function that is transmitted by the
bipolar cell. Three kinds of evidence suggest that it should not
be a pure subtraction: firstly, the requirement of Land that the
retinex operation be performed on logarithmic quantities;
secondly, the finding of Werblin & Dowling mentioned above; and
thirdly indirect psycho-physical evidence that the reconstitution
stage is computed using logarithmic quantities (see 3.5 below,
and Part III). There is however some room for doubt about whether
the bipolar signal is a raw division or its logarithm, though it
is probably not exactly either of these. On the available
evidence, the most likely candidate seems to be the logarithmic
quantity, but it may be carried with variable gain.

3.4 Evidence about the order of subsequent operations

It is clear that a large number of transformations must be
applied to the difference signal before it can be passed to the
optic nerve, and psychophysical evidence can help us to decide in
what order they are performed. From the contrast-flash
experiments of Alpern, Rushton & Torii that were reviewed
earlier, it is clear that several things are going on. They are:

3.4.1 The summation of signals from the various parts of the
surround Φ: this summation is linear in the incident energy.

3.4.2 The logarithm of the sum (or some very similar function) is
taken, both of the surround, Φ, and of the centre, λ.

3.4.3 Both signals are attenuated by an amount that depends on
the background (ignoring bleaching), by a fraction
$\theta\emptyset/(\theta + \theta\emptyset)$, where θ is the strength of the background. The
experiments do not specify whether the attenuation is performed
after 3.4.1, in which case it is a division; or after 3.4.2, in
which case it amounts to a subtraction. Rushton's (1965FL) use of
Fuortes & Hodgkin's (1964) cable theory model is a clever example
of how the two operations could be performed simultaneously.

3.4.4 Finally, the two transformed, attenuated signals are
compared, and if the difference is noticeable, a signal is
transmitted. The results of Alpern & Rushton (1967 figure 1) are
best explained by supposing that 3.4.4 is performed after all of
3.4.1 to 3.4.3.

3.5 The order suggested by this theory

        In trying to decide upon the nature of the bipolar signal,
one must first be clear about which facts are not relevant. One
result suggesting that the signal must at this stage still be a
linear function of the intensity is Alpern et al.'s (1970a)
finding that the surround signal adds linearly before interacting
with the signal from the centre. However the size of the central

region (2 degrees for the rod experiments, and 20' for the cone
ones, Alpern et al. 1970a&d) allow that the summation could all
take place in the horizontal cells, which are known to exhibit a
linear response.

A second possibly relevant finding is Rushton & Westheimer's
(1962) that the threshold of all the rods in a summation pool is
apparently raised by bleaching some of them. This seems to imply
that at some stage, a Rushton-style gain box must be used that
attenuates all the signals from a summating pool. This result is
apparently very difficult to explain on the present theory,
because if the bipolar signal really is a division, or log
division, there is no necessity to set an explicit threshold: a
constant one will suffice (see 4.1). Yet the natural place for
the gain-box to be is at the foot of the bipolar cell axon,
because this is where the pseudo light input (corresponding to
the required bleaching signal) is to be found (see Part III).
There is however a possible explanation that does not involve the
gain-box idea at all – indeed almost the only concept that is
required from Rushton's theory is the notion of two paths, one
for pseudo and one for real light. If this is correct, the
pseudo light signal will explain the results of Barlow &
Sparrock, (see Part III), and the results of Rushton & Westheimer
(1962), of Rushton (1965a and b), and of Westheimer (1968)
receive a separate explanation. These have to be dealt with in
Part IV, but they are consistent with the notion that the bipolar

signal is the log of a division.

The main argument from retinal structure suggests that whatever quantity is carried in the bipolar channel, it is composed of subquantities that can freely be added together. This is because the horizontal cell interaction with the bipolars is dispersed, which means that if the receptor-horizontal-bipolar junction performed a pure division, errors would arise because of different values of the divisor provided by the different horizontal cells at each site. The difference between $((1/2)(1/d1 + 1/d2))$, and $r/\sqrt{(d1 d2)}$ is $(\sqrt{d1} - \sqrt{d2})/\sqrt{d1 d2}$: this is especially severe when the $\sqrt{}$ are very different, which in this case occurs at boundaries – i.e. at the very places that are of interest. The observed anatomical arrangement would therefore be bad for accurate division. This argument is weakened a little by the presence of two, usually different horizontal cell processes at each site – suggesting that the divisor is their average – but it is still forceful: the argument implies that the bipolar signal is the subtraction of two logarithms, rather than the division of two numbers, though this quantity may be carried with variable gain.

3.6 Summary

It is therefore assumed that the bipolar cell carries a signal that corresponds to an H-function of some possibly variable multiple of the quantity $(log(x) - (1/\sum_{N(x^*)} w(y))(\sum_{N(x^*)}$

$w(y)\log(y)))$, for some weighting function w. The careful data
presented by Kolb (1970 figures 37 & 44) show that the diameter
of the weighting function w is about 70μ, and w decreases
with distance from the central receptor x.

4 Attenuation and thresholding of the difference signal

The heart of the retinex computation is the application of a
threshold to the difference signal, for it is this that removes
the effects of slow changes in luminance from the image.

4.1 The size of the threshold

Consider the effects of doubling the luminance of a scene
(measured on a linear, not a logarithmic scale). The energy
received from each point doubles, hence gradients double, and
hence the necessary threshold must also double. This is however
only a rough guide, because bright sunlight produces views that
are much more contrasty than the (roughly) uniform illumination
provided by a thick layer of cloud. If a histogram of gradients
from a scene is taken, the central peak may often correspond to
the thresholds that have to be removed, and it will be important
(when adequate transducers become available) to explore how the
width of this central peak depends upon various lighting
conditions. Until then, one must assume that a linear threshold

is adequate; this is consistent with psychophysical evidence on contrast detection (see section 1 of Part III). If the bipolar signal is logarithmic, a fixed threshold applied to the bipolar signal will of course behave like a threshold that varies linearly with intensity, as required.

4.2 Difference signal normalisation

Even given that the raw division signal is transformed into its logarithm somewhere in the outer plexiform layer, there remains a question about the normalisation of the difference signal. Despite the fact that the dynamic range of the difference channel is probably much greater than that of the ganglion cell axon, it should be possible to use the whole of the ganglion cell channel capacity throughout much of the working range of the retina. The best way is to arrange that the necessary normalisation takes place at the last possible moment - in this case, on the difference signal just before reconstitution. It can take place independently in the three colour channels, because the relationship between colour names and the relative strengths of the signals from the retina has in any case to be determined by some other mechanism.

5   Reconstituting the Image

The reconstitution of the image consists of computing the

steady state solution to a set of equations like:

$$x\text{✩} = x'' + 1/n \sum_{N(x\text{✩})} y\text{✩} \quad (0)$$

where the neighbourhood $N(x)$ is matched to the neighbourhood

that was used in the primary gradient extracting operation. In

the retina, the position is much complicated by the fact that the

difference signal x'' is carried over two channels, which will

be called $x''+$ and $x''-$, where $x'' = (x''+ - x''-)$. In this

situation, the inverse transformation becomes

$$x\text{✩} = (x''+ - x''-) + 1/n \sum_{N(x\text{✩})} y\text{✩} \quad (1)$$

One is naturally interested in formulations of this

operation in which the final output is split into two channels,

because of the existence of on- and of off-centre ganglion cells.

(It may be that the output has to be split for reasons like those

of section 1.2, applied to later contrast detection operations).

5.1 Possible schemes for computing the inverse

There are a number of possible ways of implementing equation

(1), and they are listed below.

5.1.1 The first is a direct implementation, using a simple

resistive network. Although this is the neatest solution, it

cannot be implemented in the retina because one has to operate

with variables that are always positive.

5.1.2 The second possibility is to solve the pair:

$$x\text{✩}+ = x''+ + 1/2n \sum_{N} y\text{✩} \quad (2)$$

$$x_{*-} = x''_- - 1/2n \sum_N y_* \qquad (3)$$

together with the condition that $x_* = (x_{*+} - x_{*-})$. This method cannot be used either, because circumstances can be constructed that require both channels to be able to carry both positive and negative signals. For example, consider a situation where the local average luminance is high, and x'' is small but negative. From equation (2), x_{*+} stays fixed, and from (3), x_{*-} becomes slightly less negative. Hence both channels need to carry signals of both signs, and this implementation is therefore not suitable.

5.1.3 To overcome this, one might propose the solution obtained by solving (1) and simultaneously converting the answer to two channels x_{*+} and x_{*-}, where $x_{*+} = x_*$ if x_* is positive, and $x_{*-} = -x_*$ if x_* is negative: both are zero otherwise. I have examined this possibility in detail, but was able to reject it. The argument is long, and will not be given in full here; but the basic reason can be briefly described. In order for this scheme to work in a way that is consistent with retinal anatomy, it turns out that the full solution x_* has to be computed by both the plus and the minus channels. It is possible to show that the full solution (with different signs) must therefore appear at the axon terminals of both kinds of midget bipolar cell: this requires that both terminals have access to both x''_+ and x''_-. But x''_+ has been transformed to x_{*+} before it can leave the plus bipolar channel, and similarly for x''_-. Hence, the negative channel cannot receive x''_+, and vice versa, so the method fails.

5.1.4   The fourth possible scheme is to try to keep the two halves of the solution apart as long as possible. This means solving the pair:

$$x_*+ = x''+ + 1/n \sum_N y_*+ \qquad (4)$$

$$x_*- = x''- + 1/n \sum_N y_*- \qquad (5)$$

At first sight, this pair of equations does not appear to compute anything useful; but one can prove that it does as follows.   By subtracting (4) and (5), we obtain:

$$(x_*+ - x_*-) = (x''+ - x''-) + 1/n \sum_N (y_*+ - y_*-) \quad (6)$$

Thus the expression $(x_*+ - x_*-)$ is in fact a solution of the equation (1). What this means is that a solution may be obtained from (4) and (5) provided that the two solutions are coupled in a subtractive way. In general, the two halves x_*+ and x_*- will both be large and positive, since there is nothing negative in either of (4) or (5) to pull them down. The solution is however not disturbed by subtracting a suitable function $f(x_*)$, provided that it is done to both x_*+ and x_*- simultaneously. In the particular case of the retina, it is important to keep the terms positive, so the amount to be subtracted must never exceed the smaller of x_*+ and x_*-: subject to this, however, a subtractive coupling between x_*+ and x_*- is permissible.

Hence we obtain the result that (4) and (5), together with the operations:

$$x_*+ \quad \text{goes to} \quad (x_*+ - f(x_*)) \qquad (6)$$

$$x_*- \quad \text{goes to} \quad (x_*- - f(x_*)) \qquad (7)$$

still represents a solution to (1), as long as the condition

$x^{*}+$ and $x^{*}-$ are kept positive   (8)

also holds. In particular, the gain in the circuit that defines

f may be chosen to keep the smaller of $x^{*}+$ and $x^{*}-$ near zero.

It is perhaps worth pointing out that the determination of f is

quite separate from the problem of fixing the DC level for the

output: nor can variations in f account for the disappearance

of stabilized retinal images, since this would correspond to

tampering with the difference $(x^{*}+ - x^{*}-)$: as long as this

difference is preserved, the output of the process is a faithful

retinexed copy of the image. (The stabilization problem is dealt

with in section 4 of Part III).

5.2 Implementation details

The method of 5.1.4 is the most satisfactory that I have

been able to devise, and it will now be shown that it provides an

explanation of many features of the inner plexiform layer.

5.2.1 The realisation of equation (4) requires a device with

feedback, because the answer, once obtained, must be applied to

the mechanisms that compute the answer at neighbouring points.

Since the retinal ganglion cells are not pre-synaptic to any

other retinal cells, the expression of the final answer $x^{*}+$

(say) cannot exist only at these cells, since if it were, it

would not be available to neighbouring ganglion cells.

5.2.2 Hence, the answer $x^{*}+$ must be computed in the bipolar cell

axon terminals, or in the amacrine cells, or both. It cannot be just the amacrine cells, because most of the ganglion cell synapses come from bipolar cells. The most likely candidate is therefore the bipolar axon terminals, with some kind of modification from the amacrine cells.

5.2.3 The complex of a dyad synapse plus a return synapse from the amacrine cell to the bipolar cell is exactly what this hypothesis requires. The reason is that the amacrine cells have to carry an expression like $\sum y_{\alpha+}$ for a neighbourhood, and add it to the quantity x^{α}_{+} that comes down the bipolar axon. At the dyad synapse, the component $x_{\alpha+}$ is transferred to the amacrine cell transverse carriers, and the expression ($x_{\alpha+}$ + $y_{\alpha+}$) is added in to complete the computation of $x_{\alpha+}$. One would therefore expect all dyad synapses that involve a stratified amacrine cell to be accompanied by an amacrine/bipolar synapse.

5.2.4 The philosophy expressed in section 1.1, that spike-carried signals must be logarithms (because of the smaller dynamic range) is consistent with this. At least some amacrine cells support spikes, and for the theory to be correct, the lateral interactions must be in terms of logarithmic quantities. (This also explains various features of bleaching adaptation: see Part IV). Conversely, if the reconstituting stage is computed in terms of logarithms, the gradient finding operation must be a division.

5.2.5 The coupling between the solutions for $x_{\alpha+}$ and $x_{\alpha-}$ must

be provided by diffuse amacrine cells, probably acting through

the amacrine-amacrine and amacrine-ganglion cell synapses. This

coupling can be done effectively in a number of ways, and one

cannot decide a priori which should be used. It is however

helpful to set out an explicit method for achieving it, because

variants on this method are then rather easy to devise. Write POS

to stand for "the positive part of": e.g. POS(4)=4, and POS(-

4)=0. (POS is useful because it captures the idea that a neuron

with no threshold can transmit at best only the positive part of

the function that it receives.) Then let us modify equations (4)

and (5) to the following:

$$x_{*}+ = x"+ \; + \; POS(1/n \sum_{N} (y_{*}+ - y_{*}-)) \qquad (4')$$

$$x_{*}- = x"- \; + \; POS(1/n \sum_{N} (y_{*}- - y_{*}+)) \qquad (5')$$

Then $(x_{*}+ - x_{*}-)$ is still a solution. Now write

$$G+ = POS(x_{*}+ - x_{*}-) \qquad (9)$$

$$G- = POS(x_{*}- - x_{*}+) \qquad (10)$$

Then one of G+, G- will be positive, and will represent the

solution.

The idea behind this formulation is illustrated in figure 5.

The subtractive interaction is done in two ways: firstly, there

is reciprocal inhibition between the stratified amacrine cells in

the top and the bottom thirds of the inner plexiform layer,

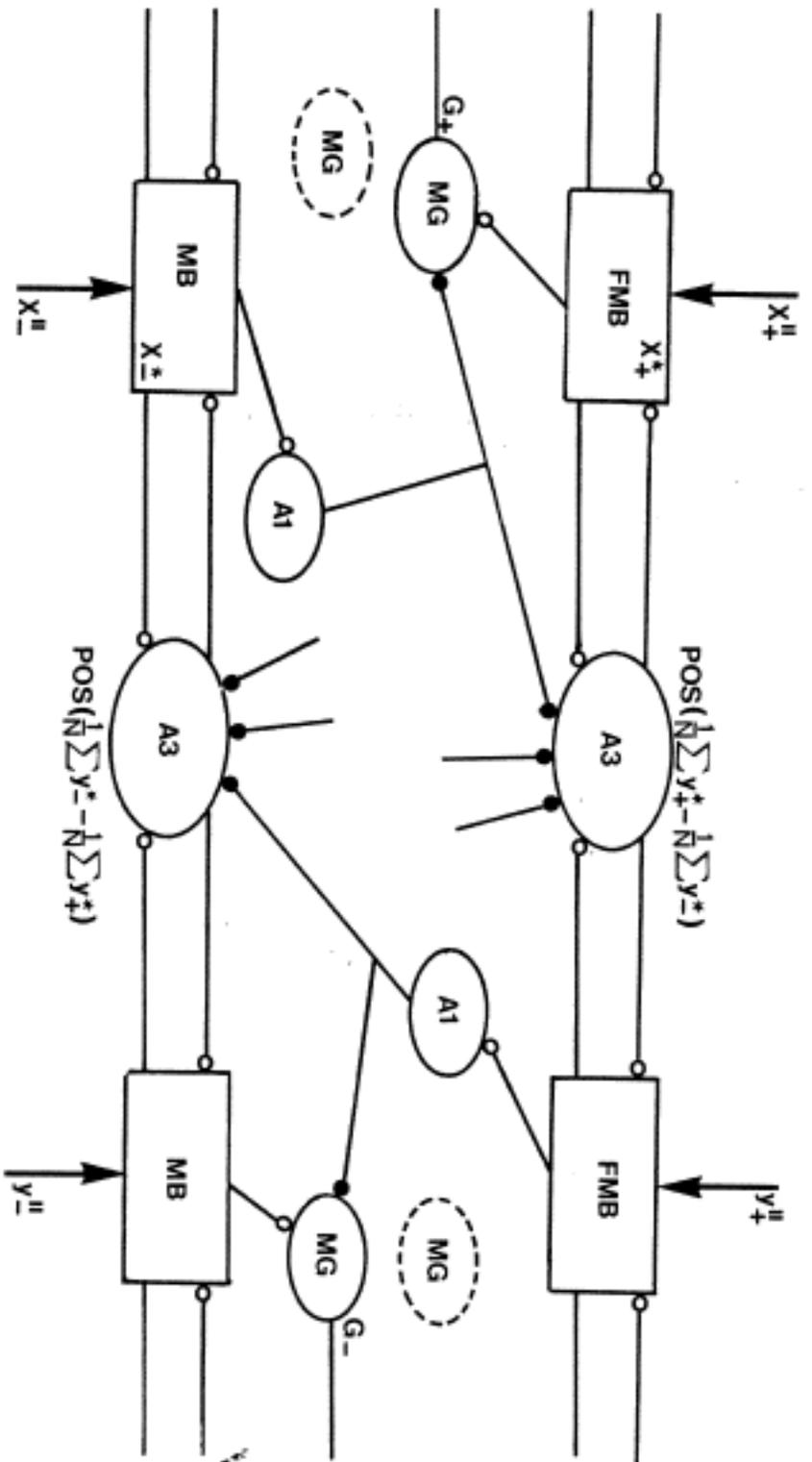mediated by the small diffuse amacrine cells. Secondly, there is

FIGURE 5

Legend to Figure 5

One way of computing the two halves of the solution in the inner plexiform layer. Open circles represent connexions that are computationally positive: filled circles, those that are negative. Cell names are as in figure 1. The association of FMB with x"+ rather than with x"- is arbitrary.

direct inhibition that acts on the ganglion cells, again mediated
by the diffuse amacrine cells. The important points about this
are:

(a) the form of the coupling between the two systems of
stratified amacrine cells is very flexible; both the kind of
coupling (i.e. the number of terms collected underneath the POS
function), and the amount, are variable provided that they
correspond closely in the plus and minus layers.

(b) the sizes of $x*+$ and of $x*-$ are kept positive: this is vital
for allowing a $x"+$ signal to escape from the bipolar terminal
and influence the solution even where $G-$ is strongly positive. It
is because $x*+$ and $x*-$ must be non-negative at all times that
direct inhibition to the bipolar terminals is undesirable.

5.3 The distribution of amacrine cell processes

From the analysis of 5.2, we see that the stratified
amacrine cells must be the ones primarily responsible for (4) and
(5), (or (4') and (5')), whereas the coupling between the two
halves must use diffuse amacrines.

5.3.1 In particular, the size of the stratified amacrine cell
processes must reflect the distribution of horizontal processes
above them. The diameter of the stratified diffuse amacrine cells
(category A3 of Part I, 1.2.1) is 20-50μ, which is consistent
with the figure of about 35μ for the radius of the horizontal
cell interaction, (see section 3.6).

5.3.2 The amacrine cells responsible for coupling the plus and the minus channels are probably the narrow field diffuse amacrine cells, (category A1), because the only other candidates, (A5), have a larger diameter, and are uncommon. •

5.3.3 From these, the following synaptic relations follow:

(a) the dyad synaptic complex should connect midget bipolar terminals predominantly with stratified amacrine cells (A1). All the synapses at these complexes should have a positive sign, from the point of view of the computation. In practice, this means that the bipolar to amacrine, and the bipolar to ganglion cell synapses should be excitatory; and the amacrine to bipolar synapse should have the same effect on the midget bipolar terminal as stimulation of the centre of that cell's receptive field.

(b) The diffuse amacrine cells should receive excitatory synapses from the midget bipolar cells, or perhaps stratified amacrine cells, in one layer, and should send inhibitory synapses to the stratified amacrine and midget ganglion cells in the other. They should not send inhibitory synapses to the midget bipolar terminals.

5.4 Ganglion cell dendrites

By the same arguments, the retinex output should come from ganglion cells with stratified dendritic distribution. It will be shown below that these should be the X ganglion cells. Y
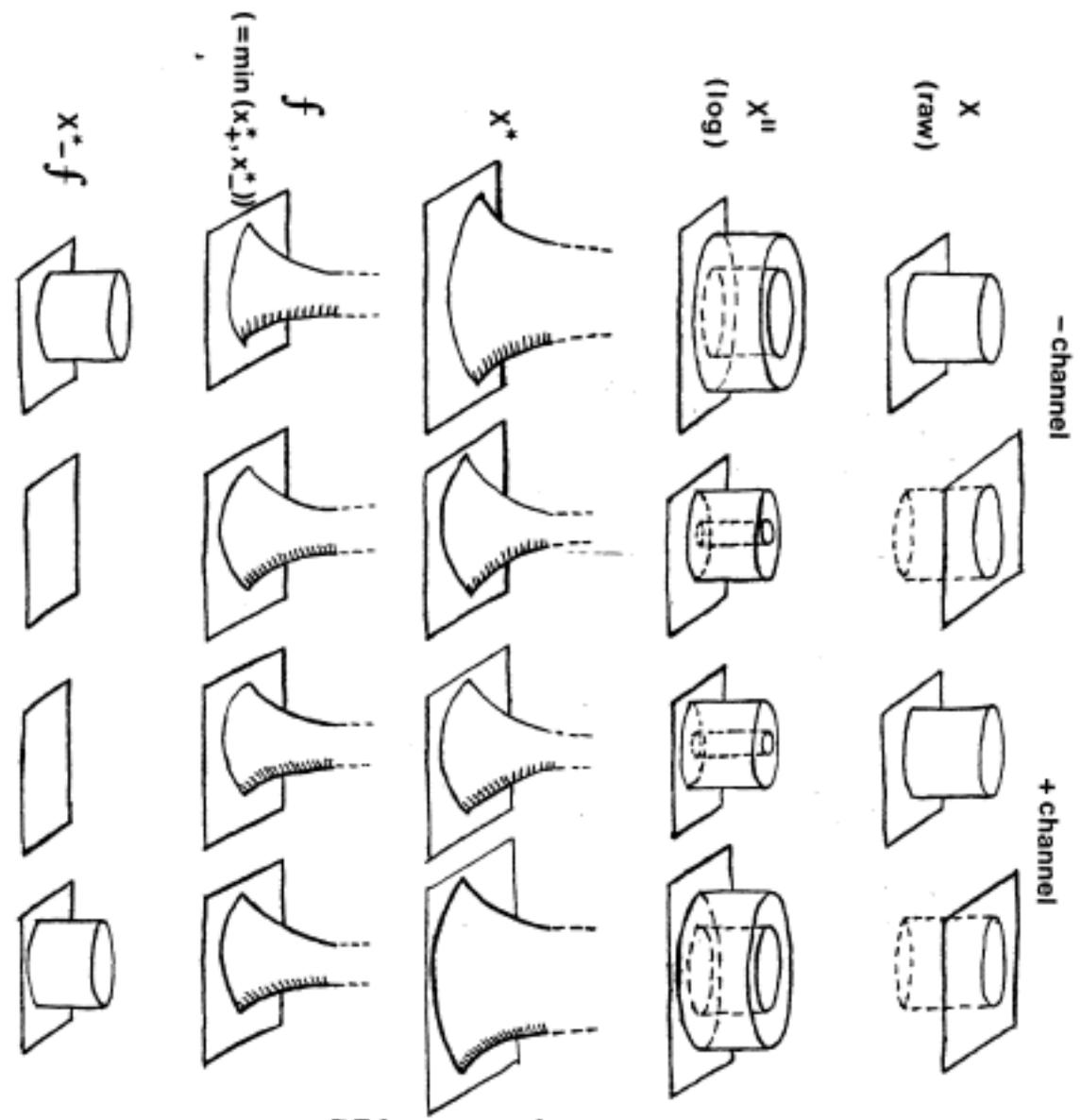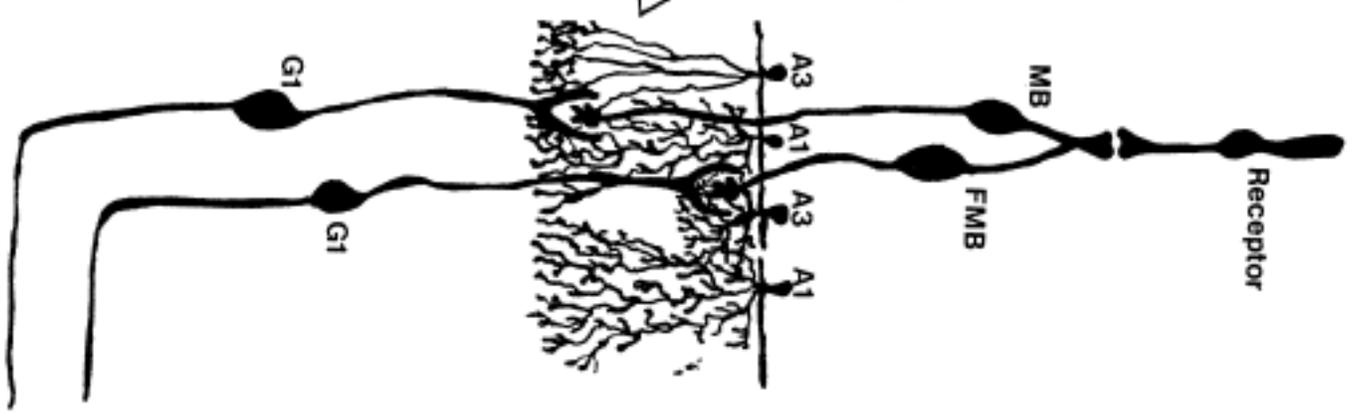
FIGURE 6

Legend to figure 6

The principal elements of the plus and minus retinal channels are shown, as analysed in the text, together with their predicted responses to a central black and to a central white spot. The solutions for x☆+, x☆-, and f(x☆) are strictly speaking undefined for the case where there is no subtractive coupling. In the retina, such coupling is provided by diffuse amacrine cells, and in its presence, both x☆+ and x☆- are properly defined. Notice that the response to a central white spot appears at what was initially the negative channel.

cells, which have a response that looks more like that which the
coupling amacrines should have,may be the diffuse retinal
ganglion cells.

The summarizing diagram shown in figure 6 illustrates the
expected responses of the various retinal components to two
simple stimuli.

PART III: Properties of the Signals from the Retina

The above analysis has dealt with the processing of a single
channel of the kind originating from a red or green cone in the
fovea; but it is clear that similar arguments may be applied to
the other channels. The same kinds of computation have to be
performed, but there is no difference in principle between low
and high resolution channels because of the linearity of the
basic equation (1). I shall therefore omit further analysis of
retinal structure, and will turn now to examine some of the many
peculiar properties of retinal output: the above interpretation
of the retina makes a number of them easy to understand.

1 The Weber-Fechner Law

The Weber-Fechner relation between overall brightness and
increment threshold is so old that it might be thought not to
need an explanation. Nevertheless, it is a curious law, depending

as it does linearly rather than logarithmically on the luminance
of the background. It is particularly curious that, although at
low levels of illumination the increment threshold varies with
the square root of the background luminance, this ceases to be
true at higher levels, although the receptors are still capable
of responding linearly with the incident energy. Why does the
square root relation, which is the optimal threshold for signal-
to-noise purposes, break down where it does - at around 3 or 4
quanta per rod per second?

The present theory provides an explanation. The purpose of
the retinex operation is to remove from the image, intensity
changes that are due to gradual changes in luminance. In the
theoretical case of a continuum, it is easy to distinguish
continuous from discontinuous changes; but if the continuum is
replaced by a discrete space, the question becomes what size of
very local gradient is large enough to be called a discontinuity?
It was seen in section 4.1 that in many conditions, the relation
between ambient luminance and the required threshold is a linear
one. But the increment threshold experiment measures exactly such
a threshold; hence the Weber-Fechner Law is a consequence of
computing the retinex function.

2 The dark-adapted retina

At low levels of luminance, the retinex function cannot be
computed. This happens when the average quantum catch at a

receptor is so low that difference signals measured by the
bipolar cells over their inherent time constant would be due to
the sparseness of the incident quanta rather than to a real local
gradient

At these low luminance levels, therefore, the retinex
computation drops out, and the retina becomes an efficient
detector of quanta, as Barlow has suggested. At the same time,
one would expect the centre-surround organization to disappear,
and illusions that rely upon the retinex process, like the Craik-
Cornsweet illusion, should no longer function. I have no specific
suggestions about how these changes are implemented: it may be
that they are related to the peculiar properties of the electro-
retinogram at low levels of illumination, (see e.g. Cone & Ebrey
1965).

3 Ganglion cell response characteristics

There are many aspects of retinal ganglion cell responses
that have not yet been discussed. Firstly, it will be evident by
now that according to the present theory, retinal ganglion cells
should behave like idealised receptor cells, rather than in the
complex manner that has been observed (see Part I, 2.5). Why are
there X, Y, and W cells, and why do they display such peculiar
characteristics? Now the computation of the retinex process is
quite a complex operation, and accuracy probably requires that
the time constants of parts of the system are quite long. It

would therefore be surprising if the retinexed output showed very good temporal resolution. This output should however be tonic rather than transient; its sensitivity (but not its response curve) should be a linear function of retinal illumination (because of the linear properties of the receptors); and the output will be relatively insensitive to flicker, or to very fast moving stimuli. Furthermore, if the computation is done accurately, there is no reason to expect the retinexed output to show McIlwain's periphery effect. Thus the conditions to be expected for the output from the retinex computation are those satisfied by X cells, but not by Y cells. In particular, if this is correct, the "colour" area of Zeki (1973) should receive its colour information ultimately from X and not from Y cells.

Why have Y-cells, and how are they formed? There is an obvious answer to this, namely that fast moving things are very important in the world. An animal with a perfect retinex system will nevertheless fail to survive if he cannot see a predator flash across his visual field, and long time constants in the computation would have this unfortunate effect. It therefore seems appropriate that another channel should be present, called the Y cells, which are especially concerned with detecting transients in the retina that are averaged out in the main retinex output. As we have seen, most inputs will cause a change in the resting levels in the amacrine layer: in the tentative theory set out above, for example, the necessity to keep

computing f efficiently would mean that a change almost anywhere
in the retina could have an effect on f elsewhere - an effect
that is subtracted out before the main retinex output channels
are reached. If this effect was applied to a second channel but
with the signs such as to add rather than subtract the extra
quantity, (perhaps by using a diffuse retinal ganglion cell),
something like the behaviour of the Y cells, with McIlwain's
periphery effect, would be the result. The question is a long and
dificult one, that deserves a separate paper: but at least the
present theory explains why Y-cells are necessary, even if not
exactly how they are formed. Analogous arguments apply to W-
cells.

4 The disappearance of stabilised retinal images

     The disappearance of stabilised retinal images is a puzzling
phenomenon, and one that would not immediately be expected from
the present theory. The retinex output, presumably the X-cells,
should not decay with time, so there seems to be no good retinal
reason why stabilised images should disappear. It is particularly
interesting that Sharpe (1972) found evidence that at least a
part of the disappearance was binocularly transferred, and hence
of necessity a central phenomenon. Now if the present theory is
correct, the X-channel will necessarily carry pseudo-images
caused originally by receptor bleaching (see Part IV); and
because of their relative brightness, if they were visible in the

normal way they would undoubtably interfere directly with perception.   Sharpe has also pointed out that the slow drift of the retina that occurs during normal fixation is essential to the perception of fine detail, yet according to the present theory, the X-channels, presumably the means of transmission of the finest detail, should be largely unaffected by this drift.

An attractive possibility is that the main contrast analysis of the striate cortex is driven by X-cells, but that the output from this area cannot influence subsequent stages unless it is accompanied by Y-cell signals from the same retinal area. This would be a convenient way of removing unwanted stabilised images, and goes some way towards understanding the parallel inputs to areas 18 and 19 that were observed by Hoffman & Stone (1971).

5 Ganglion cell receptive field organization

The conventional interpretation of the "centre-surround" organization of the ganglion cell receptive field is that it is a method for enhancing contrast across boundaries. If the retinex computation was working properly, the output of the foveal ganglion cells should be like the output from the receptors except that it will be very much more useful than the true output from the receptors, since the effects of small gradients have been removed. However, there are two reasons why, in the conditions under which such experiments are usually performed, these cells should exhibit behaviour consistent with the centre-

surround hypothesis. Firstly, the retinex computational machinery
is designed to be run in conditions of normal life: it depends,
especially at the reconstituting stage, upon lateral interaction
with reasonable signals from the rest of the retina. If a central
bright spot is presented to a ganglion cell, and then moved
around close to it while the rest of the retina is in darkness,
it is not clear what will happen. Even apart from these
considerations, however, one would expect the normal working
retinal machinery to exhibit the centre-surround effect. This is
because the DC level in the reconstitution phase is not fixed,
and is presumably chosen to be about the average of the received
luminance at the retina. It follows that the behaviour of a
midget ganglion cell will apparently be strongly influenced by
the local organization of the light on the retina, especially if
(as in well-conducted experiments) not much is happening
elsewhere on the retina. If (as also seems sensible) the DC level
is set on a local basis (giving maximum local intensity
resolution), the same will hold even if there is much happening
elsewhere. Hence it is no surprise that the ganglion cells
exhibit a centre-surround organization; this organization is a
consequence of the implementation of the retinex process, and is
not the start of contrast detection.

Because of this, it is necessary to search the literature on
the lateral geniculate nucleus to see what role the centre-
surround organization at the retina could be playing in achieving

the centre-surround organization at later stages in the visual

pathway. This question has been asked in a recent paper by Maffei

& Fiorentini (1972). These authors state that if the retinal

organization were the first step in contrast detection, the

surround at the retina should map into the surround at the

geniculate. If on the other hand, the surrounds at the geniculate

are constructed from the centres of other retinal ganlion cells,

then the contrast analysis apparently being performed at the

retina is effectively being thrown away at the geniculate.

Indirect evidence that this is in fact the case is to be found in

papers by Hubel & Wiesel (1961), and by Singer & Creutzfeldt

(1970); in the change in retinal organization during dark

adaptation (which certainly suggests that the retinal

organization is not essential for contrast detection); and in

the peculiar characteristics of the size of the surround area of

a retinal receptive field. Maffei & Fiorentini found

experimentally that the geniculate surrounds are in fact composed

of the centres of a number of retinal receptive fields. Their

argument is however weaker than they apparently believe, because

the geniculate organization seems to enhance, not contradict the

retinal organization. Thus the strongest result that they can

claim is that the retinal organization may be inessential for the

later computations.

PART IV: Bleaching adaptation

It is now time to turn to the final important cluster of
retinal phenomena, namely the vexed question of the effects on
the retina produced by bleaching its receptors.

1 Summary of the important facts

It was seen from the brief review of the subject that was
presented in section 4 of Part I that the phenomena that need to
be explained are the following:

1.1 The grating experiments of Rushton & Westheimer (1962) and of
Rushton (1965a), and the punctate background experiments of
Rushton (1965b).

1.2 The log of the increased increment threshold due to bleaching
depends on the fraction of pigment bleached $(\log(\theta/\theta_0) = kB)$;
two things are remarkable about this: firstly that the degree of
bleaching definitely determines the increased threshold; and
secondly, that a small amount of bleaching raises the threshold
by a large amount.

1.3 Barlow's hypothesis accounts for 1.2; for the similar time
course of dark-adaptational changes in spatial summation and
small field threshold; for the similar effects of bleached
receptors and of a real background; for the lasting constriction
of the pupil produced by bleaching; for the results of Rushton &
Westheimer (1962); and for the results of Barlow & Sparrock

(1964) about the subjective equivalence of backgrounds and
bleachings that have equal threshold-raising capabilities.

1.4 Rushton's hypothesis, of separate paths for the bleaching
signal and for real light signals, also accounts for all the
results that Barlow's theory accounts for. In addition, it
accounts for a number of phenomena that Barlow's theory cannot
handle.   These are Rushton's (1965b) finding that the effective
intensity of bleached receptor signals behaves like the average
of their logarithms, not the log of the average effective
intensity; it is not contradicted (like Barlow's theory) by
findings that require different paths for the bleaching signal
and for real light, though it does not necessarily explain them.
These findings are Westheimer (1968), Ernst (1968), and Naka &
Rushton (1968).

1.5 However, even Rushton's theory in its pure form is
inconsistent with some later results, published in the series of
papers by Alpern, Rushton & Torii. In experiment I of Alpern et
al. (1970c), bleaching B without background behaved as though
the rods were desensitized by the factor $b=10**12B$, not like the
equivalent background light.

1.6 Finally, although the notion of two paths seems to explain a
large number of phenomena, Rushton has given no arguments why two
such paths should have evolved, nor has he suggested what they
might be.

2 Two paths: real and pseudo light

Perhaps the most convincing evidence for the present author that Barlow's theory is incorrect was the experiments of Naka & Rushton (1968). The effect of a real background illumination is to shift the s-potential, and the effects of further stimuli appear on top of this. The effect of bleaching is however invisible (after 3 minutes) unless a light stimulus is supplied, in which case the response appears to be attenuated. The absence of a steady background potential from the bleached receptors must mean that they cannot be emitting a signal indistinguishable from that produced by real light, as Barlow's hypothesis requires.

Their other finding, that the s-potentials are already attenuated by a great amount, is I think of great significance, especially in view of the later finding in experiment I of Alpern et al. (1970c) that was referred to above. The reason is that the central issue in the controversy concerns the existence of a second path for the bleaching signal, and it turns out that the present theory has as a consequence the existence of a virtual path of about the right kind, but it depends upon receptor desensitization.   Consider the results of Naka & Rushton once more: the only effect of bleaching that they observed was that it attenuated signals due to light. Suppose that this was because of the receptors. This would explain the finding of Alpern et al. (1970c), and that result shows us that the receptors are desensitised by a factor that depends exponentially on the

bleaching. This is a regrettable property of receptors, and one
that I find surprising: but if true, it is the kind of fact of
life that one feels comfortable about having as a limit on the
performance of the visual system.

Inside the bleached area, receptor signals are attenuated by
a factor $10**kB$. Hence the local difference signal, which is the
log of a division function, is unchanged, and the bipolar signal
should be roughly unaffected. But this is not the only
consequence. Consider the effect of diffusely illuminating the
whole field of view with intensity Φ. At the boundaries of the
bleached area, the strength of the signal from the receptors will
change to $\Phi/10**kB$. This will be interpreted by the division
gradient measuring device as a boundary of size $\Phi/(\Phi/10**kB)$,
i.e. $10**kB$, which is independent of Φ, and has logarithm kB.
The reconstituting mechanism, which we saw operates on
logarithmic quantities, will therefore receive transversely
across the amacrine layer a signal that varies with kB. Thus the
bleaching effect is transmitted to the inner layer, but by a
virtual path that exists because of the way the retina normally
works.   The reconstituted image has a negative brightness, but
(like every discontinuity), will cause positive changes of size
kB to appear in both the $**+$ and the $**-$ layers. Carpenter (1972)
has shown that the appearance of after-images depends on the very
recent illumination of the retina, not on properties of the
bleaching stimulus that produced the effect; so the wrong sign

for the pseudo light's brightness is not too serious.

3 Some findings explained

The results of Barlow & Sparrock follow immediately, because
the reconstituted pseudo-image has the appropriate sized
brightness: this image however exists only in the amacrine cell
layer, where intensity is represented in logarithmic units, and
average intensity values computed there will behave like the
average of logs (as in Rushton 1965b). The electrophysiological
results of Naka & Rushton (1968) follow because of the
attenuation in the receptors, and because in the middle of a
bleached area, there is no signal from the receptors.

The result of Westheimer (1968) may be explained by looking
at the difference signal in his two situations. Let the intensity
of the small central test point be p; of the disc, d; of the
background, b; and of the equivalent bleaching disc, be d'. There
are four situations:

(1) Small real disc: the bipolar signals with and without p are
the logs of $(p+d+b)/b$ and $(d+b)/b$.

(2) Large real disc: with and without p, the bipolar signals are
the logs of $(p+d+b)/(d+b)$ and $(d+b)/(d+b)$.

(3) Small bleached disc: the bipolar signals with and without p
are the logs of $((p+b)/d')/b$ and $(b/d')/b$.

(4) Large bleached disc: $((p+b)/d')/(b/d')$ and $((b/d')/(b/d'))$.

The situations with and without bleachings are different, because

the quantities in (1) are bigger than those in (2), but those in
(3) are smaller than those in (4). The present theory thus
distinguishes the two situations: any non-linearities in the
system (which are to be expected in cases (1) and (2) if d is
large) would tend to give Westheimer's result in the case of real
light. The present model is not specific about temporal
properties of the retina - indeed this is probably quite a large
subject: the findings of Ernst (1968) may be consequences of
the receptor signal being so much attenuated.

4 Is pseudo light really the key?

Finally, the experiments of Rushton & Westheimer (1962), and
of Rushton (1965a and b) must be accounted for. As is probably
clear, all the necessary inputs are present for a gain-box type
of mechanism to operate in the amacrine layer, the gain-box
itself being probably the bipolar cell axon terminal. But this
requires that the bipolar signal be linear, and the various other
difficulties mentioned in section 3 of Part II arise. The effect
of all these difficulties is to make one ask whether these
results cannot be explained by phenomena in the outer plexiform
layer without involving the gain-box mechanism. This is a strange
question, because one of the attractions of the present theory is
that it provides the missing input that seemingly makes Rushton's
theory come true. But observe the following analysis of the
results of 1.1 above.

4.1 Rushton & Westheimer (1962)

Let us first analyse the grating experiment of Rushton &
Westheimer. Their effect holds for most people only for gratings
with periods up to about 30' at 5 deg eccentricity, which (as
they point out) is consistent with other measurements (e.g.
Hallett 1962) of the resolution of the rod channel in that
region. But according to the present theory, the raw rod signal
is not transmitted to the inner retinal layers: what is
transmitted is the sum of a large number of small logarithmic
terms log(R/H), where R is the receptor response, and H is the
horizontal cell response. Assuming that H is constant in the two
experimental cases, one can estimate the size of the signal as
follows:

Case (1): bleaching through a grating. There are n bleached
receptors, and n unbleached, so the total signal is roughly
n(logR + log(R/exp(kB1))), where B1 is the amount of bleaching
received by those rods that were bleached.

Case (2): uniform bleach. Here, all 2n receptors were bleached,
so that the signal size is about 2n(log(R/exp(kB2))).

These two expressions are the same if B1=2B2. This effect could
suffice to explain the results of Rushton & Westheimer, as
modified by Barlow & Andrews (1973).

4.2 Rushton (1965a)

In a later experiment, Rushton bleached a region of the
retina through a grating, and then measured the region's

increment threshold using the same grating at the same, and at
the opposite phase relation to the bleaching stimulus. He found
that for a grating of period 30', the threshold was the same in
both cases. Let us calculate the difference signals that we
expect according to the present theory. Suppose that the
unbleached receptor response to the test flash is R1 for the
light bars of the test grating, and R2 for the dark bars. Then if
the light bars fall on bleached receptors, the response is
$(\log(R1/\exp(kB)) + \log(R2))$; and if the dark bars fall on
bleached receptors, the response is $(\log(R1) + \log(R2/\exp(kB)))$.
In both cases, the result is the same. A similar argument can be
constructed for the other experiment (Rushton 1965b).

4.3 What does this mean?

The arguments outlined in the last two sections rest on a
large number of assumptions about the nature of the bipolar
signal, probably not all of which are correct: the bones of the
present theory would survive the disproof of many of them.
Nevertheless, it is disquieting to find that pseudo light, an
explanation of which is one of the strong points of this theory,
turns out not to be involved in the explanation of the above
phenomena, because those phenomena were part of the reason why
one was so interested in pseudo light. The reason why
explanations of the above kind have been overlooked is the
evidence about the linear nature of early retinal processes: in
particular, Ricco's Law, and the findings of Alpern et al.

Alpern, M., Rushton, W.A.H. & Torii, S. (1970c). The attenuation of rod signals by bleachings. J. Physiol. (Lond.), 207, 449-461.

Alpern, M., Rushton, W.A.H. & Torii, S. (1970d). Signals from cones. J. Physiol. (Lond.), 207, 463-475.

Barlow, H.B. (1956). Retinal noise and absolute threshold. J. opt. Soc. Amer., 46, 634-639.

Barlow, H.B. (1957). Increment thresholds at low intensities considered as signal-noise discriminations. J. Physiol. (Lond.), 136, 469-488.

Barlow, H.B. (1958). Temporal and spatial summation in human vision at different background intensities. J. Physiol. (Lond.), 141, 337-350.

Barlow, H.B. (1964). Dark-adaptation: a new hypothesis. Vision Res., 4, 47-57.

Barlow, H.B. & Andrews D.P. (1973). The site at which rhodopsin bleaching raises the scotopic threshold. Vision Res., 13, 903-908.

Barlow, H.B., Fitzhugh, R. & Kuffler, S.W. (1957). Change of organization in the receptive field of the cat's retina during dark adaptation. J. Physiol. (Lond.), 137, 338-354.

Barlow, H.B. & Sakitt, B. (1973). Doubts about scotopic interactions in stabilized vision. Vision Res., 13, 523-524.

Barlow, H.B. & Sparrock, J.M.B. (1964). The role of after-images in dark adaptation. Science, 144, 1309-1314.

Boycott, B.B., & Dowling, J.E. (1969). Organization of the primate retina: light microscopy. Philos. Trans. Roy. Soc. B., 255, 109-184.

Brindley, G.S. (1970). Physiology of the retina and visual pathway. (Physiological Society Monograph no. 6). London: Edward Arnold Ltd.

Brown, J.E. & Major, D. (1966). Cat retinal ganglion cell dendritic fields. Exp. neurol., 15, 70-78.

Bunt, A.H. (1971). Enzymatic digestion of synaptic ribbons in amphibian retinal photoreceptors. Brain Res., 25, 571-577.

Cajal, S.R. (1911). Histologie du systeme nerveux. Madrid:C.S.I.C.

Carpenter, R.H.S. (1972). After-images on backgrounds of different luminance: a new phenomenon and a hypothesis. J. Physiol. (Lond.), 226, 713-724.

Cleland, B.G., Dubin, M.W. & Levick, W.R. (1971). Sustained and transient neurones in the cat's retina and lateral geniculate nucleus. J. Physiol. (Lond.), 217, 473-496.

Cone, R.A. & Ebrey, T.G. (1965). Functional independence of the two major components of the rod electroretinogram. Nature, 221, 818-820.

Crawford, B.H. (1937). The change of visual sensitivity with time. Proc. Roy. Soc. B., 128, 552-559.

Crawford, B.H. (1947). Visual adaptation in relation to brief conditioning stimuli. Proc. Roy. Soc. B., 134, 283-302.

Denney, D., Baumgartner, G. & Adorjani, C. (1968). Responses of cortical neurones to stimulation of the visual afferent radiations. Exp. Brain Res., 6, 265-272.

De Valois, R.L. (1965). Analysis and coding of colour vision in the primate visual system. Cold Spr. Harb. Symp., 30, 567-579.

Dowling, J.E. (1960). Chemistry of visual adaptation in the rat. Nature, 188, 114-118.

Dowling, J.E., & Boycott, B.B. (1966). Organization of the primate retina: electron microscopy. Proc. Roy. Soc. B., 166, 80-111.

Dowling, J.E. & Werblin, F.S. (1969). Organization of the retina of the mud-puppy, Necturus maculosus: I. Synaptic structure. J. Neurophysiol., 32, 315-338.

Eccles, J.C. & Jaeger, J.C. (1958). The relationship between the mode of operation and the dimensions of the junctional regions at synapses and motor end-organs. Proc. Roy. Soc. B., 148, 38-56.

Enroth-Cugell, C. & Robson, J.G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. J. Physiol. (Lond.), 187, 517-552.

Ernst, W. (1968). The dependence of critical flicker fusion frequency and the rod threshold on the state of adaptation of the eye. Vision Res., 8, 889-900.

Fukada, Y. (1971). Receptive field organization of cat optic nerve fibres with special reference to conduction velocity.

Vision Res., 11, 209-226.

Fukada, Y. & Saito, H-A. (1971). The relationship between response characteristics to flicker stimulation and receptive field organization in the cat's optic nerve fibres.   Vision Res., 11, 227-240.

Fuortes, M.G.F. & Hodgkin, A.L. (1964). Changes in time scale and sensitivity in the ommatidia of limulus. J. Physiol. (Lond.), 172, 239-263.

Gouras, P. (1966). Rod and cone independence in the electroretinogram of the dark-adapted monkey's perifovea.   J. Physiol. (Lond.), 187, 455-464.

Gouras, P. (1967). The effects of light-adaptation on rod and cone receptive field organization of monkey ganglion cells.   J. Physiol. (Lond.), 192, 747-760.

Gouras, P. & Link, K. (1966). Rod and cone interaction in dark-adapted monkey ganglion cells.   J. Physiol. (Lond.), 184, 499-510.

Gray, E.G., & Pease, H.L. (1971). On understanding the organization of the retinal receptor synapses.   Brain Res., 35, 1-15.

Hallett, P.E. (1962). Scotopic acuity and absolute threshold in brief flashes.   J. Physiol. (Lond.), 163, 175-189.

Helmholtz, H. (1962) Treatise on physiological optics. New York: Dover Publications Inc. (First edition of Handbuch der physiologischen Optik published in 1867 by Voss, Leipzig).

Hoffmann, K-P. & Stone, J. (1971). Conduction velocity of afferents to cat visual cortex: a correlation with cortical receptive field properties.   Brain Res., 32, 460-466.

Horn, B.K.P. (1974). On lightness.   (To appear.)

Hubel, D.H. & Wiesel, T.N. (1960). Receptive fields of optic nerve fibres in the spider monkey.   J. Physiol. (Lond.), 154, 572-580.

Hubel, D.H. & Wiesel, T.N. (1961). Integrative action in the cat's lateral geniculate body.   J. Physiol. (Lond.), 155, 385-398.

Hubel D.H. & Wiesel, T.N. (1966). Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey.

J. Neurophysiol., 29, 1115-1156.

Kaneko, A. & Hashimoto, H. (1967). Recording site of the single cone response determined by an electrode marking technique. Vision Res., 7, 847-851.

Kolb, Helga. (1970) Organization of the outer plexiform layer of the primate retina: electron microscopy of Golgi-impregnated cells. Philos. Trans. Roy. Soc. B., 258, 261-283.

Kuffler, S.W. (1953). Discharge patterns and functional organization of mammalian retina. J. Neurophysiol., 16, 37-68.

Land, E.H. (1959). Experiments in color vision. Scientific American, 201, 16-28.

Land, E.H. & McCann, J.J. (1971). Lightness and retinex theory. J. opt. Soc. Amer., 61, 1-11.

Lennie, P. & MacLeod, D.I.A. (1973). Background configuration and rod threshold. J. Physiol. (Lond.), 233, 143-156.

McIlwain, J.T. (1964). Receptive fields of optic tract axons and lateral geniculate cells: peripheral extent and barbiturate sensitivity. J. Neurophysiol., 27, 1154-1173.

McIlwain, J.T. (1966). Some evidence concerning the physiological basis of the periphery effect in the cat's retina. Exp. Brain Res., 1, 265-271.

McKee, S. & Westheimer, G. (1970). Specificity of cone mechanisms in lateral interactions. J. Physiol. (Lond.), 206, 117-128.

Maffei, L. & Fiorentini, A. (1972). Retinogeniculate convergence and analysis of contrast. J. Neurophysiol., 35, 65-72.

Missotten, L. (1965). The ultrastructure of the human retina. Brussels: Arscia Uitgaven N.V.

Naka, K.I. (1969). Computer assisted analysis of S-potentials. Biophys. J., 9, 845-859.

Naka, K.I. & Rushton, W.A.H. (1966). S-potentials from colour units in the retina of fish (Cyprinidae). J. Physiol. (Lond.), 185, 536-555.

Naka, K.I. & Rushton, W.A.H. (1967). The generation and spread of s-potentials in fish (Cyprinidae). J. Physiol. (Lond.), 192, 437-461.

Naka, K.I. & Rushton, W.A.H. (1968). S-potential and dark adaptation in fish. J. Physiol. (Lond.), 194, 259-269.

Penn, R.D., & Hagins, W.A. (1969). Signal transmission along retinal rods and the orogin of the electroretinographic a-wave. Nature, 223, 201-205.

Rodieck, R.W. (1967). Receptive fields in the cat retina: a new type. Science, 157, 90-92.

Rushton, W.A.H. (1961). Rhodopsin measurement and dark-adaptation in a subject deficient in cone vision. J. Physiol. (Lond.), 156, 193-205.

Rushton, W.A.H. (1965a). The sensitivity of rods under illumination. J. Physiol. (Lond.), 178, 141-160.

Rushton, W.A.H. (1965b). Bleached rhodopsin and visual adaptation. J. Physiol. (Lond.), 181, 645-655.

Rushton, W.A.H. (1965FL). The Ferrier Lecture, 1962: Visual Adaptation. Proc. Roy. Soc. B., 162, 20-46.

Rushton, W.A.H. (1972). Pigments and signals in colour vision. (Invited lecture to the Physiological Society). J. Physiol. (Lond.), 228, 1P-31P.

Rushton, W.A.H. & Westheimer, G. (1962). The effect upon the rod threshold of bleaching neighbouring rods. J. Physiol. (Lond.), 164, 318-329.

Sharpe, C.R. (1972). The visibility and fading of thin lines visualized by their controlled movement across the retina. J. Physiol. (Lond.), 222, 113-134.

Singer, W. & Creutzfeldt, O.D. (1970). Reciprocal lateral inhibition of on- and off-center neurones in the lateral geniculate body of the cat. Exp. Brain Res., 10, 311-320.

Stone, J. (1972). Morphology and physiology of the geniculocortical synapse in the cat: the question of parallel input to the striate cortex. Invest. Ophthal., 11, 338-346.

Stone, J. & Dreher, B. (1973). Projection of X- and Y- cells of the cat's lateral geniculate nucleus to areas 17 and 18 of visual cortex. J. Neurophysiol., 36, 551-567.

Stone, J. & Freeman, R.B. jr. (1971). Conduction velocity groups in the cat's optic nerve classified according to their retinal origin. Exp. Brain Res., 13, 489-497.

Stone, J. & Hoffman, K-P. (1971). Conduction velocity as a parameter in the organization of the afferent relay in the cat's lateral geniculate nucleus. Brain Res., 32, 454-459.

Stone, J. & Hoffman, K-P. (1972). Very slow-conducting ganglion cells in the cat's retina: a major, new functional type? Brain Res., 43, 610-616.

Teller, D.Y., Andrews, D.P. & Barlow, H.B. (1966). Local adaptation in stabilized vision. Vision Res., 6, 701-705.

Tomita, T. (1968). Electrical responses of single photoreceptors. Proc. I. E. E. E., 56, 1015-1023.

Tomita, T., Kaneko, A., Murakami, M. & Pautler, E.L. (1967). Spectral response curves of single cones in the carp. Vision Res., 7, 519-531.

Toyoda, J., Nosaki, H. & Tomita, T. (1969). Light-induced resistance changes in single photoreceptors of Necturus and Gekko. Vision Res., 9, 453-463.

Weber, E.H. (1834). De pulsu, resorptione, auditu et tactu annotationes anatomicae et physiologicae. Leipzig: C.F.Koehler. (Cited by Brindley 1970).

Werblin, F.S. & Dowling, J.E. (1969). Organization of the retina of the mud-puppy, Necturus maculosus: II. Intra-cellular recording. J. Neurophysiol., 32, 339-355.

Westheimer, G. (1965). Spatial interaction in the human retina during scotopic vision. J. Physiol. (Lond.), 181, 881-894.

Westheimer, G. (1968). Bleached rhodopsin and retinal interaction. J. Physiol. (Lond.) 195, 97-105.

Westheimer, G. (1970). Rod-cone independence for sensitizing interaction in the human retina. J. Physiol. (Lond.), 206, 109-116.

Westheimer, G. & Wiley, R.W. (1970). Distance effects in human scotopic retinal interaction. J. Physiol. (Lond.), 206, 129-143.

Zeki, S.M. (1973). Colour coding in rhesus monkey prestriate cortex. Brain Res., 53, 422-427.