THRESHOLD SETTING BY THE SURROUND OF CAT RETINAL GANGLION CELLS

BY H. B. BARLOW* AND W. R. LEVICK†

From the Neurosensory Laboratory, School of Optometry, University of California, Berkeley, California 94720, U.S.A.

(Received 10 December 1975)

SUMMARY

1. The slope of curves relating the log increment threshold to log background luminance in cat retinal ganglion cells is affected by the area and duration of the test stimulus, as it is in human psychophysical experiments.

2. Using large area, long duration stimuli the slopes average 0.82 and approach close to 1 (Weber’s Law) in the steepest cases. Small stimuli gave an average of 0.53 for on-centre units using brief stimuli, and 0.56 for off-centre units, using long stimuli. Slopes under 0.5 (square root law) were not found over an extended range of luminances.

3. On individual units the slope was generally greater for larger and longer test stimulus, but no unit showed the full extent of change from slope of 0.5 to slope of 1.

4. The above differences hold for objective measures of quantum/spike ratio, as well as for thresholds either judged by ear or assessed by calculation.

5. The steeper slope of the curves for large area, long duration test stimuli compared with small, long duration stimuli, is associated with the increased effectiveness of antagonism from the surround at high backgrounds. This change may be less pronounced in off-centre units, one of which (probably transient Y-type) showed no difference of slope, and gave parallel area-threshold curves at widely separated background luminances, confirming the importance of differential surround effectiveness in changing the slope of the curves.

6. In on-centre units, the increased relative effectiveness of the surround is associated with the part of the raised background light that falls on the receptive field centre.

* Present address: Physiological Laboratory, Cambridge CB2 3EG.
† Present address: Department of Physiology, J.C.S.M.R., Australian National University, P.O. Box 334, Canberra City, A.C.T. 2601, Australia.
7. It is suggested that the variable surround functions as a zero-offset control that sets the threshold excitation required for generating impulses, and that this is separate from gain-setting adaptive mechanisms. This may be how ganglion cells maintain high incremental sensitivity in spite of a strong maintained excitatory drive that would otherwise cause compressive response non-linearities.

INTRODUCTION

The level of background illumination changes visual performance on tasks testing spatial and temporal summation and resolution as well as sensitivity (reviews in Barlow, 1958, 1972; Glezer, 1965). Important variations are also produced in the behaviour of retinal ganglion cells and it is of considerable interest to define these effects in order to understand how visual performance is related to ganglion cell activity. Barlow, Fitzhugh & Kuffler (1957a) gave an account of the great drop in threshold of cat retinal ganglion cells during dark adaptation and they further showed (1957b) that the familiar concentric, antagonistic, centre/surround organization of the light-adapted receptive fields became simplified after sufficient adaptation to darkness so that only the central zone could be made to yield its characteristic response. In the ganglion cells tested, a response could not be elicited from the surround, and further evidence for its weakened influence after complete dark adaptation was obtained from area-threshold curves.

These changes in spatial organization are obvious factors to be considered in discussions of spatial summation and spatial resolution, but they are not necessarily the only ones. Enroth-Cugell & Shapley (1973b) found that when background was decreased, ganglion cells with large receptive fields continued to decrease their increment-thresholds down to lower values of the background luminance than cells with small receptive fields. This provides another explanation for changes in spatial summation with luminance. Earlier work from their laboratory (Cleland & Enroth-Cugell, 1968) provided no support for changes in surround effectiveness with luminance, and recent work (Enroth-Cugell & Lennie, 1975) has emphasized that effects from the surround can, under some circumstances, be detected after complete dark adaptation. The following results were therefore assembled; they show that there can be important variations in the behaviour of different ganglion cells when background luminance is changed, but the original findings are supported. The slope (on log co-ordinates) of increment threshold curves as a function of background luminance is steeper for large than for small test stimuli, and this results from the increasing relative effectiveness of the surround at higher adaptation levels. Other experiments show that the changes in surround effective-
ness are caused by the adapting light that falls in the centre of the receptive field. It is suggested that these adjustments enable ganglion cells to transmit small changes of central luminance over many decades of mean light level, in spite of the very limited dynamic range of the volleys of impulses used to transmit the signal.

METHODS

Action potentials were recorded from single retinal ganglion cells of the cat as described by Barlow & Levick (1969a). The anaesthetic for the earlier experiments was chloralose (0·16 mg/kg/hr) plus urethane (25 mg/kg/hr) administered by continuous i.v. infusion. Later experiments were conducted under gaseous anaesthesia with 70% nitrous oxide, 27% oxygen, 3% carbon-dioxide. Eye movements were controlled by adding d-tubocurarine (0·5–1 mg/kg/hr) and gallamine triethiodide (5 mg/kg.hr) to the infusion fluid, necessitating artificial ventilation. Stimuli were provided either by electronically gated fluorescent tubes (Gerbrands & Stevens, 1964) or an electromagnetically shuttered slide-projector. Measurements of threshold were made by adjusting neutral attenuators in the light path so as to produce the weakest reliably detectable perturbation of the maintained discharge as played over a loudspeaker. In some experiments, peristimulus time-histograms (PSTHs) of the responses to repeated presentations of visual stimuli were accumulated with the aid of a multichannel analyser (Nuclear Data 180). From the PSTH the quantum/spike ratio (QSR) could be determined and a threshold measurement could be checked by application of the computational procedure already described (Barlow & Levick, 1969a).

A round artificial pupil (7 mm² area) was routinely used close to the contact lens protecting the eye, and good retinal imagery was ensured by placing a spectacle lens before the eye; the appropriate power was chosen by observing the response of a suitable ganglion cell to the passage of progressively finer grating patterns across the receptive field. Various background illuminations were provided by selecting different overhead incandescent lights and interposing different neutral gelatin filters between the eye and correcting lens. Special masks and baffles were used to ensure that the only light entering the eye was that passing through the filter and/or artificial pupil. Luminances of stimuli and backgrounds, and transmissions of filters were estimated with the aid of an SEI visual photometer. No great absolute accuracy of the estimates was required for the purposes of any of the experiments to be described.

Stimulus durations were usually manually (occasionally machine) controlled at about 1 sec, with a repetition rate of about 1 per 3 seconds. Since adaptation to darkness can be a slow process in the cat (Barlow et al. 1957b), thresholds were monitored after each change in background illumination until they became steady.

For the experiments requiring decrements of luminance (Fig. 5), various cards representing a range of grey shades from the Munsell series (Wyszecki & Stiles, 1967) were positioned immediately beneath a circular aperture cut in a grey card (reflectance about 0·5, approximately matching Munsell value 7·5). A spot of light of the same size was projected on to the aperture and adjusted in intensity so as to equate the luminance of the Munsell card within the aperture with the rest of the grey card. The match was checked with the unit under study by flipping an ophthalmic prism (4Δ) in and out before the eye, after the aperture had been centred on the receptive field. The decrement was produced by shuttering off the projected beam.
RESULTS

Effects of background luminance and stimulus parameters on sensitivity

The graphs in the left-hand column of Fig. 1 show various measures of ganglion cell performance in relation to background luminance on double logarithmic co-ordinates. In A, the threshold quantity of light (luminance increment x area of stimulus x duration) required by the ganglion cell for two particular test stimuli was determined by monitoring the impulse

![Graphs showing various measures of ganglion cell performance](image-url)

**Fig. 1.** For legend see facing page.
discharge by ear in the usual way. For the upper curve the target was of large area (4·6° diam.) and long duration (1 sec), for the lower it was small (0·29° diam.) and brief (10 msec). The curves are superficially similar to those relating psychophysically determined thresholds to background luminance, but in that case the asymptotic slopes approach 1 for large, long targets, and 0·5 for small brief ones, more closely than they do on this unit.

The other curves in Fig. 1B, C are similar, but display other measures of performance than the subjectively judged incremental thresholds so far illustrated. The use of other measures is important, since the maintained discharge and its variability, and also the form of the response PSTH, change with background luminance. Though our own experience biases us against this being an important factor, it may be thought that some of the changes so far described reflect variations in the effectiveness with which an experimenter can detect an increment, rather than changes in retinal function. Such explanations are ruled out by the other curves shown in Fig. 1B, C.

The thresholds in B were determined by a calculation based on the number of extra impulses evoked by a weak stimulus and the measured irregularity of the maintained discharge (Barlow & Levick, 1969a). The result is much the same as in A. It is clear that the slope for small brief

---

Fig. 1. A, incremental thresholds as functions of background luminance and stimulus parameters. Settings were made by monitoring impulse discharge on the loudspeaker and the values were converted to threshold quantities of light by multiplying by stimulus area and duration, to facilitate comparison with C.

B, same as A but the thresholds were calculated from PSTHs of the responses to the stimuli delivered in A, taking account of the variance of pulse number distribution measured on the maintained discharge at each level of background.

C, same as A and B, but with the ordinate displaying log QSR, obtained by dividing the quantity of light delivered by the stimuli in A by the average number of extra impulses evoked within the counting period of 200 msec or 1 sec. It is assumed that 25% of quanta incident on the cornea within the 7 mm² pupil are absorbed in rhodopsin.

D, scatter diagram of paired measurements of slopes from graphs like those in S but for large (usually 4·6°) long (0·32–1·26 sec) and small (less than receptive field centre size) brief (10 msec) stimuli. Symbols refer to different classes of units: ○, on-centre brisk-sustained; ●, off-centre brisk-sustained; □, on-centre brisk-transient; ■, off-centre brisk-transient.

E, scatter diagram of slope for QSR (e.g. upper graph in C) plotted against slope for monitored threshold (e.g. upper graph in A) for identical stimulation of each unit.

F, scatter diagram of paired slope estimates for large long and small short stimuli (as in C) with QSR as the measure of performance.
stimuli is different from that for large long ones over a wide range of backgrounds. $C$ shows the effect of the background on the average number of quanta required from a near-threshold stimulus to elicit an extra impulse (QSR). Again it is clear that log (QSR) for a small, brief test stimulus rises less steeply with log (adapting luminance) than does log (QSR) for a long duration large one.

Detailed measurements of the effect of background luminance on monitored threshold, calculated threshold, QSR, or all three have been obtained on 31 units. For large area, long duration stimuli, the slopes averaged 0.82 (25 units). For small, long stimuli they averaged 0.58 (29 units), with little difference between on-centre (0.59, $n = 14$) and off-centre (0.57, $n = 15$) units. With on-centre units, small brief stimuli gave an average slope of 0.53 ($n = 6$). On 23 units graphs of monitored thresholds were obtained as in Fig. 1A except that long duration small stimuli, not brief ones, were used for comparison with large stimuli. The pairs of slopes for each unit are plotted as a scatter diagram in Fig. 1D. In almost all cases the slope for the small stimulus was less than for the large, although the difference was small for a proportion of the units. In one instance the slopes were the same (0.57) and this unit is considered in more detail later (Fig. 3). The steepest of the curves using large test stimuli yielded a slope just over 1, and the shallowest of the curves for small stimuli had slopes just below 0.5. But the bold generalization that small, brief targets give a slope of 0.5, and large, long targets a slope of 1, would be very misleading, for two reasons. First, no unit for which both curves were completed gave the full difference: thus if the steeper curve approached 1, the shallower would usually be considerably above 0.5, and if the shallower was close to 0.5, the steeper would be less than 1. A second reason is that some curves, especially those starting shallow, tend to get steeper at high luminances, so that they cannot easily be characterized by a single figure for the slope.

The measure of performance used has only a minor influence on the shape or slope of the curves. This is brought out by the scatter diagram in $E$ where the slope of log QSR against log background (cf. $C$) is plotted against the slope when log monitored threshold (cf. $A$) is used. The points cluster near the 45° line through the origin, indicating approximate equality. The results do, however, suggest that curves are a little steeper when QSRs rather than monitored thresholds are used.

For seven units graphs like those in $C$ were obtained, with QSR as the measure of performance. The scatter diagram of the slopes is plotted in $E$, confirming the general conclusions drawn from $D$.

A special difficulty arises in doing these experiments on off-centre units. Here it is hard to detect gaps in the discharge with brief stimuli. The
variability of the maintained discharge is such that, even without stimulation, it is not unusual for pauses to occur with no impulses for times at least as long as the expected duration of a weak, brief response (50–100 msec; Levick & Zacks, 1970). To become detectable the stimulus must be increased in intensity until it causes an abnormally prolonged reduction in discharge rate, or rebound excitation. This difficulty can be avoided by using only long duration stimuli, as is done in the next section, but it means that our data are not really adequate to be sure about the differences between different types of unit with respect to the influence of stimulus duration on the slope of increment–threshold curves. We have no results to contradict the notion (see Figs. 2 and 3 below) that the most prominent differences occur in on-centre, brisk-sustained units (Cleland & Levick, 1974; X-type of Enroth-Cugell & Robson, 1966), while the least prominent change occurs in off-centre, brisk-transient (Y-type). But in the present study many units were not fully characterized.

**Area–thresholds and increment–thresholds**

The next step was to link the differences of slope to changes in the effectiveness of the surround that can be revealed by area–threshold curves. The right side of Fig. 2 shows a pair of curves obtained on an on-centre cell by plotting the logarithm of the threshold luminance for a series of centred stimulus spots against the logarithm of spot area. The lower curve refers to measurements at zero background illumination and the upper at a background sufficient to raise the increment–threshold for a small spot by almost 2 log units. Both curves have in common a short initial region where threshold fell as stimulus size increased, but at larger diameters there was substantial departure from similarity: the upper curves showed a much more pronounced rise for large spots. In both curves the rise in threshold with large spots occurred when the stimulus spread on to the surround zone of the receptive field. This unit was of the on-centre/off-surround type; light falling upon the surround diminished the excitatory effect of light simultaneously falling on the central on-zone, and the threshold was elevated. To reveal the inhibitory effect of the surround, area–threshold curves must be done with long duration stimuli, for the rise in threshold at large areas is often not seen with brief stimuli (Barlow et al. 1957b).

The shape of the upper area–threshold curve resembles that of the sustained-type of cat retinal ganglion cell (X-type of Enroth-Cugell & Robson, 1966) rather than the transient or Y-type (Fig. 2 in Cleland, Levick & Sanderson, 1973). Although other detailed tests were not carried out, the ganglion cell of Fig. 2 was most probably of the brisk-sustained type (Cleland & Levick, 1974), and we conclude that in this cell the
inhibitory effect was more prominent at high background levels than at low.

The left side of Fig. 2 shows increment–threshold curves (for the same unit) as functions of background illumination for small (0.5° diam) and large (4°) stimulus spots. As in human psychophysical experiments, the threshold for a large-area stimulus changes more steeply than for a small

![Graph](image)

Fig. 2. Incremental thresholds as functions of background luminance and spot diameter. In the left part, raising the background luminance elevated the threshold for a 4° diameter spot to a greater extent than the threshold for a 0.5° spot. Stimulus durations were 1 sec. In the right part, increment–threshold was measured as a function of spot diameter at zero background (lower right) and at 3.5 x 10⁻¹ cd/m² (upper right). The pupil area was 7 mm². As the spot spread on to the region of the antagonistic surround the threshold intensity rose; the rise was much greater at the higher background luminance. Horizontal dashed lines connect points determined under the same conditions in right and left halves of the Figure. The crosses and dot-dashed curve show the incremental threshold for the off-response to an annular stimulus of inner diameter 0.5°. Off-responses were unobtainable below 3.4 x 10⁻⁴ cd/m².

stimulus. Dashed lines connect points determined under the same conditions in the two halves of the Figure, and they make it quite clear that the steeper slope of the increment–threshold curves for large spots is caused, in this single unit, by the especially raised threshold for large spots, i.e. by increased surround inhibition at elevated backgrounds.

Also shown in Fig. 2 left, is, the increment–threshold for obtaining an off-response to a centred annular stimulus of inner diameter 0.5°. In agreement with earlier work, off-responses were unobtainable at the lowest backgrounds, regardless of increment luminance of the annulus. At background levels of about 3.4 x 10⁻⁴ cd/m² off-responses become apparent, and at higher levels the thresholds began to rise in company
with those for the 0.5° spot. The annulus–threshold curve thus provides additional evidence of a progressive augmentation of surround potency relative to the centre as background level increased. It should, however, be noted that the slight rise of the area–threshold curve (lower right) at large spot diameters would indicate that the surround still had a slight inhibitory influence at zero background intensity.

**Fig. 3.** Same experiment as in Fig. 2 but with an uncommon off-centre cell. The area–threshold curves were not significantly different in shape at the two adaptation levels, showing that in this unit there was little relative gain in potency of the surround at the higher adaptation level. The increment–threshold curves were also parallel. When the relative potency of lateral inhibition does not change with adaptation level, stimulus area does not influence the slope of the increment–threshold curves.

Not all units behaved in the manner of Fig. 2. Fig. 3 shows results of the same type of experiment carried out on an off-centre ganglion cell. Incremental thresholds for large and small stimuli differed by a constant amount at all backgrounds, and the two area–threshold curves also ran parallel. Some off-centre units showed only small changes in surround effectiveness with level of background illumination; the unit in Fig. 3 was unusual in showing none at all. It provides a demonstration that where there was no change in surround effectiveness, the slope of the increment–threshold curve was not dependent upon stimulus area.

This result may be cross-referenced to the responses of off-centre units to annular stimuli at low or zero backgrounds (fig. 11 in Cleland et al. 1973). There, PSTHs are shown of one ganglion cell (a sustained or X-type) with no detectable on-response from the surround together with another (a transient or Y-type) where both on- and off-responses persisted at low backgrounds. The unit of Fig. 3 was probably an off-centre brisk-transient
cell (Y-type) because the receptive field map showed a centre of about 1.5° diameter, and both on- and off-responses were obtained with full field flashes.

Control of surround effectiveness

The next question addressed was: 'what causes the change in surround effectiveness?' Barlow, Hill & Levick (1964) showed on rabbit retinal ganglion cells that the response from a latent surround can be brought out by applying a constant adapting light to the central zone. Such a technique has also been used by Wiesel & Hubel (1966) to bring out the surround response of monkey geniculate cells, and by Enroth-Cugell & Pinto (1972) in cat ganglion cells. The effectiveness of the technique was demonstrated with PSTHs from cat retinal ganglion cells by Cleland et al. (1973). The experiment to be described was devised to analyse the relations between light-adaptation of centre and surround on the one hand and responsiveness of the centre and surround mechanisms on the other.

In the lower left part of Fig. 4 are shown area–threshold curves for the on-response of an on-centre ganglion cell at two levels of uniform background illumination: the higher was $5 \times 10^{-2}$ cd/m² and the lower was 2 log units dimmer. For the larger spots, the curve flattens out at the lower background but shows a definite rise at the higher level. The difference is in the same sense as in Fig. 2 and again indicates that the antagonistic surround was more effective at the higher level. The shape of the area threshold relation at the higher background resembles that of the transient-type (cf. fig. 2 in Cleland et al. 1973) of cat retinal ganglion cell (Y-type of Enroth-Cugell & Robson, 1966). Although other detailed tests were not carried out, the ganglion cell of Fig. 4 was most probably of the brisk-transient type (Cleland & Levick, 1974).

A slide projector was then arranged to project steadily a series of disks of the higher luminance upon a more extensive background of the lower luminance, all the disks being centred on the receptive field. For each size of disk the threshold was measured for a small (0.4°) spot in the centre of the receptive field and also for a centred annular stimulus (2.1° inside diameter). The pair of curves in the upper left part of Fig. 4 show how these thresholds changed with the area of the added background field. Consider first the threshold for the central spot. The curve begins on the left at the point marked $S_1$ corresponding to zero diameter for the adapting field, i.e. the uniform background at the lower level. This is a duplicate determination of the point marked $S_1$ on the area–threshold graph below. The curve ends on the right at the point marked $S_2$ corresponding to the largest adapting field. This is again a duplicate determination of the point $S_2$ on the upper of the two area–threshold graphs. For intermediate sizes of the bright adapting spot, the threshold for the small
Fig. 4. Sensitization of the antagonistic surround by adapting light on the centre. At lower left are area–threshold curves upon uniform adapting fields of $5 \times 10^{-2}$ cd/m² (upper curve) and $5 \times 10^{-4}$ cd/m² (lower). In the graph at upper left, the adapting field was non-uniform: a central spot at $5 \times 10^{-2}$ cd/m² within a much larger region at $5 \times 10^{-4}$ cd/m². Incremental thresholds are plotted as functions of the diameter of the adapting spot for a small ($0.4^\circ$) centred test spot ($S_x - S_x - S_2$, circles and continuous curve, on response), and a concentric annular stimulus ($A_x - A_x - A_2$, squares and dashed curve, off response). The extent of the annular stimulus ($2.1^\circ$ inner diameter, $16.6^\circ$ outer) is indicated by the lines $A_1$ and $A_2$ in the area–threshold graph; the vertical positions of the lines give the thresholds for the annulus upon the two uniform adapting fields. On the right are PSTHs (for spot stimuli at the top, annular stimuli at the bottom) of the accumulated responses to 20 presentations of the stimuli, bin width 20 msec, scale calibrations at top right corner. The horizontal bar on each PSTH indicates stimulus duration and the labels above link the responses to measurements made under corresponding conditions and shown on the other graphs of the figure. A measure proportional to QSR was obtained from the PSTHs by computing the average extra impulses per sweep above the level of maintained discharge within a 600 msec period starting at the beginning of the response (on-response with spot, off-response with annulus) and dividing this into the stimulus luminance (unit: $\mu$cd/m² per spike). The result is plotted against adapting spot diameter on double log scales at middle right. Stimulus luminances (scotopic cd/m²), $S_x$: 0.013; $S_2$: 0.16; $S_2$: 0.13; $A_1$: 0.0081; $A_2$: 0.00048; $A_2$: 0.013.
spot rises as shown by the solid curve, reaching a value not significantly different from $S_2$ at the stage when the adapting field just covers the whole of the central zone of the receptive field, i.e. the plateau in the top curve starts at nearly the same adapting-spot diameter as the minimum of the area-threshold curve at the bottom. Apparently, adapting light on the receptive-field centre desensitizes the centre mechanism, whereas adapting light on the receptive-field surround has little influence on the sensitivity of the centre mechanism. This confirms and extends the results obtained by Cleland & Enroth-Cugell (1968), Enroth-Cugell, Lennie & Shapley (1975) and Sakmann, Creutzfeldt & Scheich (1969).

Now consider the threshold for an off-response to the annular stimulus. In the lower left part of the Figure, the lines $A_1$ and $A_2$ indicate by their common horizontal extent the thickness of the annulus (use log diameter scales) and by their vertical positions the thresholds for an off-response at the two uniform backgrounds: $5 \times 10^{-4}$ cd/m$^2$ and $5 \times 10^{-2}$ cd/m$^2$ respectively. In the upper part of the Figure, as the diameter of the adapting spot increased the threshold for the annulus moved along the dashed line from $A_1$ to $A_2$. It should be especially noted that the smaller adapting spots lowered the annulus–threshold. As the adapting field was extended to cover the antagonistic surround, the annulus–threshold then proceeded to rise as expected.

The situation was studied in more detail by accumulating PSTHs of the responses at the critical points indicated by open symbols in the top left graph. These are shown on the right side at the top (spot stimulus) and bottom (annulus), and they yielded the measurements of the QSR for increments which are plotted on the graph in between. The points approximately match in their relative positions the corresponding points for the monitored thresholds (top left). The sensitization to the annular stimulus by small adapting spots is thus reflected in the QSR as well as in the threshold determined by ear. It may also be appreciated directly from the PSTHs at the bottom right: the off-response to the annulus in the presence of the $1^\circ$ adapting spot ($A_x$) is clearly larger than that with the uniform low background ($A_1$), yet the stimulus luminance for $A_x$ was smaller by a factor of about 17.

Similar observations were made on several on-centre units with essentially similar results. The apparent sensitization of surround responses by centre adaptation was sometimes greater, sometimes less than in Fig. 4. This agrees with the observations of Cleland et al. (1973, fig. 11) that there were significant variations in the form of the surround responses unmasked by adaptation of the centre, depending upon the type of ganglion cell (on-centre or off-centre, sustained or transient).
Working point and response gradient

A further lead as to the possible role of the antagonistic surround in ganglion cell function comes from the following results. A small circular test area centred on the receptive field of a unit was given an increment (on-centre cell) or decrement (off-centre cell) of illumination. From PSTHs of the responses, the average number of impulses per stimulus duration was calculated and plotted on linear scales against the actual luminance of the test area during the stimulation. This display differs from the usual where *increments* of response above the maintained discharge are plotted against *increments* (or decrements) of the stimulus above (or below) the background level, and is intended to emphasize a relatively neglected aspect of the response characteristic.

Fig. 5. Response functions of on-centre (○) and off-centre (●) ganglion cells for small spot stimulation (0.29°, 0.96 sec for on-centre cell; 1°, 1.26 sec for off-centre) at different backgrounds. The points plot the average number of impulses (for ten repetitions) occurring throughout the duration of the stimulus as a function of the luminance of the test spot. The level of background illumination in each case is indicated by the vertical dotted lines in A, B, C. The points plotted at the bottom of the lines correspond to the level of maintained discharge in a period equal to the duration of the stimulus. The results at three backgrounds are combined on a common horizontal axis in D, but the individual points are suppressed to avoid clutter.
Fig. 5 shows the results for both an on-centre and an off-centre cell at three levels of background illumination indicated by the vertical dotted at high background lines in A, B, C. Weak stimuli were used and the relations were approximately linear. It is obvious that at high back-
grounds the downward extrapolations of the relations would intercept
the horizontal axis well away from the origin at a value close to the
background luminance. In other words, decrements (or increments)
representing only a small fraction of the illumination on the test area
would be sufficient to reduce the discharge to zero.

As the background illumination was reduced the sensitivity progressively
increased as indicated by the steepening of the relations when all are
plotted on a common scale in D. At the same time the intercepts with the
horizontal axis moved relatively closer to the origin (on-centre cell) or
towards a value relatively further above the background illumination
(off-centre cell).

DISCUSSION

The results suggest that there is a dual control system in the retina that
adjusts the ganglion cell’s threshold for impulse generation, or zero-offset,
as well as the slope of its response function, or gain. The evidence for
control of the impulse–threshold is the shift of the intercept of the response
function with the zero impulse axis that has just been described in relation
to Fig. 5. These shifts result from changes in the relative effectiveness of
the surround occurring progressively over several decades of background
luminance. Raising background luminance elevates the thresholds for
detection of all stimuli, but we think that the sensitivity to stimuli con-
fined to the centre of a receptive field is maintained at a relatively high
value by virtue of the adjustments of the total excitatory input required
for impulse generation brought about by variations in relative effectiveness
of the antagonistic surround.

The questions to be discussed are: first, whether our results conflict with
any other published work; second, whether our interpretations are correct;
and third, how far they go towards explaining psychophysical changes in
spatial and temporal summation and resolution. Finally we comment on
the roles that gain and offset controls play in enabling the retina to signal
effectively over a wide dynamic range.

Possible conflicting evidence

Though others have not emphasized, and sometimes have not found,
evidence for either peripheral antagonism or changes in the amount of it
with adaptation we think that these differences can be readily explained.
First, evidence for lateral inhibition is not found in all cells by doing
area–threshold curves. As Wiesel (1960) showed, it is less evident in cells with large receptive fields and Cleland et al. (1973) showed it was less evident in transient (Y-type) than in sustained units. Cells favourable for showing lateral inhibition (small field, sustained, X-type) are probably encountered relatively less often in the optic tract than in retinal recording, and Cleland & Enroth-Cugell (1968) deliberately confined their attention to spots smaller than 3° because they were primarily interested in the properties of the field centres. Thus we do not think the absence of surround effects or changes with adaptation that they reported are surprising.

The greatest difference of slope of increment–threshold curves are found by comparing large, long duration stimuli with ones that are not only small, but also brief. Enroth-Cugell & Shapley (1973a) found a decrease from a slope of 0.9 to 0.6 simply by using 8 Hz rather than 0.4 Hz without changing the spatial extent of their 0.45° stimulus. We would have liked to attribute this to diminished lateral inhibition at the higher frequency, since it was shown previously that the upturning of area–threshold curves may not occur if brief test stimuli are used (Barlow et al. 1957b). But this cannot be right since the shallower increment–threshold curve at 8 Hz was shown by a small stimulus and in this case very little light fell on the surround; changes in its effectiveness could not therefore influence the slope. Rather, Enroth-Cugell & Shapley’s result suggests that there may be ‘delayed temporal antagonism’ as well as the antagonism from the spatially displaced surround of the receptive field. Furthermore, since Y-cells show less lateral inhibition, it is tempting to speculate that Enroth-Cugell & Shapley’s result was obtained on Y-cells, and that these cells make good with variable delayed antagonism what they lack in variable lateral inhibition.

To summarize we think that different slopes of increment–threshold curve may not be obtained when recording from Y-cells, especially if a large enough range of stimulus diameters is not used, and also if the duration of the stimuli is not changed. We failed to find it in one brisk transient (Y-type) off-centre cell, but think it can be regularly found on brisk sustained cells (X-cells), especially on-type.

An alternative interpretation

Increased light falling in the central zone of the receptive field appears to increase the effectiveness of the surround, possibly because its inhibitory influences cannot be manifested unless there is a high level of ongoing activity to be inhibited. But an alternative explanation attributes the effect of light on the centre solely to desensitization of the centre type response; in Fig. 2, it is supposed that some scattered light from the
annular stimulus falls on the centre, and perhaps also some direct light if the inner annulus diameter is only 0.5° as in Fig. 2, or if it is decentred. If the centre is at maximum sensitivity, the effect of this scattered light may be enough to antagonize and nullify the effect of the much brighter light falling on the surround. An interpretation along these lines can clearly also be applied to the result of Fig. 4.

We thought it might be possible to exclude this hypothesis by comparing the latency for 'sensitization of the surround' with the latency for 'desensitization of the centre', but these attempts failed because, as Enroth-Cugell & Lennie's results (1975) prove, the effect of light on the centre occurs very rapidly. The only remaining argument against the centre desensitization hypothesis is the continuous and graded increase in surround effectiveness that is indicated by the fact that the difference in slope of small and large increment-thresholds is maintained over a range of two or more decades of background luminance. This feature of the results does not emerge readily from the densitization hypothesis, though it could be fitted by postulating the appropriate relation between surround desensitization and background luminance. So we must allow that desensitization of the centre may be the cause of relative sensitization of the surround.

Relation to psychophysics

When the mean illumination to which the eye is adapted is changed, its performance measured psychophysically in humans changes in characteristic ways which can be summarized as improvement in spatial and temporal summation at low mean luminances, and improvement in spatial and temporal resolution at high mean luminances (Glezer, 1965; Barlow, 1972).

The fact that human thresholds for large and long stimuli are more affected by adapting backgrounds than thresholds for small brief ones is obviously related to our demonstration that the thresholds of retinal ganglion cells behave similarly, and to Enroth-Cugell & Shapley's demonstration that high temporal frequencies of stimulation also reduce the rate at which threshold rises with background. Furthermore, we think that the limiting slopes of increment–threshold curves on logarithmic coordinates of 0.5 and 1 match quite well the performance seen psychophysically, but there is another factor mediating change in performance with luminance that must be considered.

When background luminance is raised from zero there is at first no effect on increment–threshold but above a certain value of the background the threshold starts rising regularly. Enroth-Cugell & Shapley (1973b) showed that this minimum effective background, which is equivalent to the dark light measured psychophysically, varies from unit to unit, and is, to a
first approximation, inversely proportional to the effective area of the central zone of the unit’s receptive field. At very low luminances threshold will be determined by large field units capable of extensive spatial summation, but as background luminance is raised their threshold will be elevated whereas small field units will initially be unaffected. Consequently at a higher background luminance these small field units with good spatial resolution become operative.

We cannot tell how much each of the mechanisms contributes to changes of psychophysical performance with mean luminance, and it is probably worth pointing out other major gaps in our current knowledge. Is the human scotopic system adequately modelled by the cat? What role does retinal eccentricity play in determining visual performance at different backgrounds? What role does the type of neurone (sustained or transient) play? And can spatial or temporal summation, beyond that mediated by ganglion cells, occur centrally, or does the most sensitive ganglion cell, for the stimulus employed, determine threshold?

The experiment of Fig. 4 is a neurophysiological parallel of the psychophysical experiments of Crawford (1940) and Westheimer (1965, 1967) on the effect of the area of an adapting field on the threshold for a small test stimulus. A similar result was expected, namely sensitization of a ganglion cell’s centre-type response by adapting light falling in the surround of the receptive field, but this was not found. This failure is evidence against these particular adaptation effects occurring at a retinal level (see Lennie & MacLeod, 1973), and favours a central location for the underlying interactions (Sakmann & Creutzfeldt, 1969; Nakayama, 1971).

Inhibitory mechanisms and dynamic range

When background luminance is raised the threshold for the centre is raised less than the threshold for illumination of the whole receptive field. It may appear paradoxical to attribute this favouring of small stimuli, or sparing of the centres’ desensitization, to the progressive activation of an inhibitory mechanism since one ordinarily expects inhibition to have a depressive effect. The following consideration should make the situation plain.

The standard deviation of the number of impulses occurring within the time occupied by a brief response (say 0.1 sec; Levick & Zacks, 1970) is about 1 or 2 (Barlow & Levick, 1969a), and the maximum possible number of impulses within such an interval is of the order 50–100. Because the ratio of maximum output to standard deviation is only about 50 the dynamic range of the output is clearly rather small. By contrast the dynamic range of the input is very large, not only in real life but also in the laboratory experiments we are discussing. This is of course well known,
and Fig. 6 shows three ways by which it has been supposed that the retina handles the problem.

Here the scales are linear, and the top curve (Fig. 6 A) shows Fechner's well known suggestion that the output is proportional to the logarithm of the input. This has the merit that equal absolute increments of the output correspond to equal fractional increments of the input, thus accounting for Weber's Law. But because the available dynamic range is utilized to cover the whole input, only a small part is available at any given adaptation level. It is doubtful if a Weber fraction ($\Delta I/I$) of 0.1 or less could be maintained over the full range of inputs on this scheme.

**Fig. 6. Suggested methods of utilizing the narrow dynamic range of a retinal ganglion cell’s output.**

- **A**, represents Fechner's relation $R = K \log (I/C)$ where $R =$ impulses per analysis period, $K =$ a constant = 10, $I =$ luminance in cd/m², and $C =$ another constant = $7.7 \times 10^{-4}$ cd/m².
- **B**, represents parametric feed-back or gain control, assuming a linear output for simplicity; if the feed-back holds the maintained discharge low and constant at all backgrounds ($I_1$, $I_2$, $I_3$), the slope, which determines incremental sensitivity, is about the same as the slope at the corresponding luminance in A; thus the decrease of incremental sensitivity with background is just as rapid.
- **C**, shows the advantage of control of intercept as well as gain; for light on the centre the slopes decrease only as $\sqrt{\text{(adapting luminance)}}$ thus maintaining higher sensitivity at high adapting luminances. Responses from the antagonistic mechanism (surround or delayed antagonism) follow the dotted lines. Responses from stimuli applied to both follow the mean of the two lines and would look like B. For off-centre units, the interrupted lines represent centre; continuous lines the surround.
Fig. 6B shows an alternative model in which it is assumed that there is no static characteristic relating input to output, but by parametric feed-back (Fuortes & Hodgkin, 1964) or by means of a gain-control box (Rushton, 1965) the characteristic changes with adaptation level. But this does not necessarily bring any advantage with regard to sensitivity, for if the feed-back is such that the maintained discharge is held low and constant, the slope of the output is about the same as the slope at the corresponding adaptation level on the Fechner model. Impulse frequency would be proportional to luminance, which is the case for weak stimuli, but since there is no additive constant involved the line of proportionality should always pass through the origin. This is certainly not the case at moderate or high backgrounds, where a small reduction of centre illumination reduces the maintained discharge to zero, and an increase by, say, 25% is caused by less than a 25% change of centre luminance (Fig. 5).

The ganglion cell actually appears to be constantly balancing centre excitation against the surround’s inhibition and is thus able to maintain high incremental sensitivity, provided that the stimulus light is confined to the centre. Fig. 6C shows this interpretation in which the centre’s sensitivity is preserved by using the surround as a mechanism setting the threshold excitation required for impulses, analogous to the use of zero offset on a recording instrument, or of a suppressed zero when plotting a graph (Barlow & Levick, 1969b, c). The surround cannot, however, be the only factor providing zero offset and reducing the maintained discharge at high backgrounds, for restricting such backgrounds to the centre causes only a slight increase in discharge rate (Sakmann & Creutzfeldt, 1969; Enroth-Cugell & Lennie, 1975). The results of Enroth-Cugell & Shapley (1973a) when they compared 0.4 and 8 Hz stimuli suggest that there is also a temporally delayed mechanism that is activated over the whole receptive field, not just the surround. This could preserve high sensitivity to temporal transients in the same way that the surround preserves sensitivity for stimuli confined to the centre.

O’Shea & Fraser Rowell (1975) suggest that lateral inhibition of movement detectors in the locust serves the purpose of protecting excitatory synapses from habituation; if one equates adaptation with habituation our suggestion is the same.

The decline or adaptation of sensory responses to maintained stimuli is one of their best known characteristics. Because of the simplicity with which a blocking capacitor produces similar effects in an electronic circuit, and because one is so familiar with this use, capacity-coupling is the natural analogy to adaptation that springs to mind. But many of the physiological causes of failure to maintain a response, such as inactivation of a carrier mechanism, exhaustion of a transmitter substance, or fatigue, would be
quite different from the effect of a capacitor in that incremental sensitivity would decline together with the response. These causes of adaptation would be more closely analogous to a gain control mechanism than to capacity coupling. Variable lateral inhibition and delayed self-inhibition may be the nervous system's tool for controlling this potential cause of sensitivity and information loss. This point may be particularly important when trying to find the anatomical basis for physiological functions. Zero offset control requires access to separate comparison pathways, and this should have a clearly visible anatomical basis. On the other hand, if the reduction of incremental sensitivity is the more or less inevitable consequence of strong maintained excitation, this may well occur diffusely and in several stages so that the hunt for an anatomical gain box will be fruitless.

This research was supported by grant EY 00276 (formerly NB 05215) from the USPHS. Expert technical assistance was rendered by Dr Gerda Trautner.

REFERENCES


