J. Physiol. (1969), **200**, *pp.* 1–24
With 9 text-figures
Printed in Great Britain

THREE FACTORS LIMITING THE RELIABLE DETECTION OF LIGHT BY RETINAL GANGLION CELLS OF THE CAT

By H. B. BARLOW and W. R. LEVICK*

From the Neurosensory Laboratory, School of Optometry, University of California, Berkeley, California 94720

(*Received* 14 *March* 1968)

SUMMARY

1. Responses of cat retinal ganglion cells have been examined with a view to specifying the characteristics that limit the detection of light stimuli.

2. Threshold is defined as the weakest stimulus that can be reliably detected by examination of the output from a retinal ganglion cell; it depends upon (*a*) the quantum/spike ratio, which is the mean number of additional quantal absorptions required to produce an additional impulse, (*b*) the temporal course of the response, which determines the time interval within which the maintained discharge is modified, and (*c*) the statistical distribution of the number of impulses that occur in this time interval in the absence of the stimulus.

3. The quantum/spike ratio changes greatly when adapting luminance is changed, and this is the predominant factor accounting for changes in increment threshold.

4. The time course of the response changes with adaptation level and area of the stimulus. This may account for the changes in temporal integration that occur in analogous psychophysical experiments.

5. Changes in the irregularity of the maintained discharge also affect the threshold of single ganglion cells. This is only a minor factor in the conditions of most of our experiments, but it may be important when unstabilized images and non-equilibrium adaptation conditions are encountered.

INTRODUCTION

The visual system functions effectively over an enormous range of luminous intensities, and for a remarkable range of discriminatory tasks. This performance has been investigated thoroughly by psychophysical

* Present address: Department of Physiology, Australian National University, Canberra, Australia.

methods, and it is well understood at a descriptive level; as a result it is now clear that changes in the state of adaptation are not simply the result of changes in photochemical equilibrium, but are largely brought about by the plasticity of the nervous pathways of vision (Rushton, 1963). Furthermore, for adaptation to luminance changes, this plasticity must be situated in the retina, as shown by the earlier studies of the output of the retinal ganglion cells, and also by the fact that the electroretinogram adapts (Dowling, 1967).

Since much is known about this plasticity at the behavioural level of psychophysics, it is a particularly attractive situation in which to attempt a deeper analysis of physiological mechanisms, but there is a difficulty in relating the two approaches. Psychophysically, the simplest measure of the performance of a certain task is to reduce the physical magnitude of the feature that is being detected until the task can only just be performed. This is the psychophysical threshold, but there are arguments and doubts on the nature of the underlying events. The classical view is that an abrupt, qualitative, change in the output occurs at threshold, whereas this is now challenged by signal detection theory (Tanner & Swets, 1954; Barlow, 1956, 1957); from this viewpoint, a psychophysical threshold is simply the lowest level at which the response evoked by the stimulus is large enough to be reliably detected against the background noise in the sensory pathways.

We think signal/noise considerations must be important because of the maintained discharge; recordings from single sensory neurones almost always show this continuous, irregular, traffic of impulses even in darkness when no stimulus is applied (Granit, 1955; Kuffler, FitzHugh & Barlow, 1957; Bornschein, 1958; Arduini & Pinneo, 1962; Levick & Williams, 1964; Rodieck, 1967). FitzHugh (1957, 1958) has performed a preliminary analysis of the changes that occur with weak stimuli, and in this paper we consider in more detail the problems of applying signal detection theory to the trains of impulses from single retinal ganglion cells. We show that three factors determine whether a stimulus is above or below the level that causes a reliably detectable change in the maintained discharge. The quantum/spike ratio, or average number of quantal absorptions required to elicit one extra impulse, is the most important factor because it changes most, and its control obviously corresponds to the range-setting mechanism of Craik (1938), or the automatic gain control of Rose (1948), Fuortes & Hodgkin (1964), and Rushton (1965). It is not the only factor, however; the time course of the response and the statistical characteristics of the maintained discharge are also important, and we point out conditions where each of these is probably the variable responsible for changes of threshold. In later papers we shall give quantitative results on thresholds,

and also a further analysis of the maintained discharge and its statistics, using the methods developed here.

METHODS

The action potentials of single retinal ganglion cells, or sometimes optic nerve fibres, were recorded within the eye of the cat. The conjunctiva was sewn to a ring, and the cornea covered with a contact lens. Supplementary spectacle lenses were used as required, and an artificial pupil of area 7 mm^2 was routinely placed immediately in front of the contact lens. Tungsten-in-glass micro-electrodes were advanced on to the interior surface of the retina through a fine cannula penetrating the sclera near the limbus.

Anaesthesia was induced with methoxyfluorane (Metofane); for the surgical procedures anaesthesia was continued with ether and thiamylal sodium (Surital). For the experiments the cat was maintained on 70 % nitrous oxide, 27 % oxygen, 3 % carbon dioxide, and immobilized by infusion of D-tubocurarine at 0·5–1 mg/kg.hr and gallamine triethiodide (Flaxedil) at 5 mg/kg.hr. The flow rates of oxygen and nitrous oxide were continuously monitored and a constant depth of anaesthesia was reliably produced with a low risk of failure in any part of the system. Relaxant was withheld periodically to check the level of anaesthesia: at this concentration cats have brisk reflexes but no organized responses to pain-producing stimuli.

In some earlier experiments anaesthesia was induced with ethyl chloride, and maintained after surgery by continuous infusion of chloralose and urethane in distilled water (0·625 mg/kg.hr and 18·75 mg/kg.hr).

Action potentials tripped a Schmitt trigger to produce pulses of standard amplitude and duration. These were counted by a multichannel scaler (ND-180-FMITB, Nuclear Data, Inc., Schaumburg, Illinois) and for some purposes they were also recorded on tape. The multichannel scaler was arranged to accumulate post-stimulus time-histograms, histograms of numbers of impulses per unit time (number distributions), and interval histograms (Levick, 1962). Switch-selected states of the address register were used to control the stimulus, which was provided either by electronically gated fluorescent tubes (Gerbrands & Stevens, 1964) or by an electromagnetically shuttered projector. Timing thus depended upon the crystal oscillator that controlled address advance in the ND-180, and was very accurate. The states of the address register were also used to provide gating for a counter. This gate was held open for a certain number of addresses selected to match the period during which the stimulus was known from the post-stimulus time-histogram to elicit a response. In this way the counter totalled the number of impulses occurring within the analysis period. By using the same gates it was possible to obtain the number of impulses in any interval during a single trial stimulus during the course of an experiment.

There were four alternative ways of reading out the memory of the multichannel scaler. The quickest method, used most often during the course of an experiment, was to read out the memory contents pulse by pulse at 1600 pulses/sec on magnetic tape, from which the data could be re-entered into the memory for later manipulation. Full numerical print-out of the memory contents could be obtained by a teletype page printer, and this also had a tape-punch yielding the full memory contents on paper tape. This took about 8 min for 512 channels, and a pen recorder was arranged to produce graphical records such as Fig. 1 at the same time. Paper tapes from the teletype were analysed in a programmed desk computer (Mathatronics, Inc., Waltham, Massachusetts) to yield distribution parameters such as those plotted in Fig. 4.

No attempt was made to obtain great accuracy in the photometric calibrations since we were primarily interested in relative values. Luminances were determined with a visual photometer (SEI, Salford Instruments), the calibration of which was checked from time to time against a source standardized by the National Physical Laboratory. Colour tempera-

tures were measured with a photo-electric colour temperature meter (Weston 'Megatron'), and the photopic luminances were converted to scotopic using these figures. Scotopic luminances were expressed as the number of quanta of 507 nm entering the artificial pupil per second and square degree. We have provisionally assumed that the cat has scotopic luminosity identical with that of the human, and that at 507 nm, 25 % of quanta incident at the cornea and passing the pupil are absorbed by the rods. Thus, at a pupil area of 7 mm^2 a luminance of 3·43 scotopic cd/m^2 (or 1 scotopic ft.-lambert) supplies $2{\cdot}7 \times 10^6$ quanta/sec per square degree absorbed. At a colour temperature of 2854° K, 1 scotopic ft. lambert is equivalent to 0·7 photopic ft. lambert, or 2·4 photopic cd/m^2. According to the data of Vakkur, Bishop & Kozak (1963), 220 μ on the retina subtends about 1° at the posterior nodal point. All the receptive fields lay within the region backed by the tapetum.

RESULTS

This section falls into three parts. Figure 5, at the end of the first part, explains how threshold is defined as the stimulus intensity required to cause a significant change in the discharge of the single retinal ganglion cell on 50 % of trials. This treatment as a signal detection problem requires estimating three properties of the trains of impulses from ganglion cells, and the first part explains how this is done. The second part gives formal definitions of the quantities used, and describes experimental controls that validate some of the assumptions made. Finally, in the third part we show how each of the three properties is important, in particular circumstances, in determining the threshold.

The three factors limiting sensitivity

Quantum/spike ratios and the time course of the response. Figure 1 shows post-stimulus time-histograms for various conditions of stimulation. Consider first the burst of impulses occurring as a result of a brief stimulus applied to the centre of the receptive field of an on-centre unit (Fig. 1*A*). The average number of impulses caused by the stimulus can be obtained by totalling the counts in the channels occupied by the response, subtracting the average number of impulses in the same number of channels just before the stimuli were given, and dividing by the number of repetitions of the stimulus. In this case 3·0 impulses were caused on an average by each stimulus. From our light calibrations we know that each flash led to the absorption of $4{\cdot}5 \times 10^5$ quanta on average: hence the quantum/spike ratio is $1{\cdot}5 \times 10^5$.

Consider next the response to a 1 sec stimulus applied to the centre of an on-centre unit (Fig. 1*B*). As before, one can obtain the extra impulses by examining the counts in the channels occupied by the response and subtracting out the average number expected in an equal number of channels before the stimulus: the total number of quanta delivered during the 1 sec stimulus is known, and hence the quantum/spike ratio can be obtained. One can perform the same calculation on the inhibitory response

of the off unit illustrated in Fig. 1*C*, and in this case the quantum/spike ratio is negative because an increment of quanta causes a decrement of impulses.

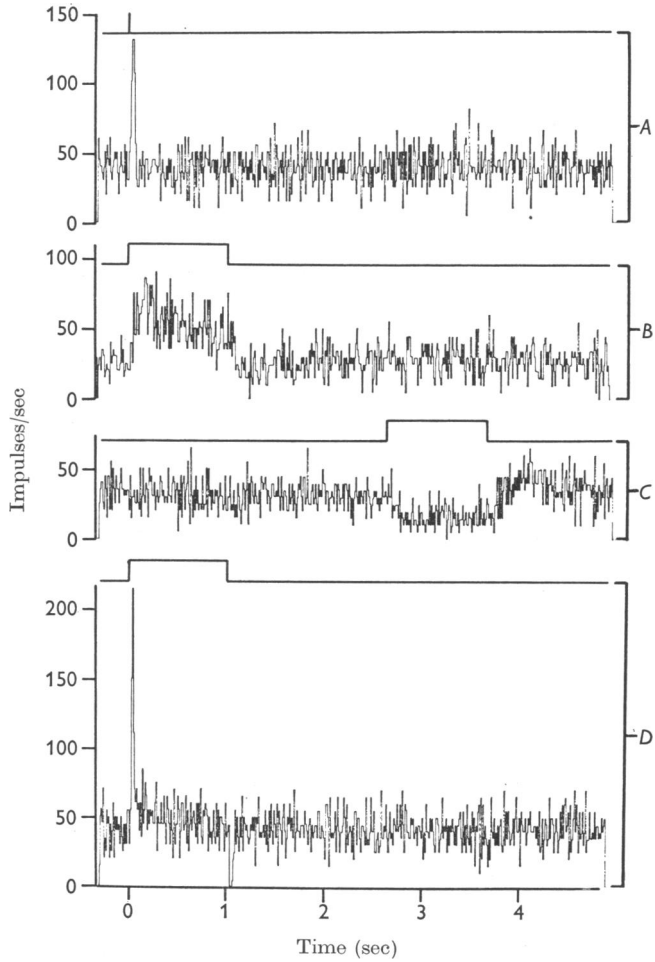

Fig. 1. Post-stimulus time-histograms for calculating quantum/spike ratios. For each pair the upper line shows the stimulus, the lower line the number of impulses in successive channels of duration 10 msec; the stimulus was repeated twenty times, and the counts have been multiplied by five to convert to impulses/sec. The number of impulses is obtained from the area of the whole response, or a selected part of it, excluding impulses attributable to maintained discharge. The number of quanta is calculated for the whole of the stimulus, or for a duration equal to the selected part of the response. *A* is a brief response to a brief, centrally located, 10 msec stimulus in an on-centre unit. *B* is a sustained response to a 1 sec stimulus at the centre of an on-centre unit. *C* is the response to a large, 1 sec, stimulus in a dark-adapted off-centre unit. *D* is the response to a large, 1 sec, stimulus in a light-adapted on-centre unit.

In both the last two cases one could obtain nearly the same result by considering only a small portion of the response, say 100 msec, and making the assumption that the extra impulses arose from quanta absorbed during an equal portion of the stimulus. However, responses are not always sustained as in these cases, and Fig. 1*D* shows a response to a 1 sec stimulus in which a large proportion of the extra impulses arose within the first 100 msec. Obviously one will get a much lower quantum/spike ratio (equivalent to less attenuation or greater 'gain') if one counts extra impulses during the peak 30 msec, assuming them to result from 30 msec of stimulation, than if one analyses the whole response. In this case the quantum/spike ratio for the peak 30 msec is $1{\cdot}4 \times 10^6$ and for the first 100 msec, $3{\cdot}1 \times 10^6$. The quantum/spike ratio for a 10 msec just suprathreshold flash (as in Fig. 1*A*) was $0{\cdot}91 \times 10^6$ for this stimulus area and adaptation level on this unit, whereas for the whole of the 1 sec stimulus it was 15×10^6.

To begin with, we define the quantum/spike ratio as the ratio of total number of additional quanta absorbed to the total number of extra spikes. However, when the response is of shorter duration than the stimulus, one can obtain an indication of the lower quantum/spike ratio obtainable with brief stimuli by counting extra impulses during a selected brief analysis period, and counting only the extra number of quanta absorbed during an equal period. It will be shown later (see Fig. 6) that the weakest incremental stimuli can be detected when one chooses an analysis period τ that includes most, but not necessarily all, of the extra impulses elicited by the stimulus. Thus the general definition of the quantum/spike ratio is the ratio of extra quantal absorptions within a period τ to the extra impulses within τ.

The usefulness of the quantum/spike ratio in characterizing the sensitivity of the retina is greatly increased by the fact that the number of additional impulses is directly proportional to the added luminance of the stimulus, provided that this is weak. This is shown in Fig. 2. The extent of the linear range varied with stimulus conditions, and is small for large stimuli that cover both centre and surround. However, for small stimuli covering the centre alone it extends to intensities several times above threshold. We shall show elsewhere that it also extends to decremental stimuli that decrease the number of impulses discharged.

Pulse number distributions. The quantum/spike ratio, and its constancy for weak stimuli, enables one to predict the average number of impulses that will result from stimuli of a known intensity, duration and area. The form of the response tells one the time interval during which these extra impulses occur.

The next problem is, 'How many impulses are required for threshold?'

If one defines threshold as the weakest stimulus that can be detected reliably, then one needs two more pieces of information in order to calculate it. The first is the distribution of the numbers of impulses per analysis period that result from the maintained discharge by itself, without stimuation. The second is the conventional criterion that defines the standard

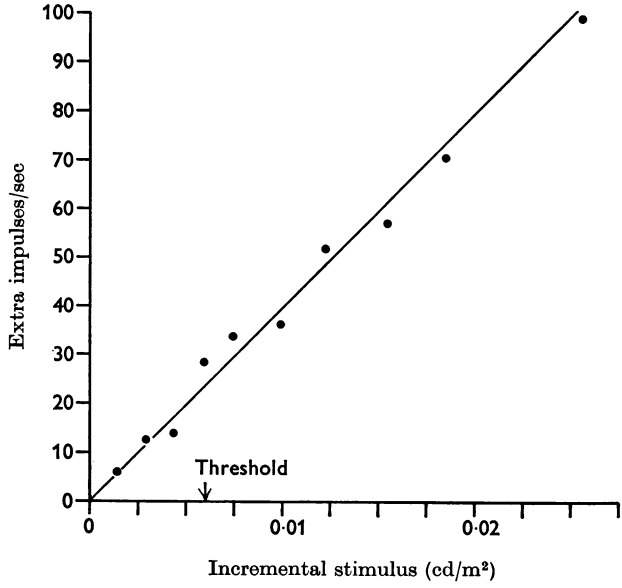

Fig. 2. Linearity for weak stimuli. Over a range from below threshold to 4 × threshold the response to central stimulation does not depart significantly from a linear function of stimulus intensity; hence the quantum/spike ratio is constant for these conditions of stimulation. The determination of threshold is illustrated in Fig. 5. Unit QQ: 4. On-centre. Background: 0·05 cd/m². Stimulus: 0·5°, 640 msec.

of reliability required, or the number of false positive responses allowed. The first point requires additional experimental evidence of the type shown in Fig. 3.

These pulse number distributions for the steady maintained discharge show the probability of occurrence, within the specified time interval, of the number of pulses plotted horizontally. Thus in Fig. 3*C* the average number within the 100 msec interval was 4·75, but on successive trials the number actually obtained varied between 1 and 9. Superimposed on each discrete distribution is a smooth curve of the Gaussian probability density function

$$P(N) = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left(\frac{-(N-\mu)^2}{2\sigma^2}\right)$$

adjusted in each case to have the same mean, μ, and standard deviation, σ, as the experimentally obtained distributions. The similarity in shape is remarkably close in all instances. Errors arise when the blocks of the discrete histogram are coarse relative to the spread of the continuous distribution, and the continuous distribution will be non-zero for unrealistic

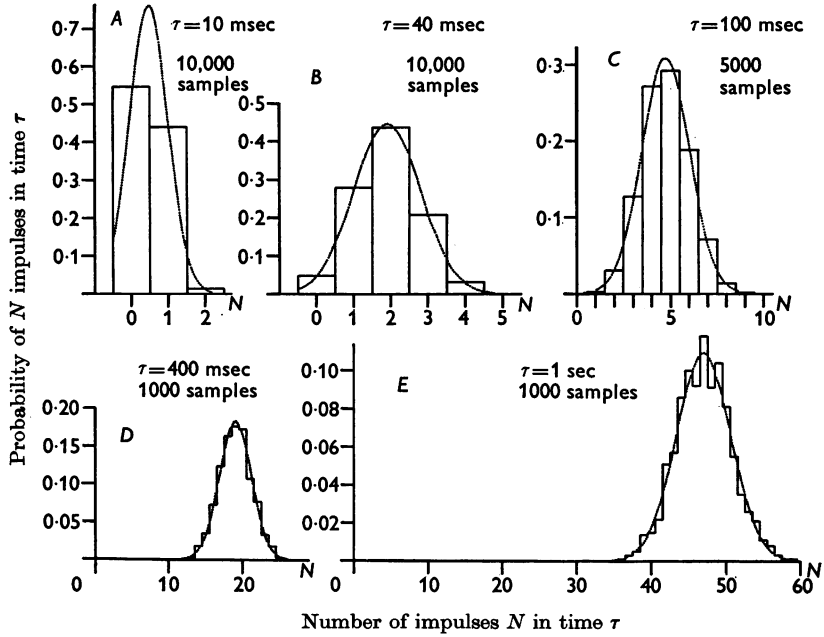

Fig. 3. Pulse number distributions for the maintained discharge. The histograms show the frequency with which a particular number of impulses occurred within successive samples of the analysis period τ. The dotted curves are Gaussian probability density functions with the same mean and standard deviation as the histograms. The responses almost always required an analysis time of 40 msec or longer, and for these values of τ the Gaussian is a satisfactory approximation to the histogram.

negative values of N. The histogram for τ = 10 msec is included to illustrate these limits to the present empirical approach, but the distributions for τ ranging from 40 msec to 1 sec are the ones of practical interest, and for these the errors introduced are negligible.

The mean and standard deviations of the experimental data are used to fit a Gaussian function to the distribution, and this enables one to calculate N_c, the criterion number of impulses that must be exceeded to decide reliably on the presence of a stimulus. The areas under the tails of the experimental distribution would be difficult to measure directly since the events with which they correspond are rare and thus would require very

large samples for accurate assessment. Means and standard deviations, on the other hand, are much more rapidly determined.

Inspection of Fig. 3 suggests that the spread of the distributions relative to their respective means becomes smaller as the mean increases. In Fig. 4 the variance has been plotted against the mean for a set of distributions, including those in Fig. 3, obtained by an analysis of a particular segment of maintained discharge at different sampling times. For any adapting

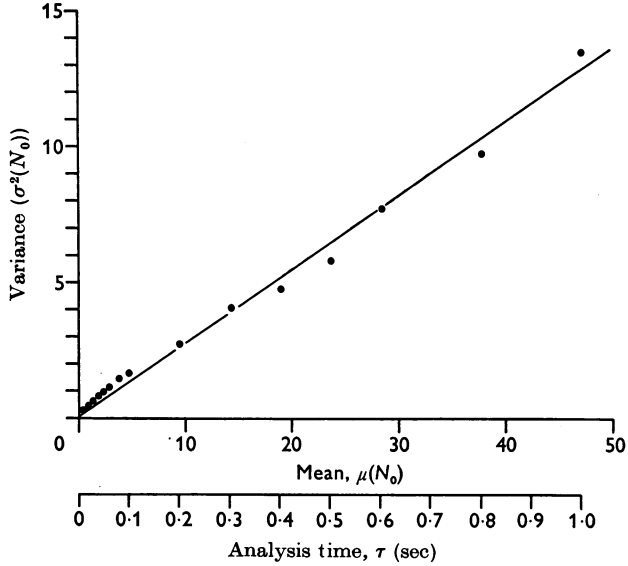

Fig. 4. Variance of pulse number distribution for different analysis times. A straight line through the origin is a reasonable approximation, but there are significant deviations when the mean number per analysis time is small. Unit Y: 3. Background 40 cd/m^2.

luminance the standard deviation is nearly proportional to the square root of the mean number occurring within the sampling periods. However, the constant of proportionality varies for different units, and for the same unit at different adaptation levels. These properties of the maintained discharge will be considered further elsewhere (H. B. Barlow & W. R. Levick, in preparation); for present purposes, the linear relation shown in Fig. 4 enables one to calculate the parameters of the pulse number distribution for any analysis period from the mean and variance determined at one period. A slightly more accurate approximation could be obtained from determinations at two periods, for the straight line passes close to the origin, but not quite through it.

Calculation of threshold. The method of calculating threshold is illustrated in Fig. 5. A post-stimulus time-histogram is obtained for a stimulus

intensity within the range of expected threshold values (see also control 1 below). This histogram is shown at the top of the figure, and the form of the response enables one to select a time interval of duration τ, beginning shortly after the stimulus is delivered, within which the majority of extra

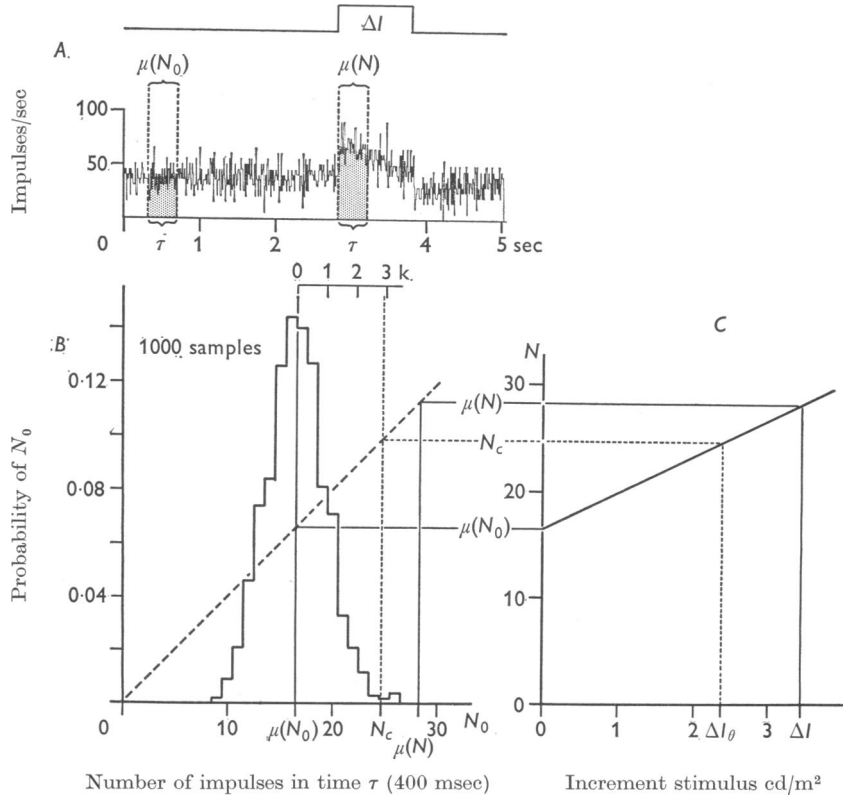

Fig. 5. Calculation of threshold. *A* shows a post-stimulus time-histogram for a weak stimulus. An analysis period of 400 msec was selected because this included the main part of the response. *B* shows the pulse-number distribution for this value of τ, compiled from a recording of the maintained discharge. The mean, $\mu(N_0)$, was 16·5, and the standard deviation, $\sigma(N_0)$, was 2·83. Hence $N_c = \mu(N_0) + 2{\cdot}88\,\sigma(N_0) = 24{\cdot}6$. In *C* the values of $\mu(N_0)$ and N_c are plotted vertically, and incremental stimulus intensity horizontally. The actual stimulus, ΔI, caused an average, $\mu(N)$, of 28·3 impulses; the threshold stimulus, ΔI_θ, required to cause N_c on average is obtained by interpolation. Unit BB: 4. On-centre. Background 22 cd/m²; 0·6° centred spot; 20 trials summed.

impulses occur (see also control 2 below). Knowing the interval of importance for this particular experiment one can obtain a pulse number distribution for this analysis interval, τ, which is illustrated at bottom left of the figure (see also control 3 below).

The mean of this distribution is $\mu(N_0)$, and its standard deviation $\sigma(N_0)$. From these one finds the criterion number of impulses, N_c, which is exceeded with the desired low probability. We have employed $N_c = \mu(N_0) + 2{\cdot}88\,\sigma(N_0)$, which gives a false positive response rate of 2/1000 if the distribution is exactly Gaussian. Now one interpolates or extrapolates to obtain ΔI_θ, as illustrated at bottom right.

We want to emphasize that averaging techniques are used to determine the optimal analysis period, and to estimate the quantum/spike ratio, but we calculate the threshold for a single presentation of the stimulus. If the cat analysed its optic nerve impulses efficiently, it could detect a stimulus of intensity ΔI_θ on 50 % of trials without giving more than the required low percentage of false positive responses; we avoid assuming that the cat averages many responses, or the responses of many fibres, and our own use of averaging improves our accuracy, but does not spuriously lower our estimate of the cat's threshold.

Definitions and controls

In this part we give a more formal statement of the theory. This is followed by experimental validation of some of the assumptions.

τ = the duration of the analysis period; this is sufficient to include the major part of the response.

I = steady luminance to which the eye is adapted.

M = average maintained discharge rate in impulses/sec at I.

N_0, $\mu(N_0)$, $\sigma(N_0)$ = number of impulses occurring within an analysis period in the absence of the stimulus, and the mean and standard deviation of this number: $\mu(N_0) = M\tau$.

C = experimentally determined proportionality factor relating $\sigma(N_0)$ and $\mu(N_0)$ at a particular adapting luminance: $\sigma^2(N_0) = C\mu(N_0) = CM\tau$.

N, $\mu(N)$, ΔN = number of impulses within the analysis period when a single stimulus of luminance ΔI is delivered, the mean of this number, and this number expressed as an increment on the maintained discharge: $N = \Delta N + \mu(N_0)$.

N_c, ΔN_c = criterion number of impulses within τ that must be exceeded for the decision that the stimulus is present, and this number expressed as an increment on the maintained discharge: $N_c = \Delta N_c + \mu(N_0)$.

k = signal/noise ratio determining the reliability of the threshold decision and the proportion of false positive responses: $N_c = \mu(N_0) + k\sigma(N_0)$.

ΔI, A, T = the luminance, area in degree^2, and duration in sec, respectively, of the stimulus. The luminance is in units such that $\Delta I . A . T$ = number of quanta of 507 nm striking the cornea and passing the pupil that would have the same scotopic effectiveness as the stimulus.

F = fraction of quanta striking the cornea and passing the pupil that are absorbed in rods: $\Delta I.A.T.F$ = number of quanta absorbed from the stimulus.

s = quantum/spike ratio; the average number of extra quanta absorbed in a time, τ, divided by the number of extra impulses elicited in an equal, appropriately placed, period:
$s = \Delta I.A.\tau.F/\Delta N$, for $T \geqslant \tau$,
$s = \Delta I.A.T.F/\Delta N$, for $T \leqslant \tau$.

ΔI_θ = threshold luminance, for which
$\mu(N) = N_c = \mu(N_0) + k\sigma(N_0)$.

The problem of calculating the threshold luminance, ΔI_θ, is broken down into three stages. First, the form of the average response is used to determine the time interval, τ, within which the maintained discharge will be significantly modified if the stimulus is suprathreshold; secondly, the statistical parameters of the maintained discharge are used to calculate the critical number N_c that must occur within τ in order to make a reliable decision that a stimulus was present; thirdly, the value of s, the quantum/spike ratio, is used to calculate the stimulus luminance required for the average response to equal N_c.

There are problems in each of these three stages that are discussed in the next section on controls. τ is determined by inspection of a post-stimulus time-histogram of a weak response, and the standard deviation of the pulse number distribution for that value of τ is obtained from the relation, shown in Fig. 4,

$$\sigma(N_0) = \sqrt{(CM\tau)} = \sqrt{(C\mu(N_0))}, \tag{1}$$

where C is determined by analysis of a segment of maintained discharge at one particular analysis period. The critical number is then given by

$$N_c = \mu(N_0) + k\sigma(N_0),$$

where k is taken to be 2·88, equivalent to 0·2 % false positive detections if the pulse number distributions are exactly Gaussian. The number of additional impulses required for threshold, $\Delta N_c = N_c - \mu(N_0)$, is now fixed, and from a particular post-stimulus time-histogram for a value of ΔI near threshold one determines the quantum/spike ratio; then

$$\Delta I_\theta.A.\tau.F = s\Delta N_c = sk\sigma(N_0) = sk\sqrt{(CM\tau)}. \tag{2}$$

For particular stimulus and adaptation conditions, all the quantities on the right-hand side can be found, following choice of τ, and thus ΔI_θ can be calculated. The equation expresses the dependence of threshold upon quantum/spike ratio, the reliability required, the irregularity of the maintained discharge, and the analysis period, which is matched to the

time interval within which the response occurs. However, it is important to realize that the form of the response changes with stimulus parameters and adaptation level; consequently τ must also be changed to obtain the lowest threshold. These changes must be known before equation (2) can be used to predict how threshold will depend upon the conditions of the stimulation and adaptation.

Certain assumptions and apparently arbitrary procedures are used in the calculation of threshold described above, and controls will now be described to check their validity. Other assumptions are considered in the discussion.

(1) *Choice of analysis time.* The analysis time is chosen from inspection of the post-stimulus time-histogram, and this may seem arbitrary. However, we have investigated how this choice influences the calculated value of threshold with the result shown in Fig. 6. It will be seen that if a reasonable choice is made, the exact value chosen for the analysis time has little influence, and this is especially true if the post-stimulus time-histogram is recorded from a near-threshold stimulus.

Ideally, in detecting a weak response, one would not take a definite analysis period, but rather one would inspect the record by continuously calculating the integrated cross product with a time-weighting function that exactly matches the expected distribution of excess impulses. This improved procedure may be more closely approximated by the analogue method mentioned in the discussion (see p. 22).

(2) *Pulse number distributions when mean rate changes.* The pulse number distributions are normally obtained from segments of the maintained discharge recorded before or after the period of stimulation. However, we have found that even weak stimuli at infrequent intervals may displace the mean rate of the unstimulated discharge, which also may drift unpredictably. Thus the parameters deduced from results such as those shown in Fig. 4 may not apply at the time when one is attempting to measure threshold.

The part of the post-stimulus time-histogram preceding the stimulus provides a check on the maintained discharge. If the mean rate was found to be very different from that of the recorded maintained discharge, the results were discarded. If the change was small, equation (1) was used to estimate $\sigma(N_0)$, even though this relation was validated for variation of τ, not for variation of M. The part of the post-stimulus time-histogram preceding the response also gives some information on $\sigma(N_0)$, and it can be estimated at the appropriate value of τ by combining groups of neighbouring channels until the effective bin width is equal to τ. By this method we have statistically tested the hypothesis that C in equation (1) does not change for variations in M of the type we are discussing. Though these

tests are not strong, they have not negated the null hypothesis, and we feel that errors introduced from possible variations in C cannot be important in this paper.

(3) *Choice of intensity for post-stimulus time-histograms.* If the response increment were exactly proportional to incremental stimulus intensity

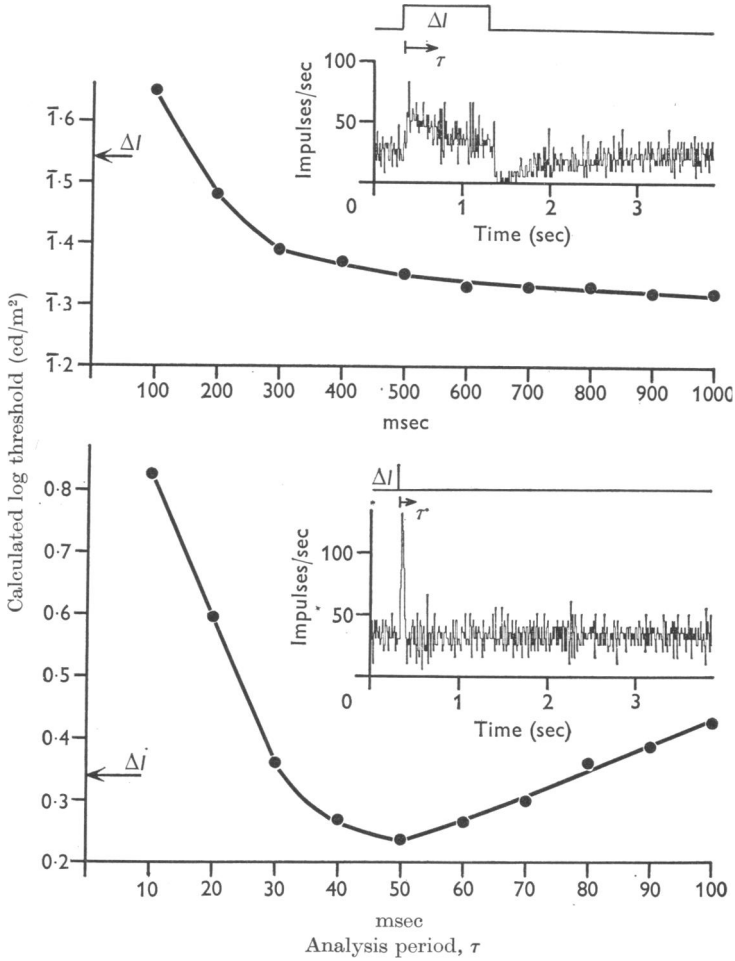

Fig. 6. Effect on the calculated threshold of varying the analysis time. In the top section a 1 sec stimulus of the intensity shown on the vertical scale produced a fairly well sustained response. The values of log threshold are plotted vertically when calculated for the analysis time shown horizontally. In this case prolonging the analysis time beyond 400 msec to include all the extra impulses elicited by the stimulus only lowers the calculated threshold by 0·05 log. units. In the lower section a brief stimulus was used, and the optimal analysis time was 50 msec; values between 35 and 70 msec gave thresholds differing by less than 0·05 log unit, but a large mismatch would elevate the threshold very much more. Unit AA: 1. Background 3 cd/m².

(quantum/spike ratio exactly constant) the intensity of stimulus chosen for the post-stimulus time-histogram would be immaterial, except that a bigger response would reduce sampling errors in estimating the excess impulses. However, for large-area stimuli the quantum/spike ratio is not independent of stimulus strength and the choice of intensity for the post-stimulus time-histogram is indeed critical; this must be chosen as near as possible to the true threshold so that the range of extrapolation or interpolation is minimal. For small stimuli, linearity is sufficiently good for such errors to be unimportant.

Conditions where each factor causes threshold changes

Our main purpose in developing an objective definition of the 'threshold' of a single ganglion cell was to find what change of single unit behaviour causes a change in the psychophysical threshold. The examples we shall present here lead us to believe that changes in time course of the response, statistical parameters of the maintained discharge, and quantum/spike ratio may each exert an influence on over-all performance.

Changes of quantum/spike ratio and increment threshold. In Fig. 7 log

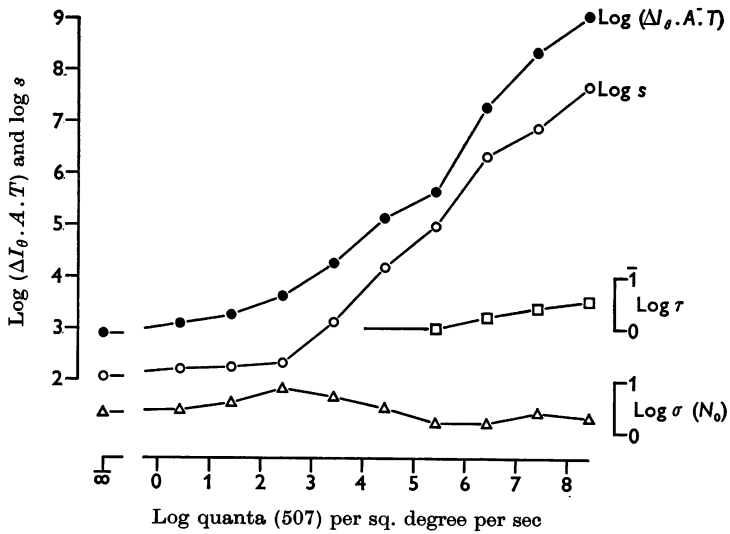

Fig. 7. Quantum/spike ratio and incremental threshold. The calculated values of incremental threshold (●) are plotted as a function of background intensity on logarithmic coordinates. Quantum/spike ratio (○), analysis period τ (□), and standard deviation of pulse number distribution $\sigma(N_0)$ (△) are plotted on coordinates chosen so that changes in log. threshold are the sum of the changes in these underlying factors. Clearly quantum/spike ratio is the dominant factor, and analysis period and irregularity of the maintained discharge contribute in some degree to details of the shape of the incremental threshold curve. Unit EE: 1. On-centre.

(threshold) for a large, long-duration stimulus is plotted against log. (adapting luminance), and these objectively measured thresholds obviously rise in a similar way to the subjectively determined psychophysical threshold of humans. The values of the three factors that determine the objective threshold are also plotted, and the scales for log s, log τ, and log $\sigma(N_0)$ have been chosen so that changes in log $(\Delta I_\theta . A . T)$ are the sum of changes in these three factors; thus it is easy to see that two factors make only minor contributions to the shape of the increment threshold curve. The rise of log $\sigma(N_0)$ accounts for the gentle rise of threshold at very low backgrounds, and the shortening of the response accounts for the steeper rise at very high backgrounds. Changes of quantum/spike ratio are obviously vastly the most important.

Changes in temporal integration and the duration of the response. Psychophysical evidence shows that the upper limit of complete temporal summation (Bloch's Law: $\Delta I . T$ = constant) is reduced by raising the adaptation level, and also by enlarging the test stimulus that is to be detected. These changes can be qualitatively explained by changes in the duration of the response.

Levick & Zacks (1968) have shown that when a constant energy stimulus is prolonged in duration and reduced in intensity, the response stays the same up to a certain critical duration. Because the form of the response is unchanged the optimal analysis period stays the same, and so does ΔN_c, the critical excess number of impulses, and s, the quantum/spike ratio. For short duration stimuli equation (2) is

$$\Delta I_\theta \, ATF = sk\sqrt{(CM\tau_{\min})}. \qquad (2a)$$

If A and F are constant, it follows that $\Delta I_\theta . T$ will be constant, which is Bloch's Law of complete temporal summation. As the stimulus is further prolonged the response gets broader, and thus the analysis period required to count all the extra impulses will start to increase above its lower limiting value, $\tau_{\min}$. The number of extra impulses required for detection will rise because τ and $\sigma(N_0)$ increase; hence $\Delta I_\theta . T$ will rise even if s stays unchanged. At still longer durations the response may terminate before the stimulus is over, and the number of additional impulses becomes independent of stimulus duration. Under these conditions ΔI_θ will reach a constant lower limiting value. Thus psychophysical performance with regard to temporal integration becomes understandable, though we have not yet made it quantitatively predictable.

As the area of a stimulus of constant energy is increased there is little change in the response until the spot encroaches upon the inhibiting surround. At this point the response is reduced, the quantum/spike ratio increases, and consequently $\Delta I_\theta . A$ must start to increase. Since the

effectiveness of the surround increases with the adaptation level, areal summation decreases, and we go into this elsewhere (H. B. Barlow & W. R. Levick, in preparation). Here we want to point out the probable reason why the amount of temporal summation is reduced by increasing either the stimulus area, or the adaptation level, in human vision (Barlow, 1958). The inhibitory effects from the surround seem to be more effective after a slight delay, and as a result there are complex interactions some of which are shown in Fig. 8. The over-all result is that the duration of a response to a long

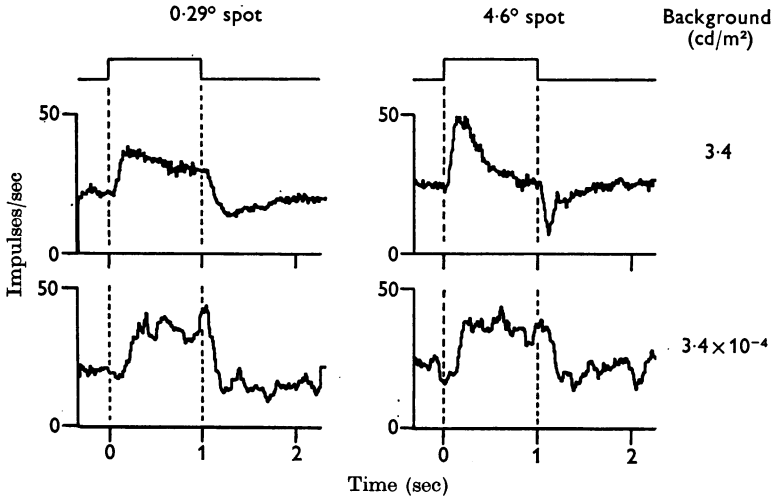

Fig. 8. 1. Effect of stimulus area on response duration at two adaptation levels. Unit EE: The response to a small stimulus spot is sustained throughout a 1 sec stimulus at both adaptation levels, but the response to a large area stimulus adapts rapidly at the upper adaptation levels. This is thought to underlie the interaction of stimulus area and adapting luminance with psychophysically measured summation time. This figure was made from 10 msec/channel post-stimulus time-histograms like those of Fig. 1 by averaging the contents of eight neighbouring channels. This is equivalent to increasing the channel width to 80 msec, and it greatly reduces the irregularity of the record. See text, p. 23.

duration stimulus is shortened either by increasing its area or by increasing the adaptation level. The cutting short of the response means that the lower limiting value of ΔI_θ is reached at smaller values of τ and thus the integration time is reduced. These effects are thought to provide an example where the form of the response is important in determining threshold changes.

Changes in maintained activity causing threshold change. Finally, we shall give an example where a difference of threshold is caused by a change in the noisy background of maintained activity. The top part of Fig. 9 shows two post-stimulus time-histograms for twenty repetitions of a 10 msec

stimulus. The channel width was 50 msec, which is close to the optimal analysis time for a brief flash under these conditions. The stimulus time was adjusted so that very nearly all the additional impulses caused by a stimulus occupied the same channel; thus the single channel indicated by

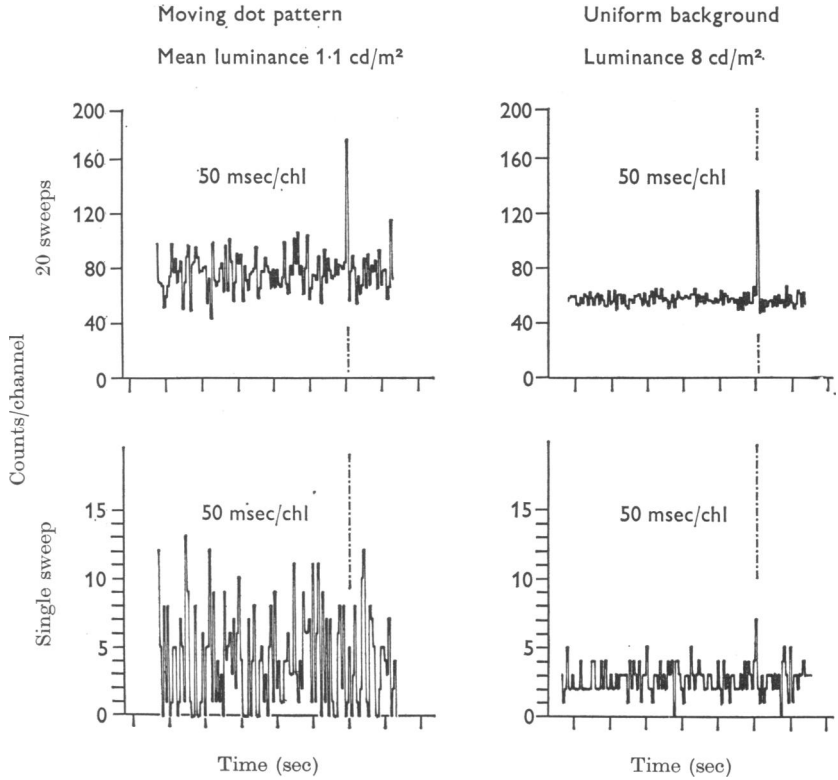

Fig. 9. Irregularity of the maintained discharge and threshold. RR: 2. On-centre. The irregularity of the background activity was raised in the left records by superimposing the stimulus upon an irregularly moved pattern of dots; as a result the stimulus could not be reliably detected in a single trial (lower record), though it was easily seen when twenty sweeps were accumulated (upper record). The irregular excitation raised the quantum/spike ratio, and a uniform background of seven times the mean luminance of the moving dot pattern produced a slightly higher ratio, as shown in the upper right record. The lower record shows that this response, though smaller than the one on the left, was readily detectable on a single trial. The moving dot pattern elevated the threshold about twenty times compared with a uniform field of equal mean luminance. The output noise was raised about threefold, the quantum/spike ratio about sevenfold. For all these records the channel width was 50 msec, which was the optimum analysis period for the responses evoked by the brief stimulus. Because channel (chl) width was equal to the analysis period, the irregularity of the base line preceding the stimulus gives a valid impression of the standard deviation of the appropriate pulse number distribution, which is the noise against which the signal must be discriminated.

the dotted line is the 'signal', and the preceding 100 channels give an indication of the 'noise'. Comparing the two halves, it will be seen that the amplitude of the signal is nearly the same, and the stimulus was in fact unchanged. The noise, however, is much greater for the left histogram. This increased noise was caused by moving a projected pattern of dots over the translucent surface of the fluorescent stimulator. Below the post-stimulus time-histograms are shown single responses to the same stimulus in the two conditions; it is clear that the response is lost in the noise in the left-hand sample, but can be picked out in the right-hand sample, and repetitions showed that this was nearly always the case. Incidentally, the base line in this figure gives a valid indication of the background noise only because the channel width was chosen to be equal to the optimal analysis time. The same is true for accumulated histograms, but when channel width is not equal to the analysis time, the irregularity of the base line is misleading (see also p. 23).

From the two records one can calculate thresholds, by the method described. The standard deviation of the pulse number distribution is 3·3 times as great in the noisy record, and the critical increment required, ΔN_c, is raised by the same factor. Since the quantum/spike ratio is a little lower for the noisy record, the threshold was raised by a factor of 2·5.

Figure 9 thus demonstrates a difference in threshold that is attributable to a difference in the maintained discharge; the quantum/spike ratio by itself would have caused a small change in the opposite direction. However, it would be misleading to suggest that a noisy visual background such as was used in this experiment causes threshold elevation only by increasing the noise in the retinal output. The uniform background was adjusted in luminance to produce nearly the same quantum/spike ratio as with the moving dot pattern, and we found that it had to be about seven times the mean luminance of the dot pattern to do this. In this luminance range Weber's Law held approximately for this unit. The moving dot pattern therefore elevated the threshold by a total factor of about twenty compared with a uniform background of the same mean luminance, and this twentyfold change resulted from about sevenfold increase in quantum/spike ratio and threefold increase in output noise. This result rather suggests that the irregularity of the output is important in controlling the quantum/spike ratio, or 'gain', of the retina.

DISCUSSION

This paper has been concerned to a large extent with the methodological problem of how one should measure the sensitivity of a single ganglion cell. The strengths and weaknesses of our approach need first to be stated in

order that the significance of our results can be assessed. Following FitzHugh (1957, 1958) what we have done is to examine the responses to various strengths of stimulation with a view to specifying the intensity at which a single neurone begins to signal reliably that a stimulus was delivered. Few would deny that the visual responses of an animal are mediated by messages along its optic nerve fibres; thus our attention to the informational aspects of these messages makes it unnecessary to postulate a 'psychophysical linking hypothesis' (Brindley, 1960), and obviates the philosophical argument that would be likely to follow such an intrinsically untestable assumption. On the other hand there are some weaknesses: first, we have treated the case where the signal to be detected is known exactly, whereas the cat is presumably interested in responses of any duration, at any time, in any of its many nerve fibres. We made this assumption that the expected signal is known exactly because it is the simplest case; in the more realistic situation where there are many possible positions, durations, and times of occurrence of a response, it is easy to see that one would have to demand a much lower rate of false positive responses in each individual channel, and the value of k would have to be increased. However, such improved reliability does not require a very big increase in stimulus strength; for instance, a 50% increase of threshold criterion would raise k from 2·88 to 4·32, and the false response rate would improve from 1 in 500 to about 1 in 100,000.

Another possible objection is that we have calculated results for optimal analysis of the discharge of a single neurone, whereas the actual analysis performed by the higher centres is likely to fall short of this. Also we have neglected the large amount of information potentially available in the detailed timing of impulses in the discharge of a single neurone, and in the timing of impulses in one neurone relative to impulses in another neurone. We do not know if these factors are important; the extent to which single unit thresholds account for behavioural results can be regarded as a test of the hypothesis that they are not. However, one must also remember that, in so far as we can make this comparison at all, we are only matching results on the human scotopic system with single unit analyses of the cat's scotopic system. Thus our results are necessarily preliminary, but they show where to look for the physiological factors underlying visual behavioural capacities, and also how these factors change adaptively with the visual input.

Before discussing methodological points some actual figures from the experiment of Fig. 7 will be briefly quoted in order to bring out the range of values of the quantities involved in calculating threshold.

Values of quantum/spike ratios and critical increments. The values of the quantum/spike ratio for a large-area, long-duration stimulus are illustrated

in Fig. 7, and it will be seen that they range from 100 up to 5×10^7. These figures are high because the large stimulus is badly matched to the receptive field: when tested with a small, brief, stimulus confined to the centre, the quantum/spike ratio drops to the neighbourhood of unity at zero background (H. B. Barlow & W. R. Levick, in preparation). The retina becomes capable of counting single quanta, and this must obviously be taken into account when considering theories of the absolute threshold.

If one absorbed quantum can cause an extra impulse it is natural to ask the question how many extra impulses are required for reliable detection. The value of this critical increment, ΔN_c, depends upon many factors, but a range from 1 to 20 gives a correct idea of its order of magnitude and its variation when adaptation conditions are held steady at different levels. This is, of course, rarely the situation in the normal, non-laboratory, use of the eye, nor would it be true in a psychophysical experiment if one was looking at a non-homogeneous field with unstabilized vision. The experiment with moving dot background suggests that the variability of the retinal output under such conditions might necessitate a threefold increase in ΔN_c.

The fact that several extra impulses are almost always needed to raise an optic nerve fibre's signal out of the noise has an important consequence. Previously (Barlow, 1965) it was suggested that the quantum/spike ratio limited sensitivity by adding to variance, in the same way that the grouping of scale readings to the nearest graduation can increase measurement errors and limit their accuracy. In its crude form this is wrong, because the maintained discharge never becomes so regular that one can reliably detect a single interpolated impulse, and other sources of variance are more important than that introduced by 'neural quantization'. This result makes it important to ask what causes the irregularity of the maintained discharge at high adaptation levels, for it is this irregularity, rather than the quantization error, that limits sensitivity. We have more information on this (H. B. Barlow & W. R. Levick, in preparation) and it appears that the statistical fluctuations in quanta absorbed in the centre are a major contributor to the irregularity of the maintained discharge. At all events the present analysis shows that the quantum/spike ratio is the main factor that adjusts the increment threshold when background luminance is changed; the mechanisms that control it are the main mechanisms of visual adaptation, and are responsible for giving the eye its dramatic dynamic range.

Comparison with other threshold methods

The method described here determines objectively the lower limit of detectability of a light stimulus, under certain assumptions about what features of the neural discharge carry information as to intensity. It is

interesting to compare four other methods. We have used three of them, and although we have made no detailed numerical comparisons we know the values given by all these three usually lie within $\frac{1}{2}$ log unit of each other. We do not think the fourth method is a valid way of estimating sensitivity, and we shall describe its shortcomings.

Subjective estimates. The main objection is simply its subjectivity, and the difficulty of specifying what change occurs at the threshold. However, this method has been used extensively, and the fact that our objective method gives nearly the same result obviously validates this past usage.

Simplified counting. We used the gates described in Methods to start and stop a counter, and thus find the number of pulses occurring within the selected analysis interval for each repetition of a stimulus. We had also a continuous record of mean rate and adjusted the intensity of stimulus until the number of pulses in the analysis interval averaged $\mu(N_0)+2\sqrt{\mu(N_0)}$. Though this setting was usually close to the correct threshold, it consistently turned out to be lower than threshold at low adaptation levels in on-centre units, and higher at high levels. This is because the dispersion of the pulse number distribution varies considerably between these conditions, and it shows that taking the irregularity of the maintained discharge into account influences the estimate of threshold.

Mean rate trace. We have some preliminary experience with an analogue technique of performing the operation illustrated in Fig. 5. Each impulse triggers a pulse of a wave form whose duration is selected to match the analysis period. These pulses add together, and the summed value at any instant indicates how many impulses occurred within the preceding analysis period. This output is displayed by a pen recorder, and deflections out of the noisy base line occur where the stimulus is suprathreshold. The noise level of the base line can be held constant with parametric feedback, and a range of pen recorders can give the results for a range of analysis periods. It is our current impression that this has all the advantages of the full digital method except that it does not give a record of the post-stimulus time-histogram.

Inspection of post-stimulus time-histogram. Heiss & Milne (1967) recently estimated the sensitivity of single fibre preparations from the cat's eye; they accumulated post-stimulus time-histograms for sixty to a hundred repetitions of a stimulus, and called 'threshold' the lowest intensity for which there was a visible amount of stimulus-locked activity. There are two serious problems with this method. First, as in all averaging methods, the number of responses accumulated makes a major difference to the effectiveness with which the signal is separated from the noise. If one attempts to correct for this by multiplying their stated result by the square root of the number of responses accumulated one obtains figures in quite

good agreement with previous estimates on single units, but the significance of this is not clear because the pupil area in their preparation is not given.

The second problem with their method is more interesting. The limit to detecting stimulus-locked activity is partly dependent on the 'noise' or irregularity in the portion of the histogram preceding and following the stimulus, but this does not give a correct estimate of the noise. This becomes obvious when one realizes that an exactly regular maintained discharge can, in appropriate conditions, give rise to a very irregular base line. To make the irregularity of the base line meaningful one should use a channel width equal to the analysis time, as in Fig. 9, or combine adjacent channels, as in Fig. 8.

Our experience with various methods of determining threshold can be summarized by saying that it is not easy to improve much upon subjective judgements based directly on the ganglion cell output. Post-stimulus time-histograms can be misleading. The simplified objective methods do provide some record of what happened, but only the full method allows one to analyse what lies behind a change of threshold.

Retinal information processing

Our results show that the weakest stimulus that a single ganglion cell can reliably signal depends upon the duration of the response and the irregularity of the maintained discharge, as well as the quantum/spike ratio. Special circumstances can be found where the first two factors probably must be taken into account in order to understand psychophysical results; the third factor, the neural 'gain' or 'attenuation', changes a great deal and its primary importance confirms what has been suspected for thirty years or more.

The discovery of the two additional factors affecting threshold is only a small reward for the labour of applying signal detection theory to this problem, and the fact that our objectively calculated thresholds are not significantly lower than those determined by the simplest subjective methods reinforces these doubts about the value of the approach. However, these rather negative results have actually thrown a very interesting light on the way the retina is processing information for utilization by the higher centres.

Detecting the minimum significant increment (or decrement) in the number of quantal absorptions requires the determination of a different criterion at every adaptation level, and this is a fairly sophisticated statistical task. But what we have found is that a relatively simple, unchanging, operation at the output will achieve fairly efficient detection of these increments. What the retina has done, then, is to 'normalize' the

changing input and present it in standardized form at the output. This must be what the neural mechanisms of adaptation are all about.

This work was supported by a grant from the United States Public Health Services (NB-05215) and was performed while one of us (HBB) held a Miller Professorship from the Miller Institute for Basic Research in Science.

REFERENCES

ARDUINI, A. & PINNEO, L. R. (1962). Properties of the retina in response to steady illumination. *Archs ital. Biol.* **100**, 425–448.

BARLOW, H. B. (1956). Retinal noise and absolute threshold. *J. opt. Soc. Am.* **46**, 634–639.

BARLOW, H. B. (1957). Increment thresholds at low intensities considered as signal/noise discriminations. *J. Physiol.* **136**, 469–488.

BARLOW, H. B. (1958). Temporal and spatial summation in human vision at different background intensities. *J. Physiol.* **141**, 337–350.

BARLOW, H. B. (1965). Optic nerve impulses and Weber's law. *Cold Spring Harb. Symp. quant. Biol.* **30**, 539–546.

BORNSCHEIN, H. (1958). Spontan- und Belichtungsaktivität in Einzelfasern der N. opticus der Katze. I Der Einfluss kurzdauernder retinaler Ischämie. *Z. Biol.* **110**, 210–222.

BRINDLEY, G. S. (1960). *Physiology of the Retina and Visual Pathway*. London: Arnold.

CRAIK, K. J. W. (1938). The effect of adaptation on differential brightness discrimination. *J. Physiol.* **92**, 406–421.

DOWLING, J. E. (1967). The site of visual adaptation. *Science, N.Y.* **155**, 273–279.

FITZHUGH, R. (1957). The statistical detection of threshold signals in the retina. *J. gen. Physiol.* **40**, 925–948.

FITZHUGH, R. (1958). A statistical analyzer for optic nerve messages. *J. gen. Physiol.* **41**, 675–692.

FUORTES, M. G. F. & HODGKIN, A. L. (1964). Changes in time scale and sensitivity in the ommatidia of *Limulus*. *J. Physiol.* **172**, 239–263.

GERBRANDS, R. & STEVENS, J. C. (1964). A high-intensity flash-source. *Am. J. Psychol.* **77**, 643–646.

GRANIT, R. (1955). *Receptors and Sensory Perception*. New Haven: Yale University Press.

HEISS, W. D. & MILNE, D. C. (1967). Single fibers of cat optic nerve: 'thresholds' to light. *Science, N.Y.* **155**, 1571–1572.

KUFFLER, S. W., FITZHUGH, R. & BARLOW, H. B. (1957). Maintained activity in the cat's retina in light and darkness. *J. gen. Physiol.* **40**, 683–702.

LEVICK, W. R. (1962). Modification of a 256-channel analyzer for neurophysiological time analysis. *Rev. scient. Instrum.* **33**, 660–669.

LEVICK, W. R. & WILLIAMS, W. O. (1964). Maintained activity of lateral geniculate neurones in darkness. *J. Physiol.* **170**, 582–597.

LEVICK, W. R. & ZACKS, J. (1968). Bloch's law for the cat retinal ganglion cell. *J. Physiol.* **196**, 1–2*P*.

RODIECK, R. W. (1967). Maintained activity of cat retinal ganglion cells. *J. Neurophysiol.* **30**, 1043–1071.

ROSE, A. (1948). The sensitivity performance of the human eye on an absolute scale. *J. opt. Soc. Am.* **38**, 196–208.

RUSHTON, W. A. H. (1963). Increment threshold and dark adaptation. *J. opt. Soc. Am.* **53**, 104–109.

RUSHTON, W. A. H. (1965). The Ferrier lecture 1962. Visual adaptation. *Proc. R. Soc.* B **162**, 20–46.

TANNER, W. P., Jr. & SWETS, J. A. (1954). A decision-making theory of visual detection. *Psychol. Rev.* **61**, 401–409.

VAKKUR, C. J., BISHOP, P. O. & KOZAK, W. (1963). Visual optics in the cat, including posterior nodal distance and retinal landmarks. *Vision Res.* **3**, 289–314.