J. Physiol. (1972), 223, *pp.* 131–150
With 6 text-figures
Printed in Great Britain

COUNTING EVERY QUANTUM

BY B. SAKITT

From the Department of Physiology-Anatomy, 2570 Life Sciences Building, University of California, Berkeley, California 94720, U.S.A.

(*Received 1 November 1971*)

SUMMARY

1. Human subjects were asked to rate both blanks and very dim flashes of light under conditions of complete dark adaptation at 7° in the periphery. The ratings used were 0, 1, 2, 3, 4, 5, and 6.

2. For one subject (B.S.) the distributions of ratings were approximately Poisson distributions. The data were consistent with each rating being the actual number of effective quantal absorptions plus the number of noise events. This subject was presumably able to count every rod signal (effective absorptions plus noise).

3. For two other subjects, the data were consistent with the ratings being one less (L.F.) and two less (K.D.) than the number of effective absorptions plus noise. They were able to count every rod signal beginning with 2 and 3 respectively. A fourth subject's erratic data could not be fitted.

4. The fraction of quanta incident at the cornea that resulted in a rod signal was estimated to be about 0·03 which is consistent with physical estimates of effective absorption for that retinal region.

5. A simulated forced choice experiment leads to an absolute threshold about 0·40 log units below the normal yes-no absolute threshold. This and other results indicate that subjects can use the sensory information they receive even when only 1, 2 or 3 quanta are effectively absorbed, depending on the individual. Humans may be able to count every action potential or every discrete burst of action potentials in some critical neurone.

INTRODUCTION

It has been known for a long time that the human visual system is very sensitive to light. The maximum sensitivity occurs after complete dark adaptation when the rods dominate the visual system. By using a small, brief stimulus located in the periphery of the retina, humans can detect light when relatively small numbers of photons are absorbed in the retina.

Hecht, Shlaer & Pirenne (1942) showed that subjects could see light 50% of the time when about 50–150 photons entered the eye on each flash.

Using the word 'see' is misleading because it implies that one either sees or does not see without ambiguity. In other words, it implies that there is a threshold for seeing. However, one can ask subjects to state whether or not they saw a light and not worry about how they make the decision as to whether or not they saw anything. This is a familiar psychophysical technique and is used in what is called a 'threshold' type experiment.

If the average number of photons striking the cornea per flash is 100, then the average number of photon absorptions will be small. The fraction is not known exactly but it is reasonable to say that it is no more than 10% (Rushton, 1956*b*). Hence the average number of photon absorptions at 'threshold' is less than 10 and may be much less. Because of the quantum fluctuations of the light, the number of absorptions, like the number of incident photons, follows a Poisson distribution.

Suppose that a person has a fixed criterion for seeing. For example, suppose one sees if 7 or more photons are absorbed in a certain area within a certain time period. Then one can adjust the light intensity such that, on the average 7 photons are absorbed on each flash. Since the number of absorptions will fluctuate in a Poisson distribution, from tables one finds that only on 55% of the occasions will the actual number of absorptions equal or exceed seven. Hence a person with an absolutely fixed criterion for seeing will only see these flashes 55% of the time purely because of the variability of the light itself.

Hecht *et al.* (1942) found that the experimental frequency-of-seeing curves could be matched to cumulative Poisson distributions whose criteria were 5, 6 and 7 for their three subjects respectively. They concluded that subjects do indeed have a fixed criterion for seeing and that the variability of response could be accounted for by the quantum fluctuations of the light itself. The hypothesis of a fixed criterion for seeing is known as the high threshold theory. There have been many elaborations of this theory as well as arguments within its contexts.

Frequency-of-seeing curves from van der Velden (1946) had shallow slopes that were consistent with a criterion of two or three. However, Brindley (1954) and Pirenne & Marriott (1955) have pointed out that additional variability will flatten the frequency-of-seeing curve. Hence the curve only tells one a lower limit to the criterion.

Hecht (1945) has suggested that a single quantum was not sufficient for vision because it might be confused with a spontaneous excitation occurring somewhere along the optic pathway whereas it was extremely unlikely that a criterion of five quantum absorptions would be reached by spontaneous excitations. He suggested that this is why the criterion was

five absorptions. Barlow (1956) has done a quantitative signal versus noise analysis showing that lowering one's criterion will lower the threshold but increase the false positive rate. He suggests that thermal decompositions of rhodopsin molecules or other random events which are indistinguishable from quantum absorptions are a possible source of noise. He concludes that at least two, and probably many more (ten to twenty) excited rods are needed to give a sensation of light and that noise in the optic pathway limits its sensitivity.

In spite of disputes over the criterion for seeing, everyone agrees that relatively few quanta are required for vision. Also all the threshold theorists quoted above agree in ruling out a criterion of one quantum on the grounds that spatial summation is not complete for large test objects whereas a criterion of one should imply a threshold independent of area.

In contrast to threshold theories, a theory of signal detection and statistical decision has been developed by Tanner & Swets (1954) and Peterson, Birdsall & Fox (1954) which assumes that there is no such thing as a threshold. This theory assumes that there exists an internal decision variable which has a *continuous* probability distribution of background or noise events as well as a different continuous distribution of events caused by a stimulus. The observer can choose any response criterion whatsoever, not necessarily integral values. The lower his response criterion is, the higher his frequency of seeing and the higher his false positive rate will be. Furthermore, the theory predicts that his hit rate will be increased by a lower criterion by more than it could be by just chance guessing. Signal detection theory has been shown to explain a great deal of auditory data and visual experiments done on moderately high backgrounds. Nachmias & Steinman (1963) have verified some of the predictions of the theory for absolute visual detection.

The purpose of this paper is to unify the signal detection and threshold approaches to the problem of absolute visual sensitivity. By presenting stimuli near the absolute threshold and enlarging the subject's repertoire from a binary decision (seen versus not seen) to a rating scheme which permits one of many responses, it is shown in this paper how these old questions about threshold criteria can be approached with more insight.

METHODS

Stimulus. The stimulus was a 29′ disk located about 7° in the temporal retina of the left eye. It was a 16 msec blue-green flash corresponding to either 66 or 55 photons on the average at the cornea, hereafter called the strong and weak stimuli. Subjects dark-adapted for about an hour before each experiment and no background illumination was used at all. In addition to the strong and weak stimuli were blank trials which corresponded to no light. The right eye was covered by an eye patch during the experiment.

3 flash strengths
0, 55, 66

Apparatus. The experiments were done on a Maxwellian view optical system illustrated elsewhere (Sakitt, 1971). The light source was a tungsten filament lamp run on a regulated power supply. The filament was imaged on a stop and this was imaged on the subject's pupil. The stop was smaller than the filament image and the effective final image in the plane of the pupil had a 2·03 mm diameter. Since this was smaller than the natural pupil, no artificial pupil was used nor was the pupil dilated. The beam passed through neutral density filters and a blue-green filter, Ilford no. 603 which transmits between 470 and 520 nm, which straddles the peak of scotopic sensitivity. In addition to the fixed neutral density filters, a wheel containing two additional filters was placed in the beam which allowed the experimenter to set either the 'strong' or 'weak' luminance.

In the target plane (conjugate to the retina) was a wheel which permitted the experimenter to insert either a 29′ disk or an opaque aperture. Both wheels were moved for every trial even if they had to be returned to their original positions. They were very quiet and could not be heard above the sounds of fans of power supplies in the room. A shutter was placed in the test beam near the first filament image. It was electronically controlled by a switch held by the subject and opened within milliseconds of the release of the switch. Phototube measurements indicated that the shutter was very reliable.

The subject's head was fixed by biting on a dental impression in dentist's wax and by leaning on a forehead rest with side bars. A small weak red light source was used as a fixation light so that the stimulus appeared in the nasal field, about 7° from the fovea in the left eye when the subject was fixating.

The beams were properly focused in a preliminary experiment and checked periodically. The subject was positioned so that the filament image was focused on the pupil and centred on the pupil during fixation.

Observers. None of the subjects needed corrective lenses. The author (B.S.) and a hired subject (L.F.) were used as the basic subjects. They were both very experienced in visual psychophysical experiments. A third hired subject (K.D.) who was relatively inexperienced in vision experiments was used in the preliminary runs but was unable to remain a subject for reasons unrelated to the experiment. During the preliminary run with K.D., the wheels controlling the target and intensity were noisy but the noises were uniform and did not seem to give any cues to the subject. These auditory noises were eliminated before the runs using all other subjects. The data for all three subjects were used for this study. A fourth subject gave erratic data and was not used further.

Calibration. The retinal luminance for each beam was calculated according to the technique described by Westheimer (1966) and Rushton (1956*a*). All filters were removed from the beam. The luminance of a white screen placed a known distance from the place where the subject's pupil would normally be was measured with an SEI meter. The SEI meter had been calibrated against a 100 foot lambert standard source whose calibration was traceable to the National Bureau of Standards. This gave the retinal illumination in photopic trolands which was converted to scotopic trolands by knowing the colour temperature of the lamp. The scotopic effectiveness of the coloured filter was determined by psychophysical methods on the apparatus, using various techniques to insure that only the scotopic system was being used. This value agreed with that obtained by calculation using the known spectral transmission of the filter. The neutral density filters were calibrated either in a Perkin–Elmer Spectrophotometer or with a Gamma photometer. The time course of the flash was measured with a phototube.

The absolute luminances are not known as exactly as the relative luminances. The relative luminance was determined *in situ* with a Gamma photometer leaving the

blue-green filter in the beam and comparing the densities of the two neutral density filters that determined the weak and strong stimuli. The ratio of the luminances was 1·20. The estimates of the actual intensities were scotopically equivalent to 66 and 55 photons of wavelength 507 nm incident at the cornea on the average per flash, for the strong and weak stimuli.

Procedure. An experimental run consisted of a block of 160 trials composed of eighty blank trials, forty strong stimuli and forty weak stimuli. The subject knew these *a priori* probabilities but these trials were presented in a random order. Brief breaks were taken occasionally during the run. After a block was finished, the subject rested for 5–10 min and another block was run. Two blocks were done each day.

During the practice sessions, the task of each subject, including the author, was to consciously think about one's own process of rating flashes of light. We tried to develop as many categories of sensory impressions as possible that we felt we could reliably use. We finally worked out the following rating system:

0 meant that we did not see anything;
1 meant that it was very doubtful if a light was seen;
2 meant that it was slightly doubtful if a light was seen;
3 meant a dim light;
4 meant a moderate light;
5 meant a bright light;
6 meant a very bright light.

We attempted to rate our impressions only in the area where the flash was anticipated and only when we opened the shutter.

After every rating, the subject was told what the trial was – strong, weak or blank. This was done during the practice sessions as well as during the actual experimental sessions.

Two subjects (B. S. and L.F.) had a large number of practice sessions and data were taken when a final stable rating system was developed. Another subject (K.D.) had few practice sessions (for reasons unrelated to the study) but seemed to give stable ratings. A fourth subject (C.L.) had a rating system that was so unstable it was not included in the final data although the other three subjects who were used all seemed able to give stable ratings from day to day. Data were taken for this experiment on 5 days. However, for another rating study in which B.S. and L.F. participated (to be discussed elsewhere) it was found that they continued to use the same rating systems.

Definitions

i = the rating (i = 0, 1, 2, 3, 4, 5 or 6).
$p_i(S)$ = probability of saying i when the strong stimulus is presented.
$p_i(W)$ = probability of saying i when the weak stimulus is presented.
$p_i(B)$ = probability of saying i when the blank stimulus is presented.
$P_i(S)$ = cumulative probability of saying i or greater when the strong stimulus is presented.
$P_i(W)$ = cumulative probability of saying i or greater when the weak stimulus is presented.
$P_i(B)$ = cumulative probability of saying i or greater when the blank stimulus is presented.
Q_e = average number of quanta (scotopically equivalent to 507 nm) at the cornea per flash.
N_i = number of times that the rating i was given.
$a(S)$ = average number of rod signals due to the strong stimulus.

$a(W)$ = average number of rod signals due to the weak stimulus.
$a(B)$ = average number of rod signals due to the blank stimulus.
S, W and B after a symbol will denote strong, weak or blank stimulus respectively.

RESULTS

Rating distributions. Table 1 shows the actual data for each subject. N_i is the number of trials on which the rating i was given for a particular stimulus for each subject.

For each stimulus the average rating $\bar{i}$ was calculated

$$\bar{i} = \sum_{i=0}^{6} i p_i, \tag{1}$$

where p_i is the probability of saying i for a particular stimulus. Also

$$\bar{i^2} = \sum_{i=0}^{6} (i)^2 p_i. \tag{2}$$

The variance of each rating distribution is

$$V = \sum_{i=0}^{6} (i-\bar{i})^2 p_i = \bar{i^2} - (\bar{i})^2. \tag{3}$$

Table 1 gives $\bar{i}$ and V for all stimuli for all subjects.

TABLE 1. Number of trials (N_i) on which each rating was given for each stimulus. Last two columns are average ratings ($\bar{i}$) and variances (V) of rating distributions

Subject	Signal	N_0	N_1	N_2	N_3	N_4	N_5	N_6	ΣN	$\bar{i}$	V
B.S.	Strong	70	75	66	109	63	12	5	400	2·19	2·22
B.S.	Weak	83	104	78	87	36	11	1	400	1·82	1·93
B.S.	Blank	566	192	33	9	0	0	0	800	0·36	0·38
L.F.	Strong	91	85	81	83	40	19	1	400	1·89	2·20
L.F.	Weak	133	83	78	58	38	9	1	400	1·54	2·10
L.F.	Blank	585	163	32	18	2	0	0	800	0·36	0·48
K.D.	Strong	105	74	112	106	3	0	0	400	1·57	1·35
K.D.	Weak	149	74	94	81	2	0	0	400	1·28	1·39
K.D.	Blank	697	78	19	6	0	0	0	800	0·17	0·23

Fig. 1 plots the average rating for B.S. versus the average number of quanta at the cornea. The three independent experimental points are well fitted by the relation

$$\bar{i} = 0{\cdot}0274 Q_c + 0{\cdot}36 \tag{4}$$

where Q_c is the average number of quanta at the cornea equivalent to 507 nm. Thus the average rating is linear with the average number of quanta at the cornea.

It is interesting to compare this with the expression for a, the average number of rod signals.

The concept of noise in the absence of light has been proposed by Hecht (1945), Barlow (1956) and by Tanner & Swets (1954) and is discussed in great detail by them. The following analysis will assume that noise events do occur. Consider the hypothesis that every effective absorption produces exactly one rod signal and that the noise adds to this. Then the average number of rod signals is

$$a = fQ_c + x = f(Q_c + X_c), \tag{5}$$

where f is that fraction of the incident quanta that produce rod signals, x is the average number of noise events (rod signals in the dark) and X_c is

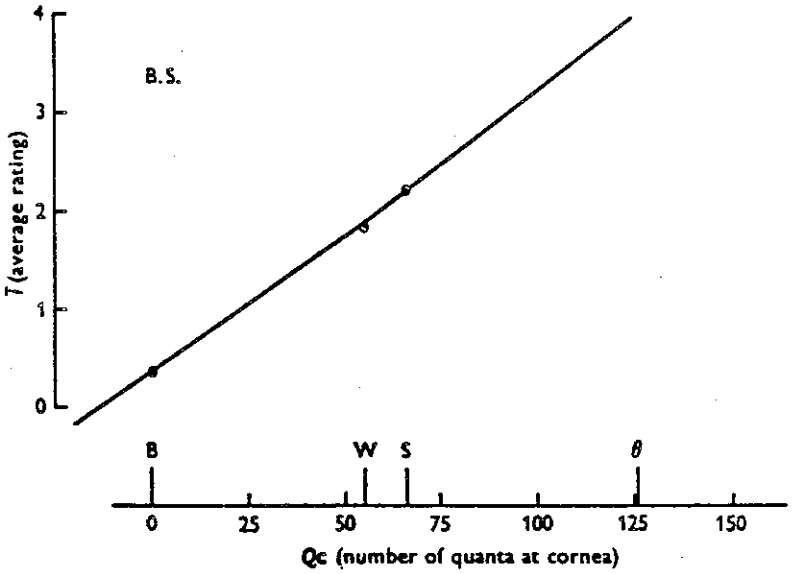

Fig. 1. Plots $\bar{i}$, the average rating against Q_c, the average number of quanta (equivalent to 507 nm) at the cornea for B.S. The straight line through the points is $\bar{i} = 0{\cdot}0274Q_c + 0{\cdot}36$.

the dark light or the average number of equivalent quanta at the cornea in the dark. From now on the expression 'effective quantum absorption' will mean an absorption that produces a signal.

The great similarity between eqns. (4) and (5) suggests that the average rating $\bar{i}$ may be proportional to the average rod signal a. If we assume that the constant of proportionality is one, then for B.S., $f = 0{\cdot}0274$, $x = 0{\cdot}36$ and $X_c = 13{\cdot}1$. Then for subject B.S. we have

$$a = 0{\cdot}0274Q_c + 0{\cdot}36, \tag{6}$$

or

$$a = 0{\cdot}0274(Q_c + 13{\cdot}1). \tag{7}$$

This value of f is consistent with the average value of 0·03 found by

This analysis assumes that the rating equals the number of absorbed photons.

Barlow (1962) for the quantum efficiency at 7° in the periphery. Also Rushton (1956*b*) has estimated that 10% of the incident quanta are absorbed at 20°. Since the rod density at 7° is 75% that at 20° (Østerberg, 1935), only 7·5% of the incident quanta are absorbed at 7°. Hagins (1955) has evidence that only half of all quantal absorptions are effective in producing a signal which implies $f = 0{\cdot}038$.

On the assumption that the average rating $\bar{i}$ is equal to a, the average rod signal predicts a value of f that is consistent with these other estimates of f within the experimental uncertainties. Consider a more radical version of this hypothesis: *the rating in any trial is the number of rod signals produced by the real and dark light* within a particular retinal region and within a critical period of time.

The number of effective absorptions, like the number of incident quanta, is Poisson distributed. The noise events will be Poisson distributed if they are due to thermal decompositions of rhodopsin or if they are due to a possible noise event occurring with a very small probability from each of a large number of independent units. Even if the noise is not rigorously Poisson, the rod signal a, usually dominated by the real light contribution, will be close to a Poisson distribution if not exactly one.

If the rating is equal to the number of rod signals, the ratings should follow a Poisson distribution. From Table 1, it is seen that B.S. has the average rating $\bar{i}$ approximately equal to the variance V for all three stimuli. Since a Poisson distribution has its mean equal to its variance, it seemed worth while to pursue this. The mean rod signals $a(S)$, $a(W)$ and $a(B)$ were estimated as the respective values of $\bar{i}$, the average ratings. The experimental points in Fig. 2 are the cumulative probabilities $\bar{P}_i$ for giving a rating of i or more plotted against the log (rod signal) = $\log a$. The solid curves are theoretical cumulative Poisson probabilities for criteria of 1, 2, 3, 4, and 5, which are given by

$$\bar{P}(c, a) = e^{-a} \sum_{n=c}^{\infty} \frac{a^n}{n!} \tag{8}$$

which are the cumulative probabilities of c or more rod signals occurring when a is the average number occurring. To the extent that the experimental points lie on these theoretical Poissons, it is seen that a rating of 1 or more is equivalent to a criterion of 1 or more rod signals, a rating of 2 or more is equivalent to a criterion of 2 or more rod signals, etc. To a fair approximation, the rating is equal to the number of rod signals which is just the number of effective quantum absorptions plus noise events.

The absolute threshold (50% seeing), obtained by a conventional yes–no method in a different study on the same apparatus (Sakitt, 1971), occurred at $Q_c(\theta) = 126$. Using Fig. 1 or eqn. (6) implies $a(\theta) = 0{\cdot}0274\ (126) + 0{\cdot}36$

= 3·81. This point, labelled θ in Fig. 2, lies on the Poisson curve for a rating of 4 or a criterion of 4 rod signals. This is interesting because it is the lowest criterion for which B.S. had no false positives out of 800 trials and is the normal criterion used by B.S. in yes–no experiments.

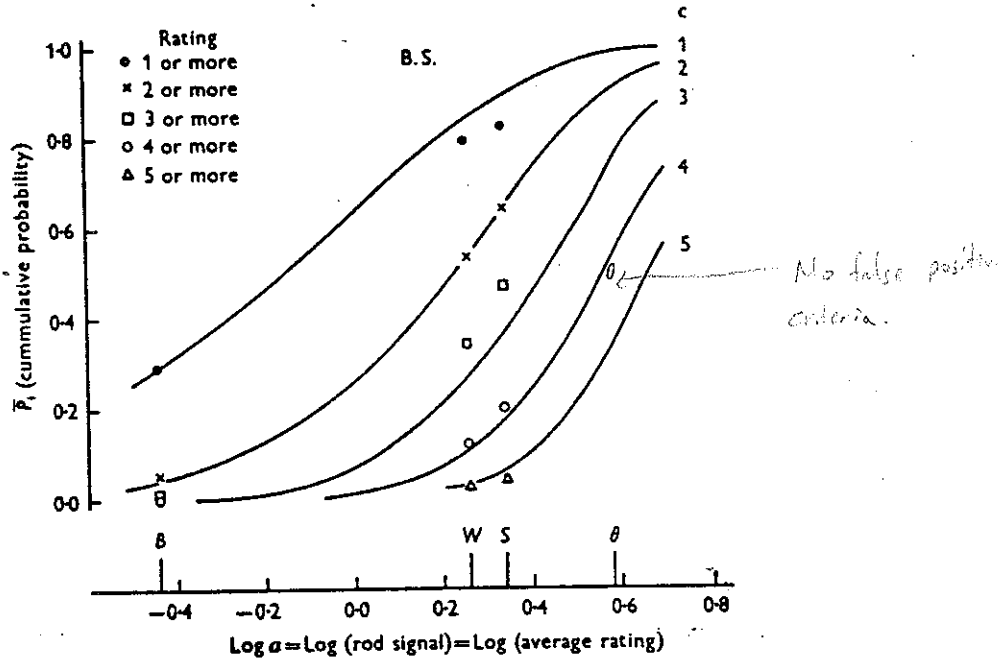

Fig. 2. The experimental points are the cumulative probabilities $\bar{P}_i$ for B.S. giving a rating i or more. The abscissa is the log of the rod signal which is the same as the log of the average rating. The points labelled B, W and S are at the values of a for the blank, weak and strong stimuli respectively. The symbol θ refers to the absolute threshold as described in the text. The smooth curves are theoretical cumulative Poisson probabilities $\bar{P}(c, a)$ that c or more rod signals occur when a is the average number occurring.

Usually when one is attempting to match cumulative Poissons, there is some leeway because the lateral shift is not known (equivalent to log f). In this case, there is *no* leeway because of the assumption that the values of log a are determined by the average ratings. The fit of the experimental points to the theoretical curves is only fair; but it should be emphasized that all the sixteen experimental cumulative probabilities (including the absolute threshold) lie near the appropriate cumulative Poisson curves although not a single arbitrary parameter was used to fit the data; and that the ratings were linear with light intensity in the same manner as the rod signals are.

From Fig. 2, it is seen that if B.S. used a criterion of 3, the absolute threshold (50 %) would occur at $a = 2{\cdot}70$ and from eqn. (6), $Q_c = 87$. A

criterion of 3 would only have a false positive rate of 1%. If the criterion was only 2, the false positive rate would be slightly over 5%, but the absolute threshold would occur at $a = 1{\cdot}70$ and $Q_c = 50$ quanta. For criteria of 2 or more, the values of the absolute threshold and the false positive rates are all consistent with the range found by observers in many studies reported throughout the visual literature. But now consider a criterion of 1. It leads to a value of $a = 0{\cdot}70$ at 50% seeing which means $Q_c = 12{\cdot}6$ and the false positive rate is 29%. This false positive rate would not be permitted in a yes–no experiment. It is possible to count a single effective quantal absorption as the rating of 1 does for B.S. However, if observers are discouraged from giving too many false positives, they won't use such a low criterion.

Since the rating of 4 was used in yes–no experiments as a criterion for 'seeing', it might be argued that there exists an internal sensory threshold (not necessarily 4 rod signals) that corresponds to the rating of 4. Perhaps a subject could invent lower ratings without any sensory information by randomly giving the rating 3 on some fraction of the occasions for which the threshold (rating of 4) was not reached. From Fig. 2, the strong stimulus is 'seen' with a criterion corresponding to a rating of 4 only 20% of the time, but is rated 3 about 47% of the time. Yet the false positive rate has only gone up to 1%. Hence the rating of 3 corresponds to a lower criterion than a rating of 4 and is not due to chance guessing. Similarly by comparing the ratings of 2 and 3, it is seen that the rating of 2 corresponds to a lower criterion than the rating of 3. Hence one can say that ratings of 2 or more correspond to 'seen'. Now consider a rating of 1. If it is not based on sensory information, then of those trials that are not rated 2 or more, the fraction rated 1 would be a constant, equal to the guessing rate. That is $p_1/(p_0+p_1)$ would be the same for the strong and blank stimuli. The actual values are 0·52 for the strong and 0·25 for the blanks. Hence the rating of 1 is *not* due to guessing but corresponds to the lowest criterion used. Hence it was possible for B.S. to use three different criteria below the normal one used in yes–no experiments.

The results from the data on B.S. are: (1) the ratings are linear with the intensity of the stimulus; (2) the average ratings correspond to the average number of effective quantum absorptions plus noise events; (3) lower ratings correspond to genuinely lower criteria and are not due to chance guessing; (4) the experimental cumulative probability of saying i or more is approximately equal to the theoretical Poisson cumulative probability that i or more rod signals occur.

The conclusion is that for this subject the rating is the number of effective quantum absorptions plus noise events. This subject was able to count every single rod signal.

From Table 1, it is apparent that L.F.'s data could not be matched to a Poisson since the averages are not equal to the variances. However, suppose that L.F. started counting not at 1 rod signal, but at 2. Even if the rod signals are Poisson distributions, the ratings would not be. The expected values would be

$$\bar{i} = \sum_{n=1}^{\infty} (n-1)p(n, a) = a - \bar{P}(1, a) \tag{9}$$

and

$$\overline{i^2} = \sum_{n=1}^{\infty} (n-1)^2 p(n, a) = a^2 - \bar{i}, \tag{10}$$

where a is the mean of the rod signal Poisson distribution, $p(n, a)$ is the probability that exactly n events occur and $\bar{P}(1, a)$ is the probability that at least 1 event occurs. Applying eqns. (9) and (10) yields values of $a(S)$, $a(W)$ and $a(B)$ of 2·77, 2·45 and 0·98 respectively.

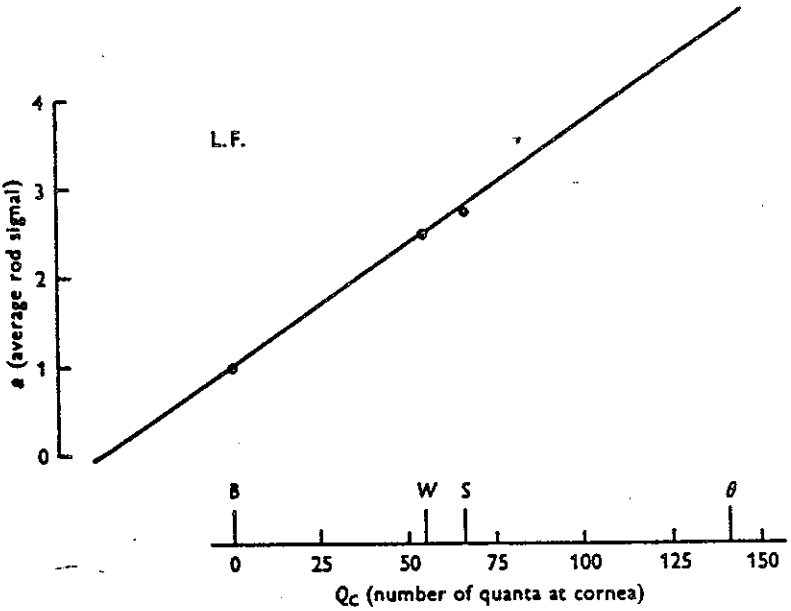

Fig. 3. Plots the values of the rod signal a as obtained from the rating distributions of L.F. against Q_c, the average number of quanta (equivalent to 507 nm) at the cornea. The straight line through the points is $a = 0{\cdot}0270\ Q_c + 0{\cdot}98$.

Fig. 3 is a plot of a versus Q_c for L.F. The straight line through the three points is given by

$$a = 0{\cdot}0270Q_c + 0{\cdot}98, \tag{11}$$

or

$$a = 0{\cdot}0270(Q_c + 36{\cdot}3). \tag{12}$$

The points fit the line well. The hypothesis that L.F.'s ratings were one less than the number of rod signals leads to a linear relation between rod signal and stimulus.

The points in Fig. 4 are the experimental probabilities of giving a rating of i or more plotted against $\log a$ = log (rod signal). The values of a were determined by eqns. (9) and (10) as described above. The continuous curves are the theoretical cumulative Poisson probabilities of c or more rod signals occurring when a is the average number occurring.

There is a moderately good fit of the points to the Poisson curves although there are *no* arbitrary parameters used to fit the data. Hence the data are consistent with the hypothesis that L.F.'s rating of any trial was one less than the number of effective absorptions plus noise.

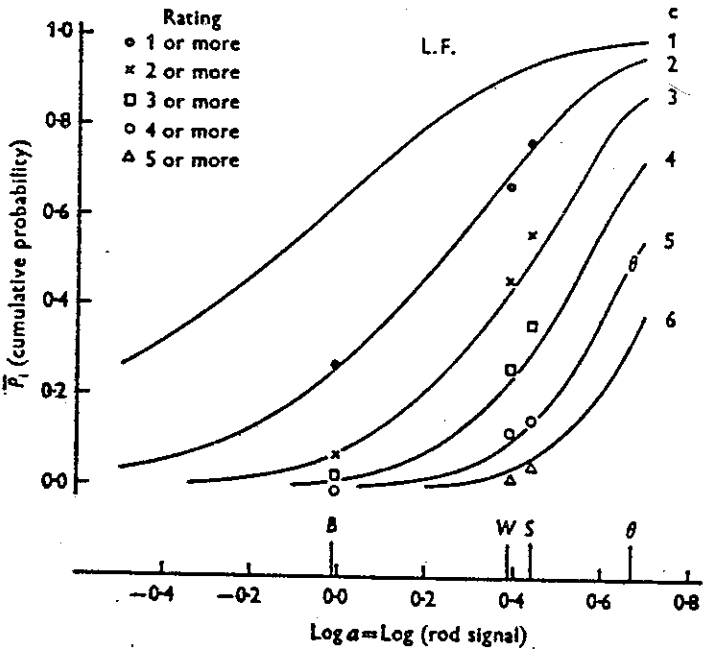

Fig. 4. The experimental points are the cumulative probabilities $\bar{P}_i$ for L.F. giving a rating i or more plotted against $\log a$ = log (rod signal). The smooth curves are the theoretical cumulative Poisson probabilities $\bar{P}(c, a)$ that c or more rod signals occur when a is the average number occurring. Symbols have same meaning as in Fig. 2.

In a previous study (Sakitt, 1971), the absolute threshold (50 % seeing) for this subject was found to be 141 quanta (equivalent to 507 nm) by a yes–no method. This determines $a(\theta)$ as 4·79. This gives rise to an experimental point in Fig. 4 which lies on the curve for a rating of 4, which is equivalent to a criterion of 5 rod signals.

By arguments similar to those used for the B.S. data, it is seen that the low ratings are not due to chance guessing but correspond to low criteria.

Attempts to fit the K.D. data to this type of '2 plus' Poisson would not give a linear relation between rod signal and stimulus. However, the data almost fitted a '3 plus' Poisson distribution with $a(S) = 3{\cdot}70$, $a(W) = 3{\cdot}20$ and $a(B) = 1{\cdot}20$. Fig. 5 is a plot of a versus Q_c for K.D. The straight line through the points is

$$a = 0{\cdot}0378Q_c + 1{\cdot}20, \tag{13}$$

or

$$a = 0{\cdot}0378(Q_c + 31{\cdot}8). \tag{14}$$

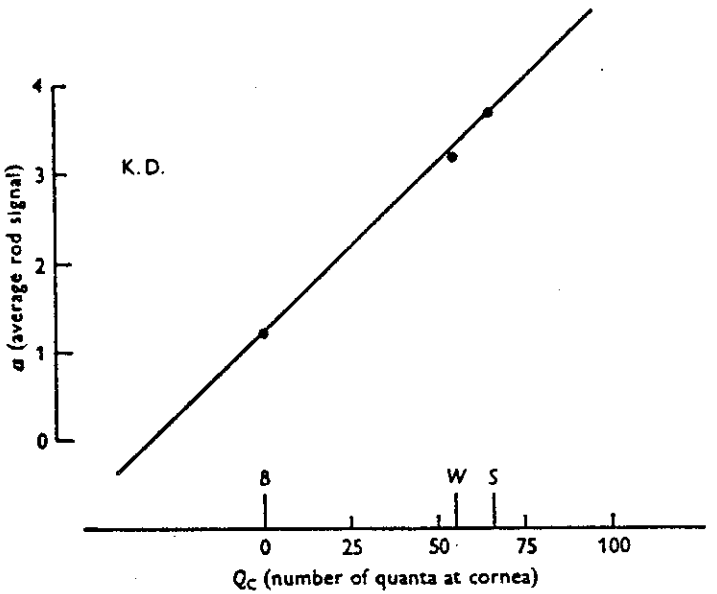

Fig. 5. Plots the values of the rod signal a as obtained from the rating distributions of K.D. against Q_c, the average number of quanta (equivalent to 507 nm) at the cornea. The straight line through the points is $a = 0{\cdot}0378Q_c + 1{\cdot}20$.

Hence the rod signal a is linear with the intensity of the stimulus. Fig. 6 plots the experimental cumulative probabilities for K.D. and the theoretical curves for '3 plus' Poisson distributions. The points lie close to the theoretical curves and are consistent with the hypothesis that the rating was two less than the number of rod signals but that only ratings up to 3 were used. There is no absolute threshold data for K.D. but if the rating of 3 (5 rod signals) were used, then the absolute threshold would be $Q_c = 68$ and the false positive rate would be 0·0075 which is very low.

The analysis of the L.F. and K.D. data raises the possibility of trying to fit the B.S. data to a '2 plus' or '3 plus' Poisson. The latter cannot be done. A '2 plus' Poisson can be fitted approximately but appears, by eye, to be a definitely worse fit than the ordinary Poisson. A '2 plus' Poisson for B.S.

implies $x = 0{\cdot}93$ and $f = 0{\cdot}0320$. This gives a value for the noise that is closer to that of the other subjects which is an argument for it. On the other hand the mean is approximately equal to the variance for the strong, weak and blank stimuli only for the B.S. data. This strongly suggests that only the B.S. data can be fitted to an ordinary Poisson and the fact is that the fit is better than for the '2 plus' Poisson.

Forced choices. Suppose the strong stimulus is presented in one of two time intervals, the other interval containing the blank. Suppose the order of presentations is random. The subject could be asked to choose which interval contained the strong stimulus.

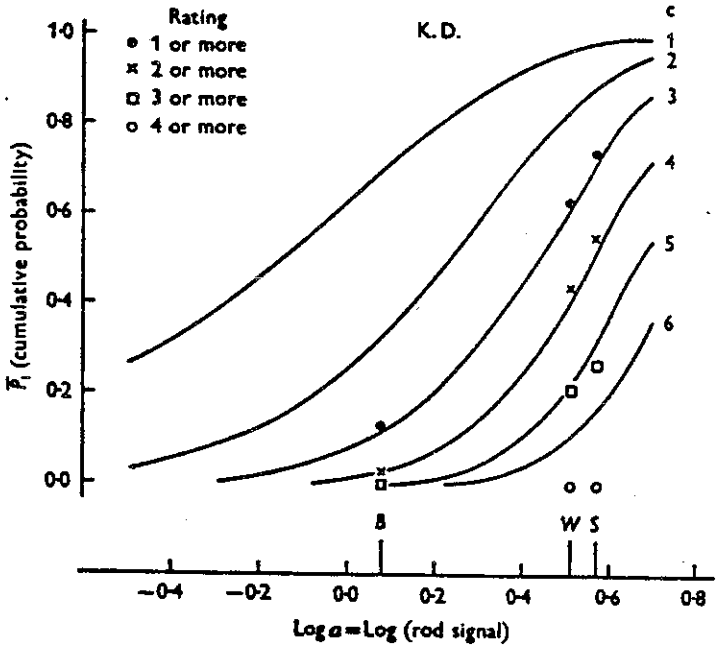

Fig. 6. The experimental points are the cumulative probabilities P_i for K.D. giving a rating i or more plotted against $\log a$ = log (rod signal). The smooth curves are the theoretical cumulative Poisson probabilities $P(c, a)$ that c or more rod signals occur when a is the average number occurring. Symbols have same meaning as in Figs. 2 and 4.

One way of choosing would be to rate each interval according to the procedure of the rating experiment described above. Then the subject would choose the interval with the highest rating as the best bet for the strong stimulus. In case of a rating tie, the subject would guess randomly. Using this strategy, the fraction of correct answers, G, in this simulated forced choice experiment would be

$$G(S-B) = \sum_{i=1}^{6} p_i(S)\,[1-P_i(B)] + \frac{1}{2}\sum_{i=0}^{6} p_i(S)p_i(B). \qquad (15)$$

The first term sums all possibilities when the interval containing the strong stimulus gets a higher rating than the blank. The second term is due to chance guessing when ties arise.

Similarly the fraction correct can be calculated for weak versus blank and strong versus weak. Table 2 sums the results for simulated forced choices calculated from the current series of rating experiments.

TABLE 2. Percentage correct in a simulated two alternative forced choice experiment

	B.S.	L.F.	K.D.
Strong *vs.* Blank	84·9	80·8	83·1
Weak *vs.* Blank	80·9	74·3	77·6
Strong *vs.* Weak	57·3	56·9	56·5

Sometimes forced choice experiments with a light versus a blank are done as a substitute for yes–no threshold experiments. Threshold is then defined at the 75% correct level. The reasoning is that 'threshold' occurs at 50% seeing but in a forced choice experiment the subject guesses right half the time when nothing is seen, bringing his per cent correct at 'threshold' to 75%. If this reasoning is correct, the same intensity that gives 50% seen in a yes–no experiment will produce 75% correct in a forced choice. Consider the data from Table 2. B.S. has almost 81% correct in a forced choice between the weak (55 quanta) and blank. Hence the 75% correct level is below 55 quanta at the cornea. But a conventional yes–no experiment resulted in 50% seeing with 126 quanta at the cornea. Thus the forced choice threshold is more than 0·36 log units below the yes–no threshold. Similarly for L.F., the yes–no absolute threshold is at 141 quanta at the cornea but the forced choice threshold occurs at about 55 quanta at the cornea. The forced choice threshold for L.F. is 0·41 log units less than the yes–no threshold. If real forced choice experiments were done, subjects could always use their rating system and get identical results to the simulated experiment. It is possible that they could do even better. Blackwell (1953) has found that increment thresholds obtained by forced choice were significantly smaller than those obtained with a yes–no procedure.

The fact that 75% correct in a forced choice experiment occurs at an intensity level that is much lower than the level of 50% seeing confirms that subjects can use sensory information received from flashes for which they report 'not seen'.

DISCUSSION

Quantum counting. The results of these experiments indicate that the ratings given by B.S. are the actual number of rod signals which are the number of effective quantal absorptions plus noise events. The ratings given by L.F. are the number of rod signals minus 1 and for K.D., the number minus 2. The evidence is based on the linearity of the rod signal with stimulus intensity and the fact that the cumulative probability of giving any rating is close to the theoretical cumulative probability that the corresponding number of rod signals occur. The conclusion is that subjects can adopt any criterion for 'seeing' even a criterion of 1 in some cases.

It is important to remember that the rating systems used here were not completely arbitrary. Part of the experiment consisted of practice sessions where the task of all the subjects was to decide how many different types of sensory impressions they each received when blanks and dim flashes of light were presented. Then numbers were assigned to describe these different sensations, as described in the Methods section. The results are consistent with the hypothesis that each different type of sensory impression corresponded to a different number of rod signals. These could be transmitted by either single action potentials or by discrete bursts of action potentials in some critical neurone.

Previous counter arguments. Brindley (1954) and Pirenne & Marriott (1955) have pointed out that a shallow frequency-of-seeing curve may not mean a low criterion but may be produced by biological variability. If the shallowness is due to variability then they have shown that the curve will shift over. That is, variability will decrease the slope and the apparent criterion but will *never* decrease the threshold. However, in the present study, as the criteria (ratings) decreased so did the absolute threshold. It was shown that subjects could greatly reduce their thresholds by using lower ratings. It was also shown that as the rating decreased the percentage 'seen' increased much faster than the false positive rate did. This result plus the result of the simulated forced choice experiment indicates that subjects are receiving information even when their 'normal' criteria are not reached.

A previous argument against criteria of less than 4 was given by Brindley (1954). Large test objects covering many degrees on the retina probably excite many independent detectors. Brindley proved that the slope of the frequency-of-seeing curve for a finite number of detectors must be *less* than the slope of a theoretical curve for an infinite number of detectors, each having the same criterion c. He showed that the slope of the curve of Bouman and van der Velden (1947) using a 23° field was *greater* than the limiting slopes for criteria of 2 or 3. This seemed to be proof that

Bouman and van der Velden's hypothesis of a criterion of 2 per detector was false providing the threshold was determined by many independent detectors. The latter assumption seems reasonable for such a very large test field. Hence the criteria of the individual detectors must be at least 4.

The answer to this apparent paradox is simply that subjects do not use the same criterion for all types of test objects. Assuming that each independent detector in the retina occupies about a degree, a 23° field stimulates at least 500 detectors.

A subject with a criterion of 2 and a noise level of 0·36 (like B.S.) would have a false positive rate of 0·052 if only one detector was being stimulated and that this fact was known. However, if any of 500 independent detectors could be excited, the false positive rate is the probability that at least one detector has two noise events and would be

$$1-(1-0{\cdot}052)^{500} = 1-3\times10^{-13} = \text{very nearly } 1.$$

A false positive rate of one is definitely unacceptable.

Suppose the criterion were 4. Then the false positive rate for one detector would be 0·0005 which is very low and acceptable, but the false positive rate for 500 such detectors would be 0·25 which is very unacceptable.

In threshold experiments, subjects know the size and location of the test stimulus. If it is relatively small (less than 1°), they can have a low criterion since they won't say yes to noise events occurring several degrees away from the expected location of the test. For large test objects, subjects must raise their criteria so that the probability that at least one detector reaches the criterion in the dark is very small. Hence the argument about large test fields is not valid about the criteria used with small fields.

There is an additional argument that everyone has used against a criterion of one quantum absorption. If the criterion were one, then complete spatial summation should occur for all stimuli of any area. The answer to this argument is that a criterion of one leads to a high false positive rate and hence is not normally used. However, some human subjects can count a single quantum even though this is not their normal criterion.

Comparison with ganglion cell results. Recent experiments by Barlow, Levick & Yoon (1971) were done on cat retinal ganglion cells under conditions of complete dark adaptation. For one unit, they plotted the cumulative probability of producing 3 or more, 6 or more, 9 or more, etc. spikes against log (stimulus plus dark light). These points fit theoretical cumulative Poisson curves with criteria of 1, 2, 3,..., etc. based on only two parameters, equivalent to the f and x used here. They also found that the variance of the pulse number distribution divided by the mean was 3. Their conclusion was that an effective absorption can produce more than one impulse, and in the above case, three impulses.

If we compare the human data from B.S. to the cat data, the variance for each rating distribution of B.S. is about equal to the mean. It seems possible that every effective absorption produces one action potential in a critical neurone of B.S. and that the ratings are the sum of the number of action potentials produced by the stimulus plus those produced by noise events. For the other subjects counting begins at two and three absorptions. It may be that it takes more than one rod signal to fire an action potential for L.F. and K.D. or it may be that they start counting spikes at a higher level in order to reduce the false positive rate for the lowest criterion.

Biased experimenters. It is usually thought that trained observers do 'better' than naive ones in psychophysical experiments. One sign of a 'good' observer is that he never, or rarely, says 'seen' when no stimulus is presented in a yes–no experiment. In other words, trained observers are able to have a zero or extremely small false positive rate. This is not discussed in the classic paper by Hecht, Shlaer & Pirenne (1942) but is discussed by Pirenne & Marriott (1962, pp. 323ff). In this latter paper, it is revealed that:

"The curves published by Hecht *et al.* were all 'good' curves, that is, they were (relatively) steep curves obtained under the most reliable conditions.

They were given by subjects who never answered 'seen' in response to 'blanks'. Steeper curves than these could not be obtained, but 'bad' curves having a greater range of uncertain seeing were obtained by Hecht *et al.* in certain cases. Two subjects at first gave extremely shallow curves, some of which were even shallower than the Poisson sum for $c = 1$. The curves gradually became steeper as the subjects repeated the measurements, finally corresponding to $c = 5$ to 7, but they never became smooth enough to be included in the publication. In the case of more reliable subjects; occasional 'bad' curves are generally associated with fatigue or other conditions which seem likely to conduce to high biological variations."

Since it is known that biological variability *may* flatten a frequency-of-seeing curve, it has been common practice for observers to call shallow slopes 'bad' data. Furthermore false positives are considered 'bad' in yes–no experiments. So what we, as experimenters have been doing is training naive observers to raise their criteria until their false positive rate is too low to be even reliably measured. This procedure that we have been using has biased our results towards high criteria. Even those experimenters claiming a criterion of two or three must have been doing this. The rating of one in the B.S. data would not have been 'permitted' in any yes–no experiment because of the high false positive rate. Also, if one neglects noise, the apparent slope of the curve looks shallower than that

for a criterion of one, and so would have been attributed to biological variability rather than to noise events.

The bias of experimenters leads to high criteria but the only conclusion should be that we are capable of having high criteria, not that we are compelled to.

Accuracy of fit. Although all three subjects were able to count their rod signals, there was variability in the starting point. It is possible that this is due to innate physiological differences between the observers. It is also possible that some observers are reluctant to begin counting at one rod signal because they have a bias against giving too many ratings greater than zero (not seen) for the blanks. Therefore, it may be significant that the author B.S., who was able to respond to a single rod signal, was well aware that false positives are not necessarily 'mistakes' and presumably was not very biased against them. The results may depend on observer bias and experience as well as the instructions to subjects about developing a rating system.

Calculation of χ^2 for the B.S. data indicates that P is less than 0·01. It is apparent from this as well as by just looking at Figs. 2, 4 and 6 that the fit of the experimental points to the theoretical Poisson curves is worse than what would be expected purely on sampling error for all three subjects. Since the probabilities for the strong and weak stimuli are based on 400 trials each, typical 95% confidence limits are $\pm 0{\cdot}05$ from the experimental value. There are probably systematic errors that are of the same order of magnitude. Aside from the obvious ones of fluctuations of the intensity of the apparatus from day to day and the inaccuracies of calibrations which probably produced errors of a few per cent, there must be *some* biological variability. Although the three subjects had stable rating systems, it is possible that a few per cent of the time they made incorrect ratings. The one subject who could not be used gave unstable ratings. The rating distributions on certain days varied so much that it is likely that unreliable ratings were being given often. However, even for the three stable subjects, the subjective difficulty in rating experiments makes it seem very possible that there were small fluctuations in the rating system, with biases. For example a subject might be biased against giving the rating one and have a tendency to under use it. It would only require a very small bias (a few per cent) to account for all the experimental points. The best one can do in principle is to be an ideal quantum counter. However, one can also do worse because of inattention, biases, etc. However, the fact that all the experimental points lie near the theoretical Poisson curves indicates that, to a large extent, people can count every quantum, even if they are not infallible.

I wish to thank Professor Horace B. Barlow and Professor Gerald Westheimer for helpful discussions in preparing this manuscript. This investigation has been supported by Grant EY 00276 from the National Eye Institute, U.S. Public Health Service.

REFERENCES

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

BARLOW, H. B. (1962). Measurements of the quantum efficiency of discrimination in human scotopic vision. *J. Physiol.* **160**, 169–188.

BARLOW, H. B., LEVICK, W. R. & YOON, M. (1971). Responses to single quanta of light in retinal ganglion cells of the cat. *Vision Res.* **11**, suppl. 3, 87–101.

BLACKWELL, H. R. (1953). Psychophysical thresholds: experimental studies of methods of measurement. *Bull. Eng. Res. Inst. U. Mich.*, no. 36.

BOUMAN, M. A. & VAN DER VELDEN, H. (1947). The two-quanta explanation of the dependence of the threshold values and visual acuity on the visual angle and the time of observation. *J. opt. Soc. Am.* **37**, 908–919.

BRINDLEY, G. S. (1954). The order of coincidence required for visual threshold. *Proc. Phys. Soc.* B **67**, 673–676.

HAGINS, W. A. (1955). The quantum efficiency of bleaching of rhodopsin *in situ*. *J. Physiol.* **129**, 22–23*P*.

HECHT, S. (1945). Energy and vision. In *Science in Progress*, Fourth Series, pp. 75–97. New Haven: Yale University Press.

HECHT, S., SHLAER, S. & PIRENNE, M. H. (1942). Energy, quanta and vision. *J. gen. Physiol.* **25**, 819–840.

NACHMIAS, J. & STEINMAN, R. (1963). Study of absolute visual detection by the rating-scale method. *J. opt. Soc. Am.* **53**, 1206–1213.

ØSTERBERG, G. (1935). Topography of the layer of rods and cones in the human retina. *Acta ophthal.* suppl. 6.

PETERSON, W. W., BIRDSALL, T. G. & FOX, W. C. (1954). The theory of signal detectability. *Trans. IRE Professional Group on Information Theory*, PGIT-4, 171–212.

PIRENNE, M. H. & MARRIOTT, F. H. C. (1955). Absolute threshold and frequency-of-seeing curves. *J. opt. Soc. Am.* **45**, 909–912.

PIRENNE, M. H. & MARRIOTT, F. H. C. (1962). The quantum theory of light and the psycho-physiology of vision. In *Psychology: A Study of a Science*, vol. 1, pp. 288–361, ed. KOCH, S. New York: McGraw-Hill.

RUSHTON, W. A. H. (1956*a*). The difference spectrum and the photosensitivity of rhodopsin in the living human eye. *J. Physiol.* **134**, 11–29.

RUSHTON, W. A. H. (1956*b*). The rhodopsin density in human rods. *J. Physiol.* **134**, 30–46.

SAKITT, B. (1971). Configuration dependence of scotopic spatial summation. *J. Physiol.* **216**, 513–529.

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

VAN DER VELDEN, H. A. (1946). The number of quanta necessary for the perception of light in the human eye. *Ophthalmologica* **111**, 321–331.

WESTHEIMER, G. (1966). The Maxwellian View. *Vision Res.* **6**, 669–682.