Study of Absolute Visual Detection by the Rating-Scale Method*

JACOB NACHMIAS AND ROBERT M. STEINMAN
Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania
(Received 11 February 1963)

Dark-adapted observers were asked to use a 5-point rating scale in reporting whether they had seen a briefly exposed 1° square slightly off the fovea. The square, at constant luminance, was presented only on half the trials (*s* trials); nothing was presented on the remaining trials (*n* trials). The results were compared to two theories of absolute visual detection: (1) A two-state low-threshold theory according to which one fixed threshold exists, which, however, may be exceeded on an appreciable proportion of catch trials. (2) Statistical-decision theory, according to which (a) no fixed thresholds exist at all, but merely decision criteria that the observer can alter in accordance with instructions, and (b) the internal events arising on *s* and *n* trials may be described by two Gaussian distributions. The second theory received much stronger support from the data obtained in these experiments.

1. INTRODUCTION

EXPERIMENTAL determinations of the absolute visual threshold usually consist of finding a stimulus whose presence is detected by the observer on approximately 50% of its presentations (or some other percentage between 0% and 100%). Since investigators are interested in these thresholds usually only as indicators of the responsiveness of the visual system to antecedent or contemporary parameters of stimulation, they frequently do not state explicitly what they believe to be the processes underlying detection itself. Implicitly, however, two major assumptions often appear to be made.

(1) Somewhere before it reaches the "decision center," physico-chemical activity originating from sense organs encounters at least one transmission barrier. This activity must exceed some threshold magnitude if it is to pass the barrier and cause the observer to see a flash of light. Therefore, all stimulus presentations on which activity of subthreshold magnitude is initiated are indistinguishable. The fact that stimuli of some fixed nominal intensities are sometimes, rather than always or never, seen may mean that their intensity is unavoidably varying (quantal effects), or that their effect on the receptors is varying, or that the threshold magnitude is varying, or some combination of these factors.

(2) The threshold can be exceeded only on trials on which a stimulus is presented. On trials when no stimulus is presented in absolute detection experiments, either no spontaneous activity occurs, or if it does, its magnitude very seldom, if ever, exceeds threshold. In increment detection experiments, activity from the background is assumed never or very rarely to exceed threshold. False positives, therefore, are due to guessing. Catch trials are given to check that one is dealing with a "good" (nonguessing) observer, or in order to assess the guessing tendency of a not-so-"good" observer.

Although both assumptions are commonly made at the same time, they are clearly independent of each other; showing one to be wrong does not automatically invalidate the other. Consider for example the best known theories of absolute visual detection, that of Hecht, Shlaer, and Pirenne[1] and that of van der Velden.[2] The two assumptions are present in both theories which account for a variety of visual phenomena on the basis of quantal fluctuations alone. More recently Barlow[3] presented evidence against the second assumption from an experiment with the method of constant stimuli in which an observer was allowed to respond "possible" as well as "seen" and "not seen"; when "possibles" are added to "seens" the resultant false positive rate increases only slightly, whereas the resultant hit rate increases greatly for stimuli of medium intensity. If there were no spontaneous activity, "possible" responses would merely represent guesses and would be given most frequently on catch trials, where the opportunity to guess is greatest. The evidence from Barlow's experiment makes it highly unlikely that "spontaneous activity" is wholly absent, or never exceeds threshold magnitude. However, as becomes apparent in Sec. 3.1, this finding by itself does not warrant the further conclusion that no real thresholds exist at all, merely "threshold criteria" under the observer's control.

Beginning in 1953, a number of investigators[4–6] have been suggesting that thresholds, in the sense delineated in Assumption (1) above, either do not exist, or are so low as not to materially influence observers' behavior in detection experiments. They have extended the theory of statistical decision to cover detection of signals by human observers, and they have applied

* This investigation was supported by Research Grant B-3682 from the National Institute of Neurological Diseases and Blindness, Public Health Service.

[1] S. Hecht, S. Shlaer, and M. Pirenne, J. Gen. Physiol. **25**, 819 (1942).
[2] H. A. van der Velden, Ophthalmologica **111**, 321 (1946).
[3] H. B. Barlow, J. Opt. Soc. Am. **46**, 634 (1956).
[4] M. Smith and E. Wilson, Psychol. Monog. **67**, whole no. 359 (1953).
[5] W. W. Peterson, T. G. Birdsall, and W. C. Fox, IRE Trans. PGIT-4, 171 (1954).
[6] J. A. Swets, W. P. Tanner, and T. G. Birdsall, Psychol. Rev. **68**, 301 (1961).

this theory successfully to an impressive variety of data, mostly from experiments in hearing,[7] but also from experiments on brightness-increment detection.[6,8]

This "signal detectability theory" has been pitted by its proponents most frequently against a "high-threshold theory," one that embodies both Assumptions (1) and (2) above. Invariably, "high-threshold theory" has fared badly in these confrontations, regardless of whether the data come from experiments on the detection of sound or light increments,[6,9] on the intelligibility of speech, or on recognition memory.[10] The outcome is by no means so certain when the contender is a "low-threshold theory," one which concedes that spontaneous or background activity can cross threshold on an appreciable proportion of catch trials. For example, Luce[11,12] showed that results from some detection experiments can be accounted for about equally well by statistical-decision theory and low-threshold theory. The aim of the present investigation was to examine absolute visual detection in relation to these two theories.

2. METHOD

2.1 Apparatus

The stimulus object used in all experiments was a uniform white square, 1° on a side, which was presented briefly on a background that was completely dark except for a small, red fixation point, located $\frac{1}{2}°$ to the right of the lower right corner of the square. The optical system was Maxwellian, and of conventional design. Light came from an automobile headlight bulb (GE 1183) run at 4.7 V dc (5.4 A). The neutral density wedge and filters were calibrated *in situ* with a photomultiplier photometer (Photovolt 501 M with tube C). Absolute luminance calibration, which was performed several times in the course of the experiments, involved two steps: (1) photometry with a Macbeth Illuminometer of a large field, uniformly and diffusely lighted; (2) visual matching of a 1° square portion of this large field with the test square, when the two squares were juxtaposed optically and both viewed through a 1.9-mm artificial pupil.

The stimulus square was exposed for 52 msec usually, and for 102, 220, or 524 msec in a few experimental sessions. Sectored disks, driven at 4 rps by a Bodine synchronous motor, produced the two briefer exposures. A vane shutter mounted on a rotary solenoid was located optically in series with the sectored disk. The solenoid was activated by a circuit containing an external timer and microswitches operated by cams mounted on the shaft driving the sectored disk. In this way, single flashes could be delivered to the observer at predetermined intervals. The solenoid shutter alone, activated through a Hunter timer, produced the two longer exposure durations. The observer heard the double click of the solenoid, which served to define the exact temporal location of every trial. Before each "catch trial," the experimenter manually interposed a silent blanking shutter.

The observer steadied his head with the help of a biting board and looked through a 1.9-mm artificial pupil with the right eye. He and the experimenter were located in different rooms, but could speak to each other by means of a two-way intercom.

2.2 Observers

Two undergraduates at the University of Pennsylvania (S. L., A. H.) and the junior author (R. S.) were the observers in this study. None of them knew the purpose of the investigation when they made their observations. All of them had at least four practice sessions before the data to be reported were gathered. This paper is based on all of the data obtained after the practice sessions.

2.3 Procedure

In brief, the procedure was as follows: The observer dark adapted for at least 15 min. Thereafter, he heard a warning buzzer every 10 sec, followed in a couple of seconds by the double click of the solenoid to indicate the occurrence of a trial. The observer knew that on half the trials the square stimulus would be presented at fixed luminance and duration, and that on the other half nothing would be presented. He used a 5-point rating scale to report his judgment as to whether the stimulus had been presented on the just completed trial. Immediately thereafter, he heard a buzzer if in fact a stimulus had been presented.

s (stimulus) and n (no stimulus) trials occurred in a sequence that was random except for either of two types of restrictions: (1) for observers R. S. and S. L., every block of 32 trials contained an equal number of s trials and n trials. Cursory examination of their data fails to show any gross effect of this restriction. (2) For observer A. H., run lengths of only one to five trials appeared but their relative frequency of occurrence was the same as in a completely random sequence of 200 trials with $P(s)=P(n)=0.5$. The response to the trial following the completion of a run of five trials of a kind was discarded.

The rating scale has been used recently in a variety

---

[7] For references see W. P. Tanner and T. G. Birdsall, J. Acoust. Soc. Am. **30**, 922 (1958).

[8] W. P. Tanner and J. A. Swets, Psychol. Rev. **61**, 401 (1954).

[9] J. A. Swets, J. Acoust. Soc. Am. **31**, 511 (1959).

[10] J. P. Egan, Tech. Note, Contract No. AF 19(604)-1962 (1958).

[11] R. D. Luce, "Detection and Recognition," *Handbook of Mathematical Psychology*, edited by R. D. Luce, R. R. Bush, and E. Galanter (John Wiley & Sons, Inc., New York, 1963), Vol. 1.

[12] R. D. Luce, Psychol. Rev. **70**, 61 (1963).

of psychophysical experiments.[6,13,14] The usual practice in such experiments is to specify to the observer the relative frequencies with which the different response categories are to be used. In this study observers were merely asked to keep three rules in mind: (1) "Give '*5*'s' and '*4*'s' when you see something, '*2*'s' and '*1*'s' when you do not, and '*3*'s' when you are quite uncertain." (2) "Make fewer mistakes with '*5*'s' and '*1*'s' than with '*4*'s' and '*2*'s'; that is $P(\text{"}5\text{"}|n) < P(\text{"}4\text{"}|n)$ and $P(\text{"}1\text{"}|s) < P(\text{"}2\text{"}|s)$." (3) "Mistakes with '*5*'s' should be very few, if any."

Each experimental session consisted of approximately 400 trials and lasted $1\frac{1}{2}$ h. The 400 trials were divided by rest periods into 12 blocks of about 33 trials each. The stimuli in the first six blocks differed from those in the second six either in luminance (duration constant) or in duration and luminance (total energy constant). The stimulus conditions were distributed among experimental sessions in such a way that each condition was tested first and last in the same number of sessions. Before the first and sixth blocks of trials, the observer was allowed several "free" trials, before each of which he was told whether or not a stimulus would be presented, in order to acquaint him with the stimulus conditions to follow.

3. THEORETICAL CONSIDERATIONS

3.1 Two-State Threshold Theory

Luce[11,12] has formulated a theory for "yes–no" and "forced-choice" detection experiments which explicitly involves a threshold of the type considered in Assumption (1) of Sec. 1. His theory may be generalized to the rating scale experiment in the following manner.

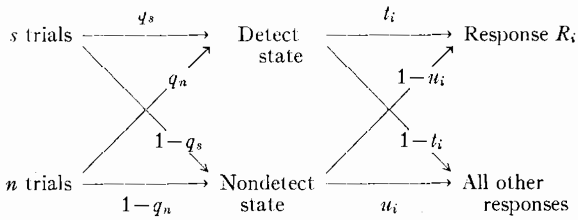
On each trial, either the threshold is exceeded, resulting in a "detect state," or it is not exceeded, resulting in a "nondetect state." Let q_s be the probability of the occurrence of a detect state on s trials with a stimulus of fixed characteristics, and let q_n be the corresponding probability on n trials. The values of these probabilities presumably depend on physico-chemical events, not directly under the observer's control. Furthermore, the value of q_n, unlike that of q_s, should not be sensitive to changes in the intensity of the stimulus presented on s trials, so long as successive trials are spaced far enough apart. [If Assumption (2) of Sec. 1 were correct, $q_n=0$.]

In contrast, the probability of a particular response, R_i, conditioned on the observer's state (detect or nondetect) can be affected by instructions, whether they be entirely verbal or involve a payoff matrix as well. Let t_i represent the conditional probability of response R_i given a detect state, and let $(1-u_i)$ represent the conditional probability of response R_i given a nondetect state. The probabilities of responding on s and n trials are then given by

$$P(R_i|s)=t_iq_s+(1-u_i)(1-q_s), \qquad (1)$$

$$P(R_i|n)=t_iq_n+(1-u_i)(1-q_n), \qquad (2)$$

where $P(R_i|s)$ is the conditional probability of R_i on s trials, $P(R_i|n)$ is the conditional probability of R_i on n trials, $\sum t_i=\sum u_i=1$, and $i=1, \cdots k$, where k is the number of alternative responses. The interrelation of theoretical constructs is also indicated in the diagram below:



In "yes–no" detection experiments, $R_i \in \mathbf{R}$= ("*yes*", "*no*")[15] and the experimenter attempts to induce different values of t_i and u_i in different experimental sessions by varying instructions with or without explicit "payoff matrices." In rating scale experiments, such as the one reported here, $R_i \in \mathbf{R}$= ("*1*", "*2*", etc.), and the experimenter attempts by means of instructions to establish simultaneously several t_i's and u_i's in one experimental session.

In applying two-state threshold theory to "yes–no" data, Luce implicitly makes an important simplifying assumption about the relation between values of t_i and u_i: there is some value t_i', $0<t_i'\leq 1$ such that for any conditions under which $t_i<t_i'$, $u_i=1$ and similarly, there is some value u_i', $0<u_i'\leq 1$, such that for any conditions under which $u_i<u_i'$, $t_i=0$. The analogous assumption for rating-scale data is that there exist two subsets of responses, $\mathbf{A}$ and $\mathbf{B}$, such that if $R_i \in \mathbf{A}$, then $u_i=1$ and if $R_i \in \mathbf{B}$, then $t_i=0$. In other words, some responses are given only when the threshold is exceeded, and some others are given only when it is not. In view of the instructions given to the observers in this experiment, one might expect that $\mathbf{A}$= ("*4*","*5*"), and $\mathbf{B}$= ("*1*","*2*"). Equations (1) and (2) then reduce to

$$P(R_i|s)=t_iq_s \quad P(R_i|n)=t_iq_n, \qquad (3)$$

if $R_i \in \mathbf{A}$, and to

$$P(R_i|s)=(1-u_i)(1-q_s), \; P(R_i|n)=(1-u_i)(1-q_n), \qquad (4)$$

if $R_i \in \mathbf{B}$.

[13] J. P. Egan, A. I. Schulman, and G. Z. Greenberg, J. Acoust. Soc. Am. **31**, 768 (1959).

[14] D. J. Weintraub and H. W. Hake, J. Opt. Soc. Am. **52**, 1179 (1962).

[15] This expression may be read as follows: R_i is a member of set $\mathbf{R}$, which consists of verbal responses "*yes*" and "*no*."

If this simplifying assumption is granted, then the threshold theory leads to three predictions which are testable by rating-scale data. The first of these predictions concerns the relative values of q_s and q_n. Whereas q_s, the probability of threshold being exceeded on s trials, should increase with stimulus luminance L the probability of threshold being exceeded on n trials, q_n, should be independent of L. These two quantities can be estimated by solving Eqs. (3) and (4), and according to the theory, they should behave as follows as L is varied:

Prediction 1.

$$q_s(L)=f(L) \quad q_n(L)=C.$$

The other predictions relate to the distribution of responses among s and n trials:

Prediction 2.

$$\frac{P(R_i|s)}{P(R_i|s)+P(R_i|n)}=\frac{q_s}{q_s+q_n}=C' \quad (R_i\in\mathbf{A}).$$

Prediction 3.

$$\frac{P(R_i|s)}{P(R_i|s)+P(R_i|n)}=\frac{1-q_s}{2-(q_s+q_n)}=C'' \quad (R_i\in\mathbf{B}).$$

When $P(s)=0.5$, the quantities on the left side of the above two equations are equivalent to the *a posteriori* probabilities of s given R_i, $P(s|R_i)$. Predictions 2 and 3 can therefore be restated as follows: If $R_i\in\mathbf{A}$, then $P(s|R_i)=C'$ and if $R_i\in\mathbf{B}$, then $P(s|R_i)=C''$.

3.2 Statistical-Decision Theory

Several extensive treatments of statistical-decision theory as applied to detection problems are already available,[4–8] so that it will be characterized here only briefly and informally. In the simplest, unidimensional case the theory contains the postulate that the internal events arising on s and n trials can be described by two continuous probability distributions $f_s(x)$ and $f_n(x)$. In "yes–no" detection experiments, the observer's task is to decide on each trial whether an observation x_i came from $f_s(x)$ or $f_n(x)$. He does this by establishing a criterion x_c such that he responds "yes" if and only if $x_i\geqq x_c$. The frequency of hits and false alarms associated with a given criterion depends on the shape and separation of the x distributions, and on the relative frequency of s and n trials. It is assumed in the theory that the observer is capable of varying his criterion over a considerable range in response to changes in the above factors, as well as to instructions about the relative importance of the two types of correct and incorrect decisions.

In a rating scale experiment with n response alternatives, the observer presumably divides the x continuum into n regions by establishing simultaneously $n-1$ criteria. His response indicates into which region a particular observation fell: that is, R_i is given when $c_i\leqq x\leqq c_{i+1}$. Consequently, to determine the frequency N of observations exceeding any one criterion, c_i, one merely adds the frequencies of observations falling in all regions to the right of c_i:

$$N(x\geq c_i)=\sum_i^n N(R_i) \quad i=2, 3\cdots n. \tag{5}$$

One way to make the theory mathematically tractable is to assume that $f_s(x)$ and $f_n(x)$ are well approximated by Gaussian distributions with means μ_s and μ_n, the former increasing monotonically with stimulus intensity. With this assumption, the theory leads to the prediction of a linear relationship between the z transforms of the probabilities that c_i is exceeded on s trials of fixed intensity, $z(c_i|s)$, and on n trials, $z(c_i|n)$, as c_i is varied:

$$z(c_i|s)=b[z(c_i|n)+d'], \tag{6}$$

where the slope $b=\sigma_n/\sigma_s$, the intercept $bd'=(\mu_s-\mu_n)/\sigma_s$ and, z, the normal deviate, is defined by

$$P=(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{z}\exp\left(\frac{u^2}{2}\right)du.$$

Such curves relating "hits" to "false alarms" are referred to as receiver operating characteristics (R.O.C. curves) or "isosensitivity" curves.[16]

The prediction of linearity can be tested by rating-scale data.

4. COMPARISON OF DATA WITH THEORY

4.1 Two-State Threshold Theory

Prediction 1.

Values of q_s and q_n were estimated from data obtained in experimental sessions in which luminance was changed (exposure duration 52 msec) after six blocks of trials. Since observers gave "5" and "1" responses very infrequently when the dimmer stimuli were being tested, "5's" and "4's" were considered

[16] The convention adopted here and in Ref. 6 is that $z=(\mu-c)/\sigma$, the higher the criterion the more negative the value of z. This convention is chosen so that the R.O.C. curves on conventionally labeled z axes will lie in the same quadrant as on conventionally labeled P axes.

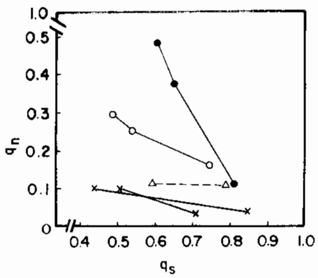


FIG. 1. Mean estimate of the probability of threshold being exceeded on n trials, q_n, as a function of the corresponding estimate on s-trials, q_s. Filled circles designate data from observer A.H., empty circles from S. L., crosses from R. S., and triangles from Monte Carlo experiments.

TABLE II. Monte Carlo experiments: assumed and means of estimated values of the parameters of two-state threshold theory.

	Experiment A		Experiment B	
	Assumed	Estimated	Assumed	Estimated
t_5+t_4	0.90		0.75	
t_1+t_2	0.00		0.00	
u_5+u_4	1.00		1.00	
u_1+u_2	0.20		0.20	
q_s	0.80	0.79	0.60	0.59
q_n	0.10	0.11	0.10	0.11

together, and so were "*2*'s" and "*1*'s" in calculating q_s and q_n from daily data. Means of these estimates are shown in Table I and Fig. 1 (solid lines). Contrary to Prediction 1 from threshold theory, q_n is neither zero nor some other constant independent of stimulus intensity. Actually it appears to vary inversely and linearly with q_s.

It is conceivable that even if q_n were actually independent of q_s, estimates of these quantities might be correlated due to some deficiency in the method of estimation from rating-scale data. This possibility was examined by generating artificial data by means of a Monte Carlo method.[17] Specifically, two sets of values were assumed for the parameters of threshold theory, as shown in Table II. Actually, all values except for q_n in Experiment B were close to estimates obtained from data of observer A. H.; q_n in Experiment B was assumed to be the same as in Experiment A. Then a table of random numbers was used to select responses with probabilities governed by these assumed parameter values. Experiments A and B were replicated nine times, with 200 trials in each replication. Average estimates of q_s and q_n from these artificial data also appear in Table II and Fig. 1 (dotted line). Note that the estimate of q_n appears to be independent of the value of q_s. Consequently it is unlikely that the correlation between estimates of q_s and q_n from real data is an artifact of the estimation procedure.

TABLE I. Mean estimates of q_s, the probability of threshold being exceeded on s trials, and q_n, the probability of threshold being exceeded on n trials.

Observer	No. of sessions	Luminance (log ml)	$\bar{q}_s$	$\bar{q}_n$
A. H.	4	−1.85	0.81	0.11
A. H.	4	−2.10	0.66	0.37
A. H.	4	−2.35	0.61	0.48
S. L.	6	−2.03	0.75	0.16
S. L.	6	−2.28	0.55	0.25
S. L.	6	−2.53	0.49	0.29
R. S.	2	−2.37	0.85	0.04
R. S.	2	−2.87	0.44	0.10
R. S.	2	−2.47	0.71	0.04
R. S.	2	−2.77	0.51	0.10

[17] S. Sternberg, "Stochastic Learning Theory," in Ref. 11, Vol. 2 (to be published).

Predictions 2 and 3

Table III contains mean estimates of *a posteriori* probabilities of s given R_i. Contrary to Prediction 2, $\bar{P}(s|\text{"5"})$ is invariably greater or equal to $\bar{P}(s|\text{"4"})$; and contrary to Prediction 3, $\bar{P}(s|\text{"2"})$ is usually greater than $\bar{P}(s|\text{"1"})$. The same pattern also emerges when the comparisons are made entirely within individual half-sessions (see Table IV). Here too, the evidence against Prediction 2 appears to be stronger than against Prediction 3.

Unfortunately, these estimates of *a posteriori* probability may be biased, especially since they are often based on small and unequal numbers of responses. Consequently the conclusions just drawn about Predictions 2 and 3 may be somewhat suspect. However, they receive support from an equivalent, unbiased test. Consider, for example, the subset of trials on which responses of either "*1*" or "*2*" were made. These trials may be classified in two ways: (a) on the basis of the response made "*1*" or "*2*" and (b) on the basis of the stimulus (present or absent). Prediction 3 implies that these two ways of classification are independent. This

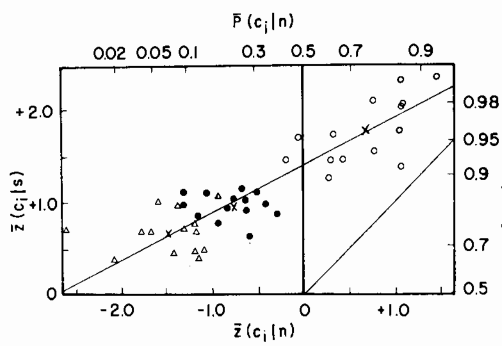


FIG. 2. $z(c_i|s)$, the z transform of the proportion of s trials on which criterion c_i was exceeded plotted against $z(c_i|n)$, the corresponding quantity on n trials. Observer A. H. Data from each of 14 half-sessions with the same stimulus energy are represented by open circles (c_2), filled circles (c_3), and triangles (c_4); means are represented by crosses.

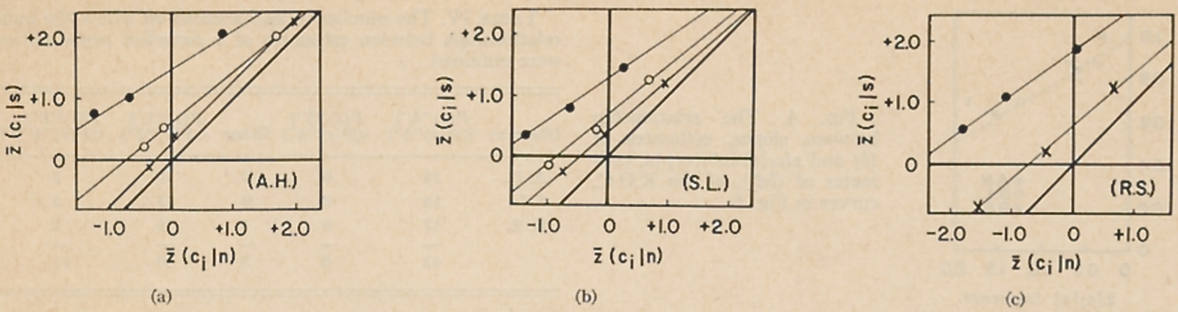


FIG. 3. R.O.C. curves on z-coordinates. (a) Observer A. H.; means of four half-sessions; filled circles -1.85 log ml, open circles -2.10 log ml, crosses -2.35 log ml. (b) Observer S. L.; means of six half-sessions; filled circles -2.03 log ml, open circles -2.28 log ml, crosses -2.53 log ml. (c) Observer R. S.; means of two half-sessions; circles -2.37 log ml, crosses -2.87 log ml.

hypothesis, as well as the corresponding one arising from Prediction 2, may be evaluated by Fisher's exact test[18] or by the χ^2 test for independence in contingency tables when appropriate. To perform these tests, trials were pooled from all experimental sessions separately for each observer. According to these tests, both hypotheses may be rejected for *each* observer ($p<0.001$).[19]

4.2 Statistical-Decision Theory

In Figs. 2 and 3 some of the data are represented in terms of statistical-decision theory: $z(c_i|s)$, the z transform of the proportion of s trials on which criterion c_i was exceeded is plotted against the corresponding quantity for n trials, $z(c_i|n)$; Fig. 2 indicates the extent of day-to-day variability in performance of one observer (A. H.) with one value of stimulus energy (-1.85 log ml at 52 msec, or -2.14 log ml at 102 msec). It contains values of $z(c_i|s)$ and $z(c_i|n)$ for $i=2, 3, 4$ obtained on each of 14 half-sessions as well as the means of these values. There are no points for c_5 because in 13 of these sessions a "5" response was never given on n trials. Twelve is the number of such sessions to be expected from a false alarm rate for "5's" of 0.002, which is the value calculated from $\bar{z}(c_5|s)$ and the line fitting the rest of the data. Only means are plotted in Fig. 3, based on sessions where only luminance was changed after six blocks of trials (exposure duration 52 msec).

The data points in Figs. 2 and 3 are well fitted by straight lines. The deviations are small and do not appear to be systematic. That is, the point for c_3 is as likely to be above as below the line joining the points for c_2 and c_4. This is true not only for the means, but also for data from individual half-sessions. The results of this experiment thus are in agreement with the prediction from statistical decision theory for Gaussian distributions, namely that R.O.C. curves are linear on z coordinates.

The slopes of the lines, which were fitted by eye to the means, are less than 1.0 in eight out of nine cases. Furthermore, the slopes of the lines (estimates of the values of b) decrease as their $\bar{z}(c_i|s)$ intercepts (estimates of the values of bd') decrease, as is shown in Fig. 4. If statistical-decision theory is correct, and if the underlying distributions are normal, this means that the standard deviations of the s distributions (i) are greater than the standard deviation of the n distribution, and (ii) increase with the means of the s distribu-

TABLE III. Mean estimates of $P(s|R_i)$, the *a posteriori* probability of s for the different response categories.

log ml	msec	No. of sessions	R_i "1"	"2"	"3"	"4"	"5"
			Observer A. H.				
−1.85	52	4	0.10	0.20	0.36	0.79	0.995
−2.10	52	4	0.36	0.35	0.52	0.60	0.97
−2.35	52	4	0.31	0.42	0.53	0.56	0.83
−1.85	52	5	0.19	0.19	0.34	0.82	1.00
−2.14	102	5	0.16	0.21	0.41	0.83	1.00
−2.43	220	2	0.21	0.23	0.46	0.75	1.00
−2.85	525	4	0.14	0.37	0.44	0.67	1.00
			Observer S. L.				
−2.03	52	6	0.16	0.31	0.46	0.77	0.995
−2.28	52	6	0.32	0.39	0.49	0.61	0.87
−2.53	52	6	0.41	0.42	0.49	0.59	0.95
			Observer R. S.				
−2.37	52	2	0.07	0.23	0.60	0.92	1.00
−2.47	52	4	0.20	0.28	0.57	0.93	1.00
−2.76	102	2	0.16	0.23	0.66	0.93	1.00
−2.67	52	2	0.37	0.38	0.49	0.76	1.00
−2.96	102	2	0.36	0.41	0.51	0.76	0.93
−2.77	52	4	0.36	0.35	0.48	0.81	1.00
−3.06	102	2	0.32	0.42	0.57	0.98	1.00
−2.87	52	2	0.31	0.42	0.61	0.80	1.00

[18] R. A. Fisher, *Statistical Methods for Research Workers* (Oliver and Boyd, Edinburgh, 1938), Chap. IV.

[19] Preliminary analysis of one experimental session indicates that successive responses made by observers are not independent; that is, they depend on factors other than the nature of the trial. If this response dependence could be reduced experimentally or corrected statistically, contingency tables of the type under consideration here would reveal no less dependence in classification than they now do.

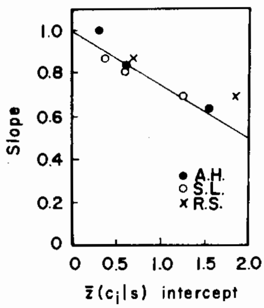


FIG. 4. The relationship between slopes, estimates of (b) and $z(c_i|s)$-intercepts, estimates of (bd'), of the R.O.C. curves in Fig. 3.

TABLE IV. The number of half-sessions on which the indicated relationships between estimates of *a posteriori* probabilities of s were obtained.

Observer	$P(s\|"4")<P(s\|"5")$	$P(s\|"5")<P(s\|"4")$	Other	$P(s\|"1")<P(s\|"2")$	$P(s\|"2")<P(s\|"1")$	Other
A. H.	28	0	0	19	7	2
S. L.	18	0	0	15	2	1
R. S.	17	0	3	16	2	2
	—	—	—	—	—	—
	63	0	3	50	11	5

tions (assuming that σ_n is independent of μ_s). Swets, Tanner, and Birdsall[6] noted such a dependence between variance and signal strength in their brightness increment experiments, and suggested that it can be described by the empirical equation, $\Delta\sigma_s=\frac{1}{4}\Delta\mu_s$. An equivalent expression is

$$1/b=1+0.25d'. \tag{7}$$

The line in Fig. 4 represents this equation; it fits the data from the present experiment reasonably well. This agreement may be fortuitous. As Swets, Tanner, and Birdsall[6] pointed out, the rate at which σ_s varies with μ_s might well be expected to depend critically on stimulus parameters.

Observers in rating-scale experiments usually have been required to use the different response categories with specified relative frequencies.[6,14] No such restrictions were employed in the present experiment, and in fact the relative frequency of each response category varied from observer to observer and from one luminance level to another. Figures 3 (a), (b), and (c) also reveal that a given criterion was not kept fixed with respect to the n distribution as luminance varied. For example, on the basis of estimates from his data it appears that Observer A. H. set c_4 at 1.29 σ_n above μ_n when flash luminance was -1.85 log ml, whereas he set it at 0.37 σ_n above μ_n when flash luminance was -2.35 log ml. The nature of this variation is revealed better in Fig. 5, in which $\bar{P}(c_i|s)$ is plotted against $\bar{P}(c_i|n)$. For each subject, the points for any criterion fall about a line originating in the upper left-hand corner. The equation for these lines is

$$P(c_i|n)=\beta[1-P(c_i|s)]. \tag{8}$$

It appears then that observers spontaneously alter each criterion so as to maintain a constant ratio between the probability that the criterion is exceeded on n trials and the probability that it is *not* exceeded on s trials. In other words, each criterion represents a particular balance between "false positives" and "misses."

5. DISCUSSION

As the analysis in the preceding section has demonstrated, statistical-decision theory can be applied quite successfully to absolute visual detection. The theory may provide powerful leverage on substantive problems such as the quantum efficiency of the visual system and the nature of visual noise, among others.

On the other hand, the two-state threshold theory described in Sec. 3.1 clearly cannot account for the results of these experiments. However, the possibility that fixed visual thresholds exist still cannot be ruled out altogether. In the first place, the difficulties of two-state threshold theory are due in large measure to the auxiliary assumption that observers can be induced by instructions to use some response categories only when threshold is exceeded and others only when it is not. Conceivably, the theory would fare much better if this particular assumption is abandoned in favor of

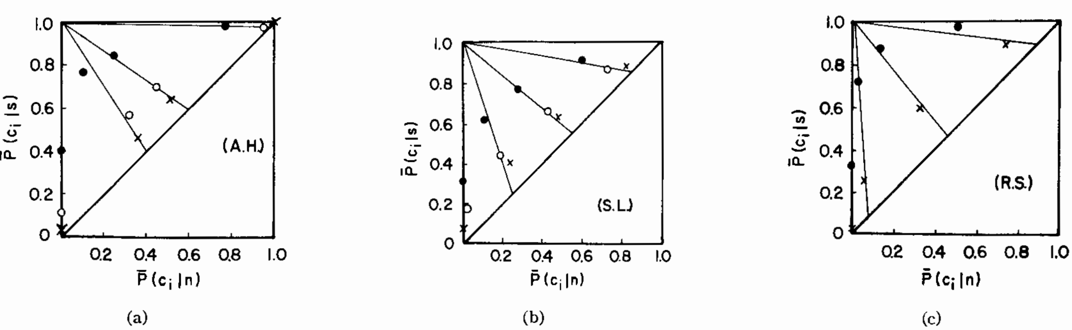


FIG. 5. Mean proportion of s trials on which criterion c_i was exceeded $\bar{P}(c_i|s)$, plotted against mean proportion of n trials on which c_i was exceeded, $\bar{P}(c_i|n)$. The symbols in this figure have the same significance as in Fig. 3.

some other kind of constraints on t_i and $(1-u_i)$, the conditional probabilities that the various response categories are used after detect and nondetect states. However, without some constraint on the behavior of t_i and u_i, the two-state threshold theory probably is untestable altogether by rating scale or "yes–no" experiments, because practically any conceivable outcome of such experiments could be accounted for by the theory.

The data here presented also do not contradict theories in which several fixed thresholds, hence more than two internal states, are postulated. Von Békésy's neural quantum theory of loudness-increment detection[20] is such a theory; its scope is greatly increased as a result of several modifications recently proposed by Larkin and Norman,[21] and Norman.[22] Norman's work also demonstrates the difficulties in distinguishing empirically between a theory with five states and statistical-decision theory, with its infinite number of states. But even if relatively few thresholds were involved in absolute visual detection, the lowest of these in all likelihood is exceeded on a large proportion of n trials.

---

[20] G. von Békésy, Ann. Physik **7**, 329 (1930); translated in G. von Békésy, *Experiments in Hearing* (McGraw-Hill Book Company, Inc., New York, 1960), pp. 329–359.

[21] W. D. Larkin and D. A. Norman, "An Extension and Experimental Analysis of the Neural Quantum Theory," in *Studies in Mathematical Psychology*, edited by R. C. Atkinson (Stanford University Press, Stanford, California, to be published).

[22] D. A. Norman, unpublished Ph.D. dissertation, University of Pennsylvania, 1962.