JOURNAL OF MATHEMATICAL PSYCHOLOGY 17, 220–235 (1978)

Pooling Peripheral Information: Averages Versus Extreme Values*

BRIAN WANDELL

University of California at Irvine, Irvine, California 92664

AND

R. DUNCAN LUCE

Harvard University, Cambridge, Massachusetts 02138

Suppose, as an idealization, that sensory intensity is coded in peripheral channels as identical Poisson pulse trains with intensity parameter a power function of signal intensity. Discrimination models based on either an average count computed over a fixed time or an average time computed for a fixed count per channel have difficulty in fitting the Weber function ($\Delta I/I$ versus I) if the free parameters are constrained to ranges determined from other experiments (magnitude estimation, reaction time). Here we study a different decision rule, namely, the most extreme observation in either the counting or timing mode. Our extremum-counting model, but not two timing ones, accounts very nicely for the Weber function. However, the ROC curves for these extremum models, which agree in shape with data of Green and Luce, yield estimates for the intensity parameter which are much larger than predicted by the power function growth used to calculate ΔI and about twice as large as those estimated from reaction time data collected in the same experiment.

Detection at absolute threshold is not governed by the activity of a single, peripheral neuron. For example, at a visual absolute threshold the amount of information available to any single rod is considerably less than the amount of information subjects display in their responses (Hecht, Schlaer, & Pirenne, 1942; Sakitt, 1972). The same is true of the peripheral auditory neurons (Siebert, 1968). Theorists generally account for this by assuming that information about these weak signals is accumulated from the activity of a set of units, not from some single, distinguished neuron, and subjects' responses arise in some way from a property—or decision statistic—derived from that accumulated information.

Most often the property assumed to be the decision statistic is a sum or average over the entire set of units. This means that some aspect of the responses of all units is added or averaged, and if the sum exceeds some criterion a detection is reported (Luce and Green, 1974; Sanderson, 1975; Siebert, 1968, 1970). The properties of summing or

* The first author was supported by a National Science Foundation predoctoral fellowship and the second author by a National Science Foundation Grant to Harvard University. Please send reprint requests to R. D. Luce, 33 Kirkland St., Cambridge, Mass. 02138.

0022-2496/78/0173–0220$02.00/0

averaging are widely understood (the Central Limit Theorem) as are the properties of the expected value operator which serves as averager in theories with a probabilistic orientation. This, no doubt, contributes to the popularity of this assumption.

An alternative assumption whose analytic properties, although also known, are generally less familiar is probability summation or, what is equivalent, decisions based on extreme observations. The simplest version assumes that the behavior of individual units, over trials, can be described by some common random variable, and the observer responds with a detection if the observation arising from at least one unit exceeds (or fails to exceed, depending on the statistic being used) a criterion value. Since some unit exceeds the criterion if and only if the unit with maximum response exceeds it, we must understand the behavior of the maximum of a set of random variables in order to derive the properties of probability summation.

A considerable amount of literature exists on the properties of extrema (Gumbel, 1958, provides a review; the fundamental papers are Fisher & Tippet, 1928; Gnedenko, 1941, 1943, 1953; von Mises, 1923, 1931). Here we summarize enough of that theory to permit us to calculate response probabilities and such derivative measures as intensity discrimination functions (Weber functions) and receiver operating characteristics (ROC curves) for a class of psychophysical models with which we have been working.

This class of psychophysical models postulates that estimates of the pulse rate on a set of channels are the underlying variable for detecting the presence or absence of a signal. The most natural statistics to use as estimates of the rate are: (1) the number of pulses (counts) that occur on a channel during a fixed time, and (2) the amount of time required in order to achieve a preassigned number of counts. Models of the former types are known as counting models (McGill, 1967) and those of the latter type, timing models (Luce & Green, 1972). In the previous work on these models information across channels was pooled via an averaging rule. Here we begin to work out some consequences of using the extreme value over the channels as the decision statistic.

In evaluating these models, there is some question of exactly which data to use. For example, Luce and Green (1974) report the results of six studies of $\Delta I/I$ versus I (the Weber function) for tone intensity, and there is considerable disagreement. Because of this, Jesteadt, Wier, and Green (1977) redid this work using a two-alternative forced-choice procedure, varying both I and f. Their results are substantially independent of f, and we use their data to evaluate our models.

Remarks on the Distribution of Extremes

As one samples an unbounded random variable, with cumulative distribution F, the observed maximum value increases without bound as the number of observations increases. This follows since for any finite value, x, the maximum will exceed x after J observations with probability $1 - F(x)^J$, which tends to unity as J increases. However, we may hope that at least for some F, as the number J of channels (fibers) grows, the shape of F^J will not change significantly. This is in analogy to the Central Limit Theorem where, despite the fact that the sum of a set of random variables increases without bound, the sum becomes approximately normally distributed. More precisely, by the shape remaining unchanged, we mean that there exists some (nontrivial) cumulative density, G, and normalizing constants, a_J and b_J, such that for large J,

$$F(x)^J = G(a_J x + b_J).$$

If such a G exists, it necessarily has the property that for a set of random variables with cumulative distribution G, the distribution of the maximum of that set is also distributed as G up to a transformation by the normalizing constants. That is,

$$G(x)^J = G(a_J x + b_J).$$

This functional equation was first used by Fisher and Tippet (1928) to solve for extreme value distributions, and it is now called the stability postulate. There are precisely three distribution functions that solve this equation. Gnedenko (1943) provided necessary and sufficient conditions on the underlying distribution, F, so that one may determine which, if any, of these three distributions is the limiting one for that F.

For the class of distributions with exponential upper tails, which includes the normal, the gamma, the exponential, and the log normal, the asymptotic distribution G is

$$e^{-e^{-\alpha(x-\beta)}}, \tag{1}$$

which is called the double exponential or extreme value distribution.[1]

For distributions that are high tailed, such as the Cauchy, the limiting distribution is

$$e^{-(v/x)^k}. \tag{2}$$

And for random variables that are bounded from above, the maximum is distributed as

$$e^{-[(\omega-x)/(\omega-\zeta)]^k}, \tag{3}$$

where ω is the upper bound, $\omega - \zeta > 0$, and $k > 0$.

In each case the various parameters of these distributions depend both on the sample size, J, and on the underlying random variables. For the psychophysical models we investigate here we assume that the sample size does not vary, but the parameters of the underlying random variables do. Three cases interest us for applications, and for these cases we now compute the dependence of the mean and standard deviation of the extreme statistics on the parameters of the underlying random variables.

[1] Yellott (1975) and McFadden (1974) have shown that if random variables associated with stimuli are distributed as the double exponential, with α constant and β dependent on the stimulus, and if choices from sets of stimuli are made according to which stimulus produces the largest observation, then the choice probabilities satisfy Luce's (1959) choice axiom. Moreover, and this is the difficult part to show, if the choice axiom holds and is generated by such a Thurstone mechanism with a shift family of distributions, then they must be distributed according to the double exponential.

The Maximum of Normals

Let $X_i (i = 1, \ldots, J)$ be independent and normally distributed with mean μ and variance σ^2. Then,

$$P[\max_{1 \leqslant i \leqslant J} X_i < z] = \left[\int_0^z \mathcal{N}(\mu, \sigma^2)\right]^J = \left[\int_0^{z-\mu/\sigma} \mathcal{N}(0, 1)\right]^J.$$

Since the maximum of a set of unit normals is distributed asymptotically as the double exponential we have (Eq. (1))

$$P[\max_{1 \leqslant i \leqslant J} X_i < z] \cong e^{-e^{-\alpha_J[(z-\mu)/\sigma - \beta_J]}}.$$

The mean and standard deviation of this distribution can be shown to be (Gumbel, 1958, pp. 173–174)

$$E[\max_{1 \leqslant i \leqslant J} X_i] \cong \mu + \sigma\left(\beta_J + \frac{C}{\alpha_J}\right),$$

$$V^{1/2}[\max_{1 \leqslant i \leqslant J} X_i] \cong \sigma \frac{\pi}{\alpha_J 6^{1/2}}, \tag{4}$$

where C is Euler's constant ($=0.5772$) and the parameters α_J and β_J are functions of the sample size, J. Graphs of them are provided in Fig. 1 (Gumbel, 1958, p. 134).

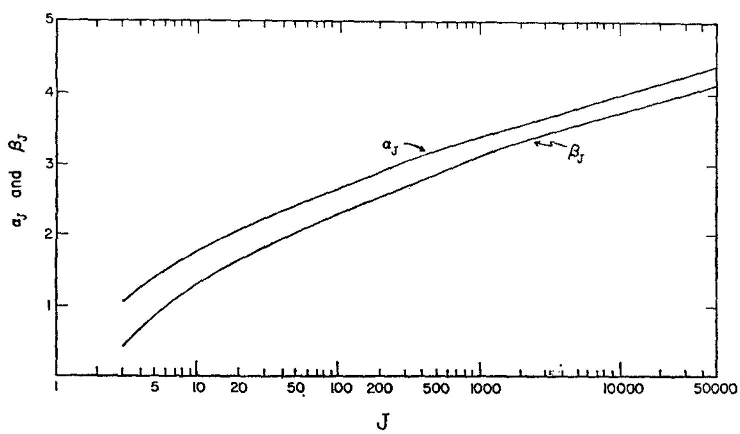

FIG. 1. The parameters α_J and the characteristic largest value, β_J, versus J for the asymptotic double exponential distribution arising from the maximum of J random variables that are distributed as the normal.

The Maximum of Gammas

To compute the maximum of a set of independent, identically gamma distributed random variables, Z_i, we first note that for a set of gammas with a rate parameter of one, that is

$$F_1(z) = \int_0^z \frac{x^{k-1}}{(k-1)!} e^{-x}\, dx,$$

the max Z_i is distributed asymptotically as

$$e^{-e^{-\alpha_{J,k}(z-\beta_{J,k})}},$$

where J is the sample size and k is the order parameter. If the rate parameter differs from one we may write

$$F_\mu(z) = \int_0^z \frac{\mu^k x^{k-1}}{(k-1)!} e^{-\mu x} = \int_0^{\mu z} \frac{x^{k-1}}{(k-1)!} e^{-x}\, dx = F_1(\mu z).$$

Therefore the maximum with an arbitrary rate parameter is distributed as

$$e^{-e^{-\alpha_{J,k}(\mu z-\beta_{J,k})}}.$$

The mean and standard deviation for this distribution are

$$E[\max_{1\leqslant i\leqslant J} Z_i] \cong \frac{1}{\mu}\left(\beta_{J,k} + \frac{C}{\alpha_{J,k}}\right),$$
$$V^{1/2}[\max_{1\leqslant i\leqslant J} Z_i] \cong \frac{1}{\mu}\frac{\pi}{\alpha_{J,k}6^{1/2}}. \tag{5}$$

(Note, in the normal approximation to the gamma, $\beta_J = \beta_{J,k} - 1$.) A graph of $\beta_{J,k}$ is given in Fig. 2 and of $\alpha_{J,k}\beta_{J,k}$, which is the parameter that arises, in Fig. 3 (Gumbel, 1958, pp. 144–145).

The Minimum of Gammas

Let Y_i be gamma distributed with rate parameter one and order parameter k. Notice that

$$P[\min_{1\leqslant i\leqslant J} Y_i < z] = 1 - P\left[\max_{1\leqslant i\leqslant J} \frac{1}{Y_i} < \frac{1}{z}\right]. \tag{6}$$

These random variables have the distribution function

$$P\left[\frac{1}{Y_i} < z\right] = P\left[Y_i > \frac{1}{z}\right] = \int_{1/z}^{\infty} \frac{x^{k-1}}{(k-1)!} e^{-x}\, dx.$$

Making the change of variable $y = 1/x$ yields

$$P\left[\frac{1}{Y_i} < z\right] = \int_0^z \frac{1}{y^{k+1}(k-1)!} e^{-1/y}\, dy.$$

Therefore the random variables $1/Y_i$ possess high-tailed densities whose maximum converges to the distribution in Eq. (2) (Gumbel, 1958, p. 259). The defining characteristic of these densities is that for some $k > 0$ and $A > 0$

$$\lim_{z\to\infty} z^k f(z) = A.$$

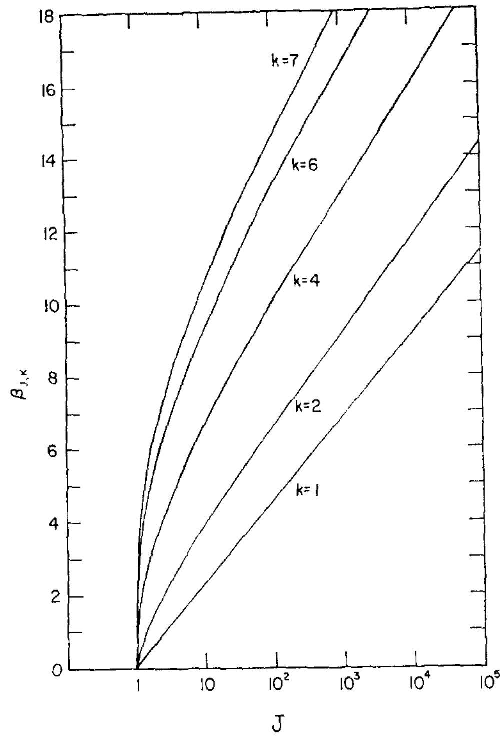

FIG. 2. The characteristic largest value, $\beta_{J,k}$, versus J for the asymptotic double exponential distribution arising from the maximum of J random variables that are distributed as a gamma of order k and intensity 1.

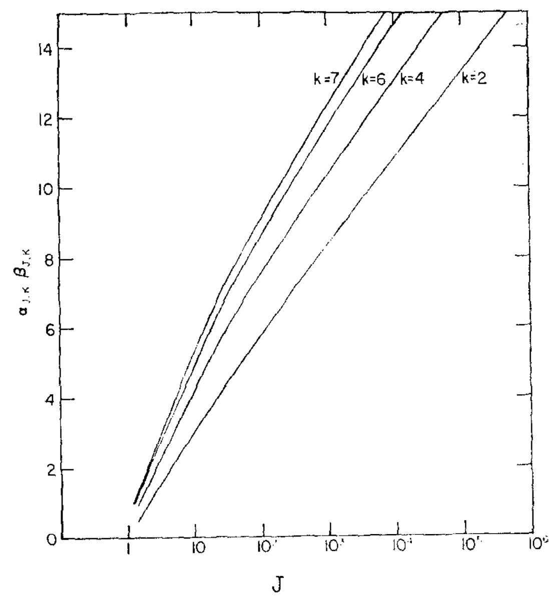

FIG. 3. The parameter $\alpha_{J,k}\beta_{J,k}$ versus J for the asymptotic double exponential distribution arising from the maximum of J random variables that are distributed as a gamma of order k and intensity 1.

Since the density of $1/Y_i$ has this property we may conclude (Gumbel, 1958, pp. 259–260)

$$P[\max_{1\leqslant i\leqslant J} 1/Y_i < x] \cong e^{-(v_{J,k}/x)^k}.$$

Substituting this into Eq. (6) yields

$$P[\min_{1\leqslant i\leqslant J} Y_i < z] \cong 1 - e^{-(v_{J,k}z)^k}.$$

This distribution function, sometimes called the Weibull, has recently been proposed for the psychometric function (Graham & Rogowitz, 1976; Green & Luce, 1975; Quick, 1974).

The effect of introducing a nonunit rate parameter, μ, can be calculated by a simple change of variable procedure identical to the first two instances. The mean and standard deviation for the arbitrary rate parameter are simply

$$\begin{aligned} E[\min_{1\leqslant i\leqslant J} Y_i] &\cong \frac{1}{\mu v_{J,k}}\, \Gamma\left(1+\frac{1}{k}\right), \\ V^{1/2}[\min_{1\leqslant i\leqslant J} Y_i] &\cong \frac{1}{\mu v_{J,k}}\left[\Gamma\left(3+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right)\right]. \end{aligned} \tag{7}$$

As we shall see, the graph of $v_{J,k}$ is not needed for our purposes.

Common Assumptions of the Models

We can now use the calculations of the previous section to compute the Weber functions for the timing and counting models based on probability summation. These models have been described in detail elsewhere (McGill, 1967, Luce & Green, 1972, Wandell, 1977); however, since the presentations have varied in some of the details of the assumptions, we list here the exact assumptions which we make for the present calculations:

(1) A physical stimulus elicits a train of pulses on a set of J channels.

(2) The pulse train on each channel is a Poisson process. The processes on different channels are independent, but statistically identical.

(3) The rate parameter of the common Poisson process is a power function of the stimulus intensity with an exponent of γ.

These hypothetical channels obviously bear some resemblance to peripheral auditory neurons. But in three important respects they differ. First, the channels do not saturate over a relatively narrow (20–30 dB) range, whereas peripheral neurons do. Second, those channels activated by a signal are all statistically identical, whereas pheripheral neurons are not. And third, the number of active channels is assumed to be constant with intensity, whereas the number of active peripheral neurons increases markedly with intensity. It is clear, then, that our channels are not meant to model peripheral neurons; rather one can think of them as hypothetical structures which either are some grouping of peripheral neurons or are later in the central nervous system. Because few quantitative studies have been made of the behavior of auditory neurons in the central nervous system and, more importantly, because the class of neurons which forms the basis of the behavioral response is unknown, physiological data are only of limited value in guiding our choice of assumptions. Thus, the models we investigate are genuinely *psychophysical* models of auditory detection. Physiological ideas have certainly influenced—but not dictated—their form.

Weber Functions Based on Observations of Maxima

Timing

If the underlying process on a channel is (well approximated by) a Poisson process (see Luce & Green, 1974), then the distribution of times to collect k counts is (approximately) gamma with order parameter k. One strategy a subject might adopt is to collect these counts, and if the maximum time exceeds a criterion, then respond "No." To derive an expression of $\Delta I/I$ under that strategy, we proceed as follows.

First, we calculate the expression for the usual d' of a "Yes–No" experiment.

$$d' = \frac{E[\max X_i \mid n] - E[\max X_i \mid s]}{V^{1/2}[\max X_i \mid n]}$$

$$= \frac{(1/\mu(n) - 1/\mu(s))(\beta_{J,k} + C/\alpha_{J,k})}{(1/\mu(n))(\pi/\alpha_{J,k}6^{1/2})}$$

$$= \left(1 - \frac{\mu(n)}{\mu(s)}\right)\left(\beta_{J,k} + \frac{C}{\alpha_{J,k}}\right)\Big/\frac{\pi}{\alpha_{J,k}6^{1/2}}.$$

(Note: the order of the expectations is such that d' is positive.)

Second, we assume, as we have been led to in other studies, that μ grows as a power function of intensity so that

$$1 - \frac{\mu(n)}{\mu(s)} = 1 - \left(1 + \frac{\Delta I}{I}\right)^{-\gamma}. \tag{8}$$

Third, we set $d' = 1$, which corresponds closely to 75% correct responses in a two-alternative forced-choice procedure. By algebraic manipulation we find

$$\frac{\Delta I}{I} = \left(1 - \frac{\pi/6^{1/2}}{\alpha_{J,k}\beta_{J,k} + C}\right)^{-1/\gamma} - 1. \tag{9}$$

If we assume that the number of active channels, J, and the order parameter, k, are independent of tone intensity, then Eq. (9) says that $\Delta I/I$ is a constant (Weber's Law) which is empirically false. So, if this model is to be made viable we must assume one or both of these parameters vary with intensity.

If we allow these parameters to vary, this will be reflected in a change of the value of $\alpha_{J,k}\beta_{J,k}$ in Eq. (9). We plot $\beta_{J,k}$ and $\alpha_{J,k}\beta_{J,k}$ in Figs. 2 and 3, respectively, for small k. To fit the data of Jesteadt *et al.* (1976) at low intensities, where $\Delta I/I \cong 0.5$, $\alpha_{J,k}\beta_{J,k}$ must be about 50. Extrapolating in Fig. 3 we see that for small k, $\alpha_{J,k}\beta_{J,k}$ equals 50 for J considerably greater than 10^5. Since there are considerably fewer than 10^5 peripheral auditory nerve fibers, the model for small k is not acceptable.

For large k we approximate the gamma by the normal with mean k/μ and standard deviation $k^{1/2}\mu$. Then

$$d' = \left(1 - \frac{\mu(n)}{\mu(s)}\right) \frac{k + k^{1/2}(\alpha_J + C/\beta_J)}{\pi/\alpha_J 6^{1/2}},$$

where α_J and β_J are plotted in Fig. 1. Thus

$$\frac{\Delta I}{I} \cong \left(1 - \frac{\pi/\alpha_{J(6^{1/2})}}{k + k^{1/2}(\alpha_J + C/\beta_J)}\right)^{-1/\gamma} - 1.$$

Table 1 gives the values of J and k needed for this model to fit the data of Jesteadt *et al.* Again, the data at low intensity require the range of J to be unreasonably large.

TABLE 1

$\Delta I/I$ for the Timing Model Using Normal Distributions

J	$\gamma = 0.3$, $k = 40$	$\gamma = 0.4$, $k = 10$
5	0.536	0.588
1000	0.144	0.167
30,000	0.103	0.111

Counting

To carry out the analogous calculation for the counting model, we make use of the fact that in any renewal process (and the Poisson in particular) the number of counts observed in time δ, $N(\delta)$, is approximately normal with

$$E(N(\delta)] = \delta/E(T),$$
$$V[N(\delta)] = \delta V(T)/E(T)^3,$$

where T denotes the time between successive pulses. Since

$$E(T) = 1/\mu \quad \text{and} \quad V(T) = (1/\mu)^2,$$

then

$$E(N) = \delta\mu \quad \text{and} \quad V(N) = \delta\mu = E(N).$$

Proceeding as before

$$d' = \frac{\delta[\mu(s) - \mu(n)] + \delta^{1/2}[\mu(s)^{1/2} - \mu(n)^{1/2}](\beta_J + C/\alpha_J)}{\delta^{1/2}\mu(n)^{1/2}(\pi/\alpha_J 6^{1/2})}.$$

Assuming again μ is a power function of I, say $\mu = \rho I^\gamma$

$$d' = \frac{(\rho\delta)^{1/2}[(1 + \Delta I/I)^\gamma - 1]I^{\gamma/2} + [(1 + \Delta I/I)^{\gamma/2} - 1](\beta_J + C/\alpha_J)}{\pi/\alpha_J 6^{1/2}}.$$

Setting $d' = 1$ and solving the resulting quadratic yields

$$\frac{\Delta I}{I} = \left\{\left[\left(\frac{\beta_J + C/\alpha_J}{2(\rho\delta)^{1/2}I^{\gamma/2}}\right)^2 + \frac{\beta_J + C/\alpha_J + \pi/\alpha_J 6^{1/2}}{(\rho\delta)^{1/2}I^{\gamma/2}} + 1\right]^{1/2} - \frac{\beta_J + C/\alpha_J}{2(\rho\delta)^{1/2}I^{\gamma/2}}\right\}^{2/\gamma} - 1. \quad (10)$$

This equation is readily fit to the data, even assuming that J is independent of intensity. One fit to the data of Jesteadt *et al.* (1977) is shown in Fig. 4. It corresponds to a multitude of possible parameter values, two sets of which are shown in the figure caption. It is obviously a very good fit.

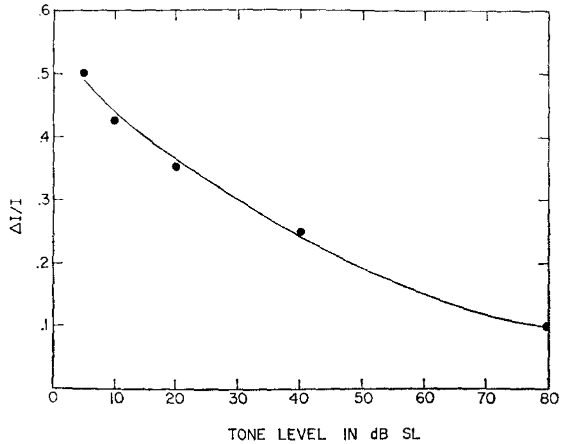

FIG. 4. $\Delta I/I$ versus I for the counting model (Eq. (10)); the data are those of Jesteadt *et al.* (1977). Various sets of parameters will generate the theoretical curve, e.g., $J = 1000$, $\gamma = 0.22$, $\rho\delta = 5$ and $J = 100$, $\gamma = 0.20$, $\rho\delta = 13$.

WEBER FUNCTIONS BASED ON OBSERVATIONS OF MINIMA

Timing

For the timing model an alternative decision rule is for the subject to respond "Yes" when the time to collect k counts on any channel is less than some criterion. Thus the probability of a detection response is the probability that the minimum of a set of gamma distributed random variables is smaller than some criterion.

Proceeding as in the previous two cases and substituting the appropriate expressions for the mean and standard deviation yields

$$d' = \frac{(1/\mu(n) - 1/\mu(s))\, \Gamma(1 + 1/k)/v_{J,k}}{(1/\mu(n))(1/v_{J,k})[\Gamma(1 + 2/k) - \Gamma^2(1 + 1/k)]}$$

$$= \left(1 - \frac{\mu(n)}{\mu(s)}\right) \frac{\Gamma(1 + 1/k)}{\Gamma(1 + 2/k) - \Gamma^2(1 + 1/k)}.$$

Substituting Eq. (8) and setting $d' = 1$ results in

$$\frac{\Delta I}{I} = \left(1 - \frac{\Gamma(1 + 2/k)}{\Gamma(1 + 1/k)} + \Gamma(1 + 1/k)\right)^{-1/\gamma} - 1.$$

Not only does this imply Weber's law for fixed J and k, as in the maximum timing model, but worse yet $\Delta I/I$ is wholly independent of J and so there is no reason for multiplying the channels. Nonetheless, the value is in the correct range: with $\gamma = 0.3$, $\Delta I/I = 1.520$ and 0.047 for $k = 2$ and 10, respectively.

Discussion

Of the three models we have examined, only the one based on the maximum number of counts provides an adequate description of the Weber function for tone intensity discrimination. This is enouraging since we have argued elsewhere (Green & Luce, 1973, Wandell, 1977) that subjects can choose among timing and counting strategies and that counting is the appropriate strategy for psychophysical experiments using stimuli all of the same duration.

The problem remains to choose between the counting model we have described here (extremum-counting model) and the usual counting model based on averaging (averaging–counting model).

The Coefficient of Variation

As Jesteadt *et al.* (1977) point out, the averaging–counting model does not describe the Weber function data very well; therefore, relative to those data the extremum-counting model seems preferable.

One way to see why the extremum-counting model accounts for the data better than does the averaging–counting model is to examine the coefficient of variation—standard deviation divided by mean—of the decision variable. This quantity (in logarithmic scale) is plotted in Fig. 5 for each model as a function of the number of channels (in logarithmic scale), with signal intensity as a parameter. (Several other assumed constants are shown in the caption.) Although the quality of the decision statistic is comparable for a small number of channels, the extremum model is distinctly poorer for a large number of channels; and more to the point, for a fixed number of channels the quality

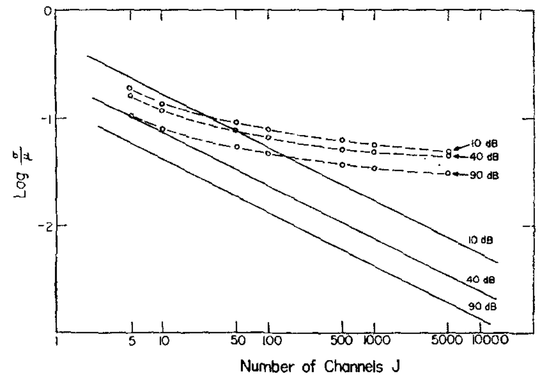

FIG. 5. A plot of the log of the coefficient of variation as a function of the log of channel size. For the averaging–counting model (solid lines) the relationship is $\log \sigma/\mu = -\frac{1}{2} \log \rho(I/I_0)\gamma$ and for the extrema-counting model (dashed lines) $\log \sigma/\mu = \log \pi/6^{1/2} - \log[\alpha_J(\rho(I/I_0)^\gamma)^{1/2} + \alpha_J\beta_J + C]$. The plots assume $\rho = 2$ and $\gamma = 0.25$.

improves with intensity less rapidly for the extremum model than it does for the averaging model. It is this reduced sensitivity to changes in intensity that makes it easier for the extremum model to account for the Weber function data than the averaging model.

Speed–Accuracy Tradeoff

A second type of data for which both models make predictions are reaction time data from speed–accuracy tradeoff experiments. However, the models make the same predictions for those data and therefore cannot be distinguished by those experiments. They both account for certain kinds of reaction time data quite well (Green and Luce, 1973; Wandell, 1977).

ROC Curves

The third type of data which can be used to compare the models are the receiver operating characteristic (ROC) generated in response-biasing experiments. Using the expression derived for the maximum of normal random variables and eliminating the criterion value from the expressions for $P(Y \mid s)$ and $P(Y \mid n)$, we obtain the following expression for the ROC of the extremum-counting model:

$$\ln \ln \frac{1}{1 - P(Y \mid s)} \cong \left(\frac{\mu(n)}{\mu(s)}\right)^{1/2} \ln \ln \frac{1}{1 - P(Y \mid n)}$$

$$+ \alpha_J \left\{\beta_J \left[1 - \left(\frac{\mu(n)}{\mu(s)}\right)^{1/2}\right] + \delta^{1/2} \frac{\mu(s) - (n)}{\mu(s)^{1/2}}\right\}. \qquad (11)$$

From Eq. (11) we see that when we plot the empirical ROC points in coordinates of $\ln \ln(1/(1 - p))$, the theory predicts the data points will fall along a straight line with slope $[\mu(n)/\mu(s)]^{1/2}$.

For comparison, the ROC for the averaging-counting model (Green and Luce, 1973, p. 152, Eq. (1)) is:

$$z(Y \mid s) \simeq \left(\frac{\mu(n)}{\mu(s)}\right)^{1/2} z(Y \mid n) + \delta^{1/2} \frac{\mu(s) - \mu(n)}{\mu(s)^{1/2}}, \tag{12}$$

where $z(Y \mid s)$ and $z(Y \mid n)$ are the standard normal deviates (z-scores) of $P(Y \mid s)$ and $P(Y \mid n)$, respectively. This model predicts that the data will fall along a straight line when plotted on double probability paper. The slope, again, is predicted to be $[\mu(n)/\mu(s)]^{1/2}$.

Figure 6 shows the data collected by Green and Luce replotted in the coordinates

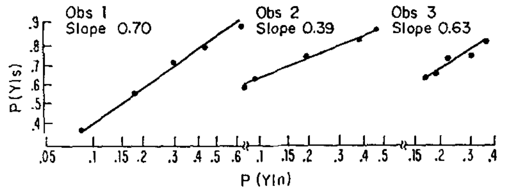

FIG. 6. $P(Y \mid s)$ versus $P(Y \mid n)$ in $\ln \ln(1/(1 - p))$ coordinates for detection of response terminated pure tone signals in noise under a 500-msec deadline on all trials. According to the extremum-counting model, these data should line on a line with slope $[\mu(n)/\mu(s)]^{1/2} < 1$.

$\ln \ln(1/(1 - p))$. The linearity is satisfactory. In the original paper Green and Luce (Fig. 4) show that the data also are satisfactorily fit by a line when plotted in double probability coordinates. What differs, however, are the estimates of $\mu(n)/\mu(s)$. These are much larger, as shown in Table 2.

TABLE 2

Estimated $\mu(s)/\mu(n)$

	Observer		
	1	2	3
Averaging	1.19	2.13	1.25
Extremum	2.04	6.57	2.52

The large numbers predicted by the extremum-counting model are difficult to reconcile with the power function relation of Eq. (8). This can be seen in Table 3 where we have computed the value of $\mu(s)/\mu(n)$ one would expect from various values of $\Delta I/I$ and γ.

TABLE 3

Predicted $\mu(s)/\mu(n)$

	$\Delta I/I$				
γ	0.1	0.3	0.5	1	2
0.2	1.02	1.05	1.08	1.15	1.25
0.5	1.05	1.14	1.22	1.41	1.73
1	1.10	1.30	1.50	2.00	3.00

The extremum-timing model can be more sharply rejected. Using the double exponential derived for the underlying gamma distribution of the extremum-timing model, one calculates

$$-\ln \ln \frac{1}{P(Y \mid s)} \cong -\frac{\mu(s)}{\mu(n)} \ln \ln \frac{1}{P(Y \mid n)} + \alpha_{J,K}\beta_{J,K}\left(\frac{\mu(s)}{\mu(n)} - 1\right). \qquad (13)$$

The corresponding averaging equation (Eq. (7) of Green and Luce) is

$$z(Y \mid s) \cong \frac{\mu(s)}{\mu(n)} z(Y \mid n) + (Jk)^{1/2}\left(\frac{\mu(s)}{\mu(n)} - 1\right).$$

Figure 7 shows their data from the experimental condition for which timing is anticipated, plotted in the coordinates $-\ln \ln(1/p)$. Not only are the slopes too large to be

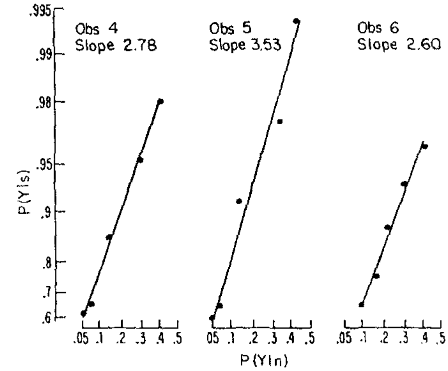

FIG. 7. $P(Y \mid s)$ versus $P(Y \mid n)$ in $-\ln(1/p)$ coordinates for detection of response terminated pure tone signals in noise under a 500-msec deadline on signal trials only. According to the extremum-timing model, these data should line on a line with slope $\mu(s)/\mu(n) > 1$.

compatible with Eq. (8), but in addition—as Green and Luce pointed out (Eq. (5))—the timing model predicts a linear relation between noise and signal mean reaction times

$$MRT_n = \frac{\mu(s)}{\mu(n)} MRT_s + \bar{r}\left(1 - \frac{\mu(s)}{\mu(n)}\right),$$

which affords an independent estimate of $\mu(s)/\mu(n)$. The reaction time estimates are very close to the ROC estimates of the averaging–timing model and quite different from those of the extremum-timing model.

We are left with the following unsatisfactory state of affairs. Green and Luce's data for the condition where the general timing model appears to be correct establishes that the extremum-timing model fails because the estimates of $\mu(s)/\mu(n)$ are so different from the ROC and the reaction time data. These predictions do not require any assumption about the dependence of μ on I. However, the $\Delta I/I$ versus I data favor the extremum-counting model over the averaging–counting model when μ is assumed to be a power function of intensity. Together, these observations suggest that the simple power function assumption may be the source of the trouble.

Acknowledgments

For their comments on an earlier draft we thank N. Graham, D. M. Green, W. Jesteadt, B. Rogowitz, B. Sakitt, and C. C. Wier.

References

Fisher, R. A., & Tippett, L. H. C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 1928, **24**, 180.

Gnedenko, B. V. Limit theorems for the maximal terms of a variational series. *Doklady Akademii Nauk SSR*, 1941, **32**, 37 (Russian).

Gnedenko, B. V. Sur la distribution limite du terme maximum d'une serie aléatoire. *Annals of Mathematics*, 1943, **44**, 423.

Gnedenko, B. V. On the role of the maximal summand in the summation of independent random variables. *Ukrainskii Matematičeskiĭ Zurnal*, 1953, **5**, 291.

Graham, N., & Rogowitz, B. E. Spatial pooling properties deduced from the detectability of FM and quasi-AM gratings: A reanalysis. *Vision Research*, 1976, **16**, 1021.

Green, D. M., & Luce, R. D. Speed–accuracy trade-off in auditory detection. In S. Kornblum (Ed.) *Attention and performance IV*. New York: Academic Press, 1973. Pp. 547–569.

Green, D. M., & Luce, R. D. Parallel psychometric functions from a set of independent detectors. *Psychological Review*, 1975, **82**, 483.

Gumbel, E. J. *Statistics of extremes*. New York: Columbia Univ. Press, 1958.

Hecht, S., Schlaer, S., & Pirenne, M. H. Energy, quanta, and vision. *Journal of General Physiology*, 1942, **25**, 819.

Jesteadt, W., Wier, C. C., & Green, D. M. Intensity discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America*, 1977, **61**, 169–177.

Luce, R. D. *Individual choice behavior*. New York: Wiley, 1959.

Luce, R. D., & Green, D. M. A neural timing theory for response times and the psychophysics of intensity. *Psychological Review*, 1972, **79**, 14.

Luce, R. D., & Green, D. M. Neural coding and psychophysical discrimination data. *Journal of the Acoustical Society of America*, 1974, **56**, 1554.

McFadden, D. Conditional logit analysis of quantitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics*. New York: Academic Press, 1974. Pp. 105–142.

McGill, W. J. Neural counting mechanisms and energy detection in audition. *Journal of Mathematical Psychology*, 1967, **4**, 351.

Mises, R. von Uber die Variationsbreite einer Beobachtungsreihe, Sitzungsber. *Berliner Math. Ges.*, 1923, **22**, 3.

Mises, R. von Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik. Leipzig, Deutsche, 1931.

Quick, R. F. A vector-magnitude model of contrast detection. *Kybernetik*, 1974, **16**, 65.

Sakitt, B. Counting every quantum. *Journal of Physiology*, 1972, **223**, 131.

Sanderson, A. C. Discrimination of neural coding parameters in the auditory system. *IEEE Transactions on Systems, Man, and Cybernetics*, 1975, **SMC-5**, 533.

SIEBERT, W. M. Stimulus transformations in the peripheral auditory system. In P. A. Kolers and M. Eden (Eds.), *Recognizing patterns*. Cambridge, Mass.: M.I.T. Press. P. 104.

SIEBERT, W. M. Frequency discrimination in the auditory system: Place or periodicity mechanisms? *Proceedings of the IEEE*, 1970, **58**, 723.

WANDELL, B. A. Speed–accuracy tradeoff in visual detection: applications of neural counting and timing. *Vision Research*, 1977, **17**, 217.

YELLOTT, J. I., JR. The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 1977, **15**, 109–144.

RECEIVED: September 29, 1977