ATTENTION AND PERFORMANCE IV
© 1973
ACADEMIC PRESS, INC., NEW YORK AND LONDON

Speed-Accuracy Trade Off in Auditory Detection[1]

David M. Green

Department of Psychology
University of California, San Diego
La Jolla, California

R. Duncan Luce

The Institute for Advanced Study
Princeton, New Jersey[2]

ABSTRACT

A modified yes–no design, with response-terminated signals, a fixed deadline, and payoffs for various stimulus–response combinations, was used to study the speed–accuracy trade off. The signal was a 1000-Hz tone presented in noise. The deadline was varied from 300 to 2000 msec, with a fixed symmetric payoff matrix, and the payoff matrix was varied with a fixed 600-msec deadline. The data are compared with four models for the speed-accuracy trade: (*a*) the fast-guess model, (*b*) a random-walk model, (*c*) a Poisson counting model, and (*d*) a Poisson timing model. When a fixed deadline is imposed on both signal and noise trials, the counting model is favored because the mean reaction time is roughly independent of both the stimulus and response and the receiver operating characteristic (ROC) plotted in normal–normal coordinates has a slope less than unity. When the deadline is imposed only on signal trials, the data favor the timing model because the mean reaction time on noise trials is a linear function of the mean reaction time on signal trials and the ROC curve has a slope greater than unity.

[1]This work was supported in part by a grant from the National Science Foundation and in part by a grant from the Alfred P. Sloan Foundation to the Institute for Advanced Study.

[2]Present address: School of Social Science, University of California at Irvine, Irvine, California.

I. Introduction

Two common measures of behavior are speed and accuracy. For many tasks they are inversely related, which is not surprising intuitively because accuracy arises from care and caution, and that requires time. Although the inverse relation is generally acknowledged to hold even for simple detection tasks, the precise basis for the trade off is in doubt. This hardly could be otherwise since it can only be formulated in terms of the observer's sensory–decision process, and that is still a matter of debate. Opinion ranges from, on the one hand, theories of sequential decision making, which suppose a steady accumulation of information and a response rule whose time of application is conditional on that information, to, on the other hand, simple threshold models, which suppose the sensory information is of an all-or-none character and that the trade off arises simply by varying the mix of two quite different response modes. Another school of thought bases its models on the assumption that signal energy is recoded into neural pulses, the rate of which is an increasing function of signal intensity. Since either the number of pulses in a fixed interval of time or the time for a fixed number of pulses to occur yields an estimate of the pulse rate, and hence of the signal intensity, two classes of models arise based on these statistics. Of course, as in all statistical situations, the more data accumulated, the more accurate is the estimate—the price being time. Thus, the speed–accuracy trade off is inherent in the basic idea of a pulse model, and the only question is whether detailed analyses of either model lead to a good account of the experimental results.

This contribution attempts to attack this problem experimentally. We first describe the experimental situation—basically, a response-terminated, yes–no design with a response-time deadline and payoffs for accuracy. Next, we outline the key predictions for the four models just mentioned. The data then are examined from each of these viewpoints, and one model is favored strongly. A modified deadline procedure is suggested to see if the observers can be induced to follow another model, as indeed they can. Independent of any model, this change in procedure has a very marked effect both on the reaction times and on the ROC curves.

II. Experimental Design

The beginning of each trial was marked clearly by a warning light, at which point a pure tone signal either was or was not added to the background noise. The observer was permitted one of two responses, yes or no, corresponding to the two possible types of trials. One unusual feature of the design is that the signal, when presented, remained on until the response occurred. The reason for using a response-terminated signal was to maintain statistically stationary conditions throughout the listening interval; this seems especially important when the effect of a variable deadline is studied. Following the response, information feedback was presented for .5 sec, and after another .5 sec the next trial began.

Each of three observers listened binaurally through earphones in a sound-treated room. The white Gaussian noise had a spectrum level of 40 dB. The signal, a 1000 Hz sinusoid, was either weak or intense—20 or 50 dB above the noise power density ($10 \log P/N_0$). With weak signals, performance is essentially at chance when a response must be made within 300 msec of the warning light and is nearly perfect with a 1000-msec deadline. The intense signal is roughly at the level generally used in trade-off experiments.

Each experimental session lasted about 2 hr and consisted of five runs. Each run was under one experimental condition (see below) and consisted of about 250 trials. A total of six sessions were run, so there are about 1500 trials per condition.

We recorded both the time from the onset of a trial to the response and the nature of the response. The two independent conditional response probabilities, $P(\mathrm{Y}|\mathrm{s})$ and $P(\mathrm{Y}|\mathrm{n})$, were estimated in the usual way. The means and variances of reaction-time distributions (MRT and VRT, respectively) also were estimated in the usual way; MRT and VRT are subscripted as needed by the presentation, s (signal) or n (noise), and/or the response, Y (yes) or N (no).

In addition to an hourly wage of \$1.88, observers received points according to a payoff matrix for accuracy and a deadline for speed; these are described in the next section. The points were accumulated and converted into a bonus at the end of each run on a competitive basis among sets of three observers as follows. Let the points accumulated by the ith subject on a given run be denoted V_i, which was always positive. His share of the \$0.25 bonus for that run was $V_i^2/\Sigma_{i=1}^3 V_i^2$.

III. Experimental Variables

Aside from the two signal levels, the two major variables were both instructional, the one intended to manipulate the speed–accuracy trade and the other, the response bias. Any response prior to the warning light was fined 25 points; as a result, such anticipations occurred with a relative frequency of less than .001. Any response following the prescribed deadline D was fined 4 points, independent of accuracy. Any response between the warning light and D was paid off according to the following matrix:

$$\begin{array}{c} \\ s \\ n \end{array} \begin{array}{c} \begin{array}{cc} Y & N \end{array} \\ \begin{bmatrix} X & -10 \\ -10 & Y \end{bmatrix} \end{array},$$

where the observers were informed of the value of (X, Y) in each condition. For reasons that will become apparent, we refer to this (standard) deadline procedure as the sn deadline (*s*ignal and *n*oise deadline).

In one experiment, (X, Y) was fixed at (10, 10), thereby producing a symmetric payoff matrix, and D was varied over the values 250, 300, 400, 500, 600, 800, 1000, 1500, and 2000 msec. The aim of this manipulation was to produce a speed–accuracy exchange. In a second experiment, D was fixed at 600 msec and the (X, Y) pair was varied over the values (20, 1)(15, 5)(10, 10)(5, 15)(1, 20) points. The aim of this manipulation was to generate an ROC curve, but without any speed–accuracy exchange. As the data will make clear, both manipulations were successful.

IV. Models

As the four models we shall test are described fully in the literature, it suffices to suggest their general nature and to state the exact predictions to be compared with data. Unfortunately, comparable predictions from the models are not available. Our policy is to accept whatever seems to be a characteristic prediction of a model, especially predictions of linear relations, and to examine the data in an absolute sense without worrying about what the other models predict. This is a sound strategy when the data unambiguously reject a model; it is much more suspect when we are inclined to accept one.

A. Fast-Guess Model

The fast-guess model (Ollman, 1966; Yellott, 1967) supposes that the observer selects on each trial one of two wholly different modes of behavior. In one, he pays attention to the signal and responds only when he has recognized it. We assume that the probability of a correct response is

a and the mean time for it to occur is μ_s. In the other mode, the observer responds as fast as he can to the onset of the signal, making no attempt to identify it; this accounts for the name of the model. He has some bias probability b for responding Y, and the mean response time is μ_g, where $\mu_g < \mu_s$. The observer uses the first mode with probability q and the second with probability $1 - q$. Any experimental manipulation that alters the probability q of paying attention generates a speed–accuracy trade off.

Let P_c and P_e denote the probabilities of correct and of error responses. Thus, if the signal is presented with probability $\frac{1}{2}$,

$$P_c = \tfrac{1}{2}P(\mathrm{Y}|\mathrm{s}) + \tfrac{1}{2}P(\mathrm{N}|\mathrm{n})$$

and

$$P_e = 1 - P_c.$$

Let M_c and M_e denote the MRTs to correct and error responses. Then it is not difficult to show that

$$P_cM_c - P_eM_e = \mu_s(P_c - P_e).$$

We use this linear prediction to test the fast-guess model.

B. The Random-Walk Model

The literature includes a variety of sequential-decision models. Of these, Laming's (1968) seems to be the best worked out. The observer divides time into a sequence of equal intervals, each of which is then treated as a fixed-interval, yes–no situation. The sensory random variables (RV) observed in the several intervals are assumed to be independent and identically distributed; of course, the distribution depends upon whether or not the signal is present. Two response criteria, β_1 and β_2, are established. The decision rule has the following character. If no decision has been reached prior to the ith observation, the mean of all i observations is compared with the criteria. If it is less than β_1, respond N; if it is greater than β_2, respond Y; and if it lies between β_1 and β_2, collect the $(i + 1)$st observation and proceed as on the ith observation. The onset of the sampling is a parameter of the model; it may begin before the onset of the potential signal.

Distinctive quantitative predictions are difficult to come by in any sequential model, and so we will content ourselves with two qualitative ones. Using Laming's (1968) labeling:

B. In a two-choice experiment that signal which elicits the faster reaction, on average, has the smaller probability of error, and conversely [p. 44].

C′. In two-choice experiments errors are faster than the same response made correctly [p. 82].

C. A Counting Model

This model (McGill, 1967) assumes that the sensory trasducer converts signal energy into one or more pulse trains, which, for a constant intensity signal, are identical Poisson processes, that is, the times between successive pulses are independent, identically distributed exponential RVs. The common intensity parameter of these Poisson processes is assumed to be an increasing function of signal intensity, given that all other stimulus parameters are constant. Put another way, the expected time between pulses is a decreasing function of signal intensity.

The observer selects a time period Δ during which the number of pulses is counted. This RV is then treated just as likelihood ratio is in the theory of signal detectability (TSD), that is, if it is larger than some criterion β, he responds Y, and if it is smaller, he responds N. Two predictions follow readily. First, since the time for initiating a response does not depend on either the stimulus condition or on the response made, the observed reaction time distributions should be the same in all four cells. Second, we derive the form of the ROC curve. Let μ and ν denote, respectively, the Poisson parameters corresponding to s and n, and let $z(\mathrm{s})$ and $z(\mathrm{n})$ be, respectively, the normal deviates corresponding to $P(\mathrm{Y}|s)$ and $P(\mathrm{Y}|\mathrm{n})$. Using the normal approximation to the Poisson process, the following approximate linear relation holds:

$$z(\mathrm{s}) \cong (\nu/\mu)^{1/2}\, z(\mathrm{n}) + \Delta^{1/2}\,[(\mu - \nu)/\mu^{1/2}]. \tag{1}$$

This agrees with the prediction of TSD but, in addition, it predicts exactly how the slope of the ROC curve decreases with increased signal intensity.

As a measure of accuracy, define d' to be the value of $z(\mathrm{s})$ corresponding to $z(\mathrm{n}) = 0$, that is, to $P(\mathrm{Y}|\mathrm{n}) = \frac{1}{2}$. Then the speed–accuracy trade off is described in terms of the size Δ of the subject-controlled observation interval by the equations

$$\mathrm{MRT} = \bar{r} + \Delta, \qquad d' = \Delta^{1/2}[(\mu - \nu)/\mu^{1/2}], \tag{2}$$

where $\bar{r}$ is the mean of the residual times not accounted for by the observation time.

D. A Timing Model

Timing models (Luce & Green, 1972) assume exactly the same pulse structure as the counting models. They differ only in the processing of the pulses. Instead of assuming that pulses are collected for a fixed time and then counted, these models assume that a fixed number of pulses are collected and the time required is measured. That time is a RV, which is treated much as in TSD except that small values—short interarrival times (IATs)—correspond to the signal rather than to the noise, because the shorter the IAT, the more likely it is that a signal is present.

We must take into account a complication that we could ignore in the counting model. We shall suppose that the stimuli activate pulse trains on J statistically identical, parallel channels (this, no doubt, is a gross oversimplification) and that the observer collects κ IATs from each channel, for a total of $J\kappa$. The response is determined by the mean of all $J\kappa$ IATs, and the time of initiation is determined by the slowest of the J channels. Thus, the response time depends on J and κ separately, not just on $J\kappa$.

Independent of the response, it can be shown (Luce & Green, 1972, Eq. 23) that the mean reaction time to the signal is of the form

$$\mathrm{MRT}_s = \bar{r} + h(J, \kappa)/\mu, \tag{3}$$

where $\bar{r}$ again denotes the mean residual time. An approximation to this function h is

$$h(J, \kappa) \simeq \kappa + 1 + (\kappa + 1)^{1/2} H(J), \tag{4}$$

where $H(J)$ is the mean of the largest of J normally distributed RVs, each with mean 0 and variance 1. A table of H is given by Tippett (1925). The equation for MRT_n is the same as Eq. (3), with ν substituted for μ. Eliminating h from these two equations yields the testable linear relation

$$\mathrm{MRT}_n = (\mu/\nu)\,\mathrm{MRT}_s + \bar{r}(1 - \mu/\nu). \tag{5}$$

Obviously, this provides not only a test of the model, but a way to estimate μ/ν and, in principle, $\bar{r}$.

A similar calculation of variances yields

$$\mathrm{VRT}_n = (\mu/\nu)^2\,\mathrm{VRT}_s + V(r)\,[1 - (\mu/\nu)^2], \tag{6}$$

where $V(r)$ is the variance of the residual distribution.

Again, using the normal approximation to the gamma, the ROC curve can be shown (Luce & Green, 1972, Eq. 31) to be approximately,

$$z(\mathrm{s}) \simeq (\mu/\nu)z(\mathrm{n}) + (J\,\kappa)^{1/2}(\mu/\nu - 1). \tag{7}$$

Although the linear form agrees with TSD and with the counting model, the prediction has the striking feature that the slope must be greater than 1 since $\mu > \nu$, which is the opposite of what is predicted by the counting model [Eq. (2)] and contrary to almost the entire body of data reported in the detection literature.

Using the same accuracy measure as in the counting model, the speed–accuracy trade off is described by Eqs. (3) and (4), together with

$$d' = (J\kappa)^{1/2}(\mu/\nu - 1). \tag{8}$$

In this case, κ is the subject-controlled parameter that affects both speed and accuracy. Note that this model predicts approximately the same trade off as does the counting model: MRT is approximately linear in the parameter and d' is proportional to its square root.

V. Data from the sn-Deadline Experiment

A. Variability

The deadline had the desired effect of varying the MRT from 100 to 1000 msec; it also affected the variability of the RTs. Figure 1 presents the scatter plot of standard deviation versus mean for all of the experimental data of this paper (including a deadline procedure not yet described). The open circles are data using a large signal-to-noise ratio, where the detection of the signal is no problem. The mean reaction times are short, 100–250 msec, and the variability is about 50 msec. The solid points are data using a small signal-to-noise ratio, where the detection of the signal is difficult. Although the correlation is far from 1, it is evident that the standard deviation tends to increase with MRT. A general rule of thumb is that, for these procedures and the weak signal, the standard deviation is about one-third of the mean. This fact must be kept in mind when comparing various conditions.

B. $P_cM_c - P_eM_e$ versus $P_c - P_e$

The linear prediction of the fast-guess model has been tested for easy-to-discriminate stimuli (Ollman, 1966; Yellott, 1967; Link & Tindall, 1971). In general, the fit is impressive except for conditions of extreme accuracy ($P_c \cong 1$), in which cases the data points are well above the extrapolated linear curve. Our data for both signal levels are shown in Fig. 2.

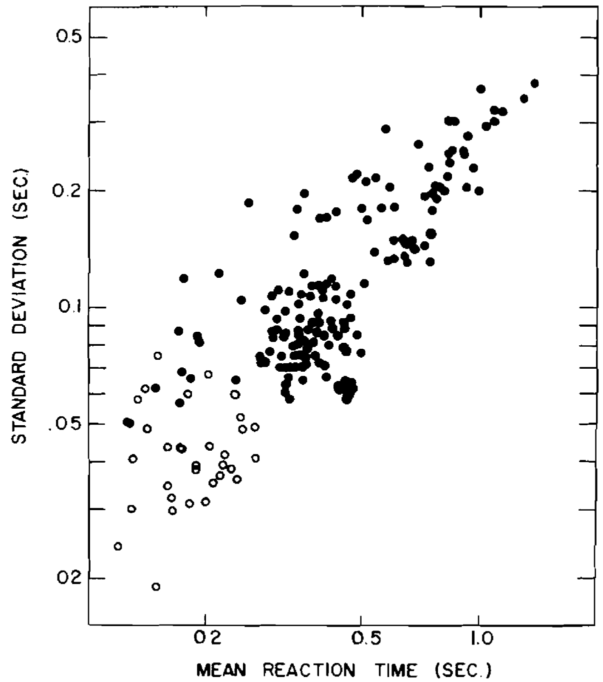

Fig. 1. Standard deviation versus mean reaction times for all conditions and observers: (●) 10 log P/N_0 = 20; (○) 10 log P/N_0 = 50.

Those for the 50-dB signal conform to those in the literature; those for the weak 20-dB signal are much more discrepant from the theory. For the latter, a linear fit through the origin summarizes only a very narrow range of these data. Thus, we conclude that the fast-guess model does not provide an adequate description of behavior in a response-terminated, sn-deadline, Y–N detection design using weak signals.

C. Mean Reaction Time

Figure 3 presents the MRT_s and MRT_n data from the variable deadline experiment for each response separately. To a good first approximation, the stimulus condition does not affect the results. To the extent that this is so—recall that the data are quite variable (Fig. 1)—it disagrees with prediction C′ of the random-walk model and with Eq. (5) of the timing model. It is in perfect accord with the prediction of the counting model.

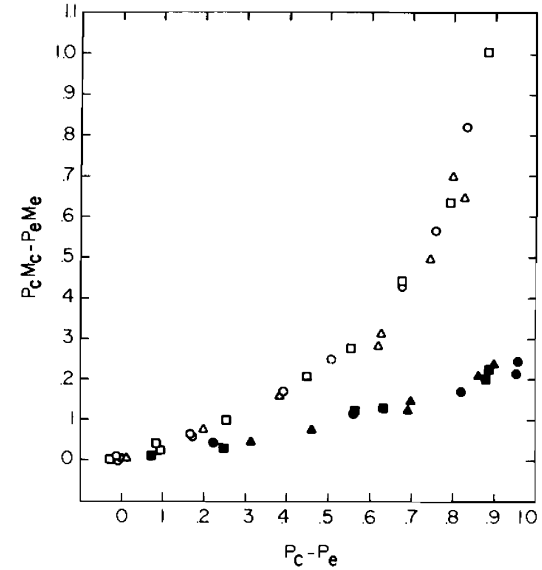

Fig. 2. Fast-guess analysis of data for sn-deadline conditions:

10 log P/N_0 =	20	50
Observer 1	○	●
2	△	▲
3	□	■

Each point was generated by a different deadline value. The solid points were obtained using an easy-to-detect signal, and the open ones using a hard-to-detect signal.

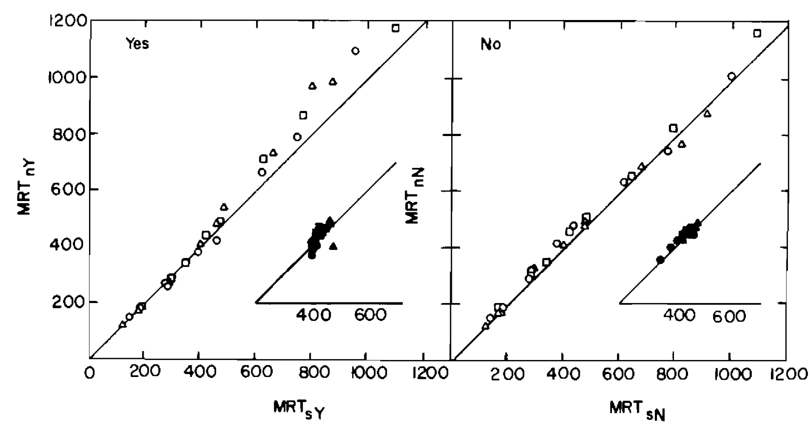

Fig. 3. The four stimulus–response mean reaction times obtained in the sn-deadline procedure: (○) Observer 1; (△) Observer 2; (□) Observer 3. The small subscript indicates the stimulus condition: s for signal, n for noise. The capital letter indicates the response: Y for yes, N for no. The open points are for conditions in which the sn deadline was varied. The solid points in the insert are the data for the condition in which the criteria was varied for a fixed, 600-msec deadline.

D. The ROC Curve

With a fixed deadline but a variable payoff matrix, the MRTs are approximately the same independent of either the stimulus or the response; see the insets in Fig. 3. The ROC curves for these conditions are shown in Fig. 4. The MRT data are inconsistent with prediction B of the random-walk model, and the ROC curves are grossly inconsistent with Eq. (7) of the timing model because the slopes are all less than 1. They are consistent with the counting model, and we obtain the values 1.19, 2.13, and 1.25 as estimates[2] of μ/ν.

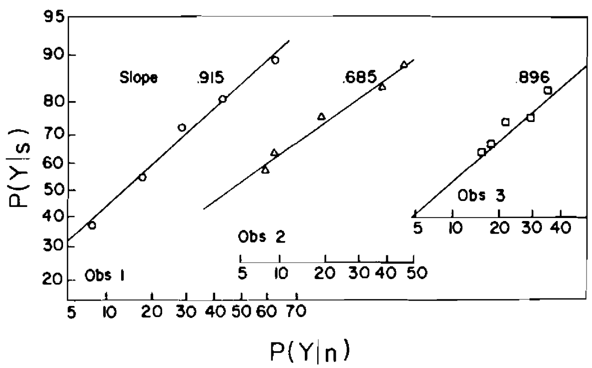

Fig. 4. ROC curves, plotted on double probability paper, for a sn-deadline of 600 msec (10 log P/N_0 = 20); (○) Observer 4; (△) Observer 5; (□) Observer 6. Each point is obtained from a different (X, Y) payoff condition.

VI. Intermediate Discussion

Of the four models, the data clearly favor the counting one. The question is, first, how seriously should we take this and, second, how general is the conclusion.

It is difficult to know whether an adequate repair of the fast-guess or random-walk models is possible. The data rejecting the fast-guess model

[2]Here and elsewhere when we examine a linear relation between two variables, both of which are random, we do both regressions and report the geometric mean of the two slopes and the arithmetic mean of the intercepts.

suggest that the observer may use more than two states and that, perhaps, he enters into some sort of sequential decision making. However, other aspects of the data did not support sequential models. It is unclear to us how seriously to take this since the predictions were simply inequalities without any indication of the magnitudes of the differences that should be found. This is, in fact, typical of sequential models; their random structure is inherently difficult to analyze, and so unless forced to it, one is inclined to search for simpler models.

Both the counting and timing models are (in this situation) very simple, and clearly the counting one is far the better of the two. That being so, one can wonder why we ever seriously entertained the timing model and, if our reasons seem adequate, what it was about this experiment that forced our observers to abandon timing behavior. Our reasons for considering timing models are both general and specific. We showed (Luce & Green, 1972) that they readily account for a wide range of psychophysical findings, including magnitude estimation as well as detection and discrimination. One new development was a theory for the detection of the onset of weak signals that are presented at random times. Unpublished simulations of that model seem, at least qualitatively, to account for some previous data (Green & Luce, 1970). Thus, we were encouraged to think it an interesting idea.

Why, then, did it fail so badly to account for the sn-deadline procedure? Although the number of IATs collected per channel, κ, is independent of the stimulus, the time to collect them is not. It takes longer to do so on n trials than it does on s trials. Therefore, either the observer chooses κ sufficiently small to avoid the deadline on n trials, in which case he does not collect as much information on s trials as he might have, or he chooses a larger κ, thereby reversing the difficulties. Clearly, in this deadline procedure, it is much more efficient to behave, if that is possible, as in the counting model: select a time, somewhat less than the deadline, and count pulses in that time. This appears to be what all three observers did.

The next question, then, is this: Can we modify the experimental design so that the timing procedure, if available, is more efficient than the counting one? Once asked, the answer is obvious: impose the deadline on only one of the stimulus conditions. It seems most natural to place it on the signal, and so we call this the s-deadline procedure. To be quite explicit: all anticipations are punished; on n trials, all responses after the warning light are paid off according to the payoff matrix; on s trials, all responses after the warning light and before the deadline D are paid off according to the payoff matrix, but responses after the deadline are fined.

VII. Data from the s-Deadline Experiment

A. $P_cM_c - P_eM_e$ versus $P_c - P_e$

The fast-guess model again is rejected, as can be seen from Fig. 5. As this plot seems little different from Fig. 2, we do not discuss it further.

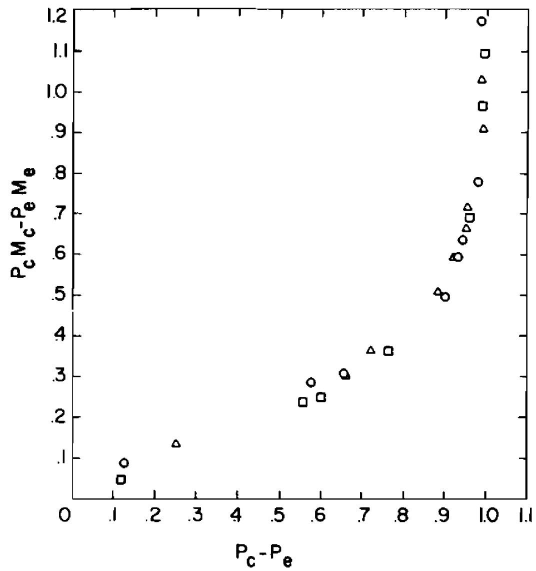

Fig. 5. Fast-guess analysis of data obtained from different s-deadline conditions.

B. Mean Reaction Time

Figure 6 is the analog of Fig. 3. Obviously, these data are less regular, but apparently MRT is no longer independent of the stimulus. The data for the N responses agree with prediction C' of the random-walk model since errors are faster than correct responses, but those for the Y responses go in the opposite direction, at least for the longer reaction times. The basic fact revealed by this plot is that $\text{MRT}_n < \text{MRT}_s$, as is shown clearly in Fig. 7. These data disagree with the counting model and agree with the timing one. The estimates of μ/ν, the slopes of the fitted line, are shown on the graph. In principle, the intercept determines $\bar{r}$; however, it appears to be too unreliable to be useful. It is easy to see why, for if $\bar{r}$ is about .2, then $\bar{r}(\mu/\nu - 1)$ is only about .06 sec.

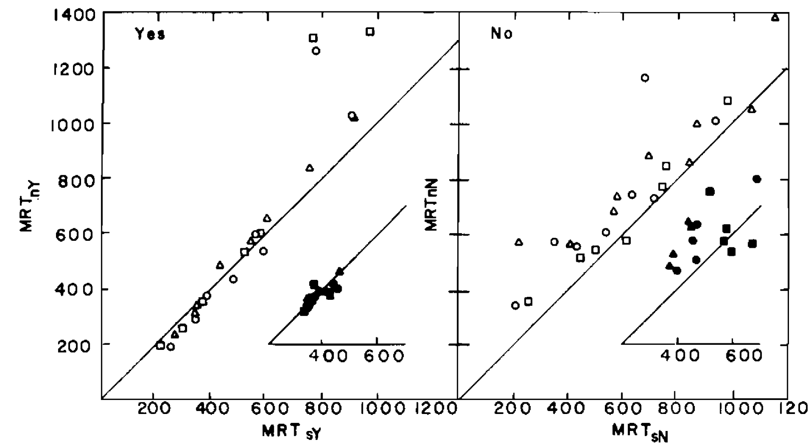

Fig. 6. The mean reaction times for various s deadlines. The figure is analogous to Fig. 3, and the same notation is used.

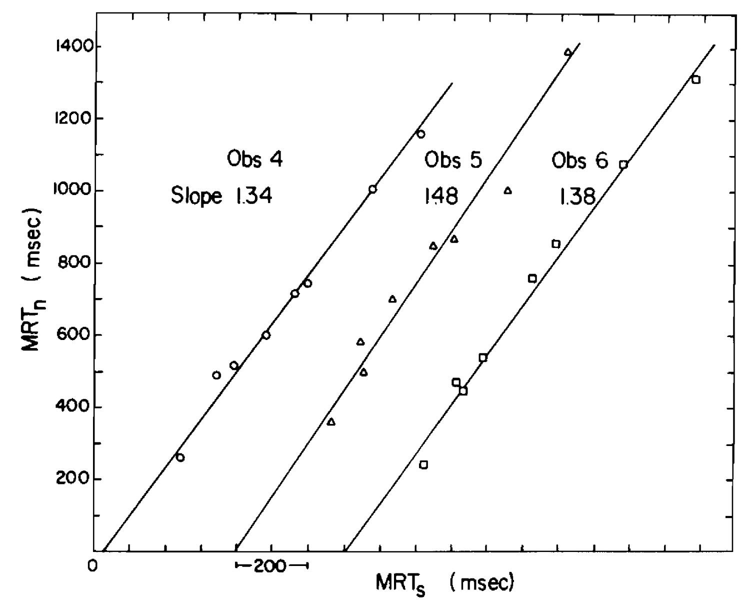

Fig. 7. The mean reaction time for noise trials versus the mean reaction time for signal trials using the s-deadline procedure. Each point was generated by a different deadline.

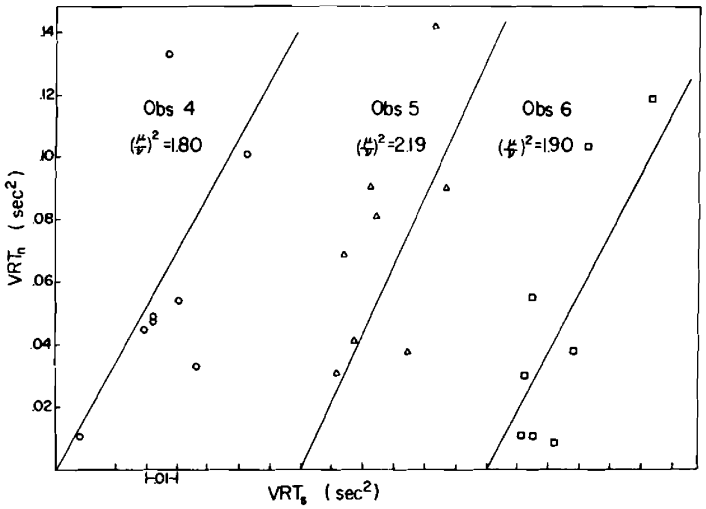

Fig. 8. The variance of the reaction times for the two stimulus conditions using the s-deadline procedure. Each point was generated by a different deadline.

The variance data are shown in Fig. 8. The theoretical curves are those obtained from Eq. (6) using the estimates of μ/ν from Fig. 7 and assuming (incorrectly, of course) that $V(r) = 0$. Given the variability of variance estimates, the fits are not bad.

C. The ROC Curve

Figure 9 presents the ROC curve for these three observers. Observe that the slopes are appreciably greater than 1, as predicted by the timing model [Eq. (7)] and in contradiction to the counting model [Eq. (1)]. In fact, the least-squares estimates of the slope of the ROC curve, which is an approximation to the ratio μ/ν, are very close to the independent estimates of the same quantity obtained from the MRT (see Fig. 7). Unfortunately, this remarkable agreement between two estimates is probably fortuitous if the timing model is correct. According to that model, the time taken to accumulate $J\kappa$ pulses is a gamma distribution of order $J\kappa$. It has a mean $J\kappa/\mu$ when the signal is presented and a mean of $J\kappa/\nu$ when noise alone is presented. Thus, the true ROC curve is constructed by integrating the gamma distribution. Using the estimates of $\mu/\nu = 1.48$, obtained from the

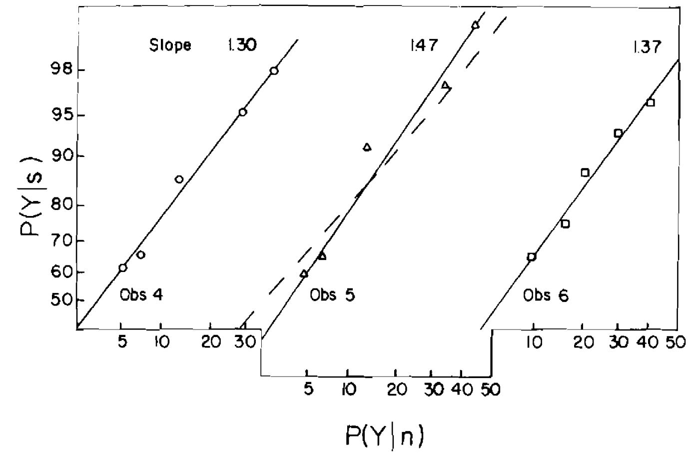

Fig. 9. ROC curves, plotted on double probability paper, for an s deadline of 600 msec. Each point was obtained from a different (X, Y) payoff condition. The dotted curve is the exact gamma curve (see text).

data for Observer 5 (Fig. 7), and estimating $J\kappa = 32$, via Eq. (7) and the data of Fig. 9, we obtain the exact prediction for the ROC curve shown as the dotted line in Fig. 9. This ROC derived from the gamma distribution is nearly linear over the major portion of the scale used in the figure, but it has a slope of about 1.20 rather than the least-squares fit to the data points of 1.47.

It is important to understand how this discrepancy arises. These two linear ROC curves actually give nearly the same fit to the data. A difference of only .01 looms very large in the standard score plots when the actual probability is less than .05 or greater than .95, and thus small differences at these extremes produce large differences in slopes. Since it is difficult to estimate such extreme probabilities accurately and since the normal approximation to the gamma has some error, the estimated slope is almost surely wrong by a sizable amount. Thus, it is not a very useful statistic except to judge crude qualitative facts such as whether the slope is greater or less than unity. It comes as an unpleasant surprise that even for gamma distribution of order 50, the normal approximation is sufficiently inaccurate to misestimate the slope by 20%. Because the gamma is not tabulated for orders beyond 50, we cannot say when this error is reduced to, say, 5%.

D. Estimation of Parameters for the Timing Model

We have estimates of μ/ν from the MRT data of Fig. 7 and Eq. (5). Using the ROC curve and that estimate of μ/ν, Eq. (8) yields an estimate of $J\kappa$. Using the MRT from strong signals as an estimate of $\bar{r}$, we can select $h(J, \kappa)/\mu$ so as to minimize the sum of the squared errors in Eq. 3:

$$[\mathrm{MRT}_s - \bar{r} - h/\mu]^2 + [\mathrm{MRT}_n - \bar{r} - (\mu/\nu)(h/\mu)]^2,$$

which yields the estimate

$$\frac{h}{\mu} = \frac{\mathrm{MRT}_s - \bar{r} + (\mathrm{MRT}_n - \bar{r})(\mu/\nu)}{1 + (\mu/\nu)^2}.$$

Another independent equation is needed to determine all of the parameters. Since we do not have one, we present in Fig. 10 for each possible value of J the corresponding values of κ and μ, where the latter follows from Eq. (3). The values of $\bar{r}$—165, 168 and 173 msec for Observers 4, 5 and 6—were the observed MRT for a very large signal-to-noise ratio ($10 \log P/N_0 = 80$). The values of $J\kappa$ were 50 for 4, 32 for 5, and 32 for 6.

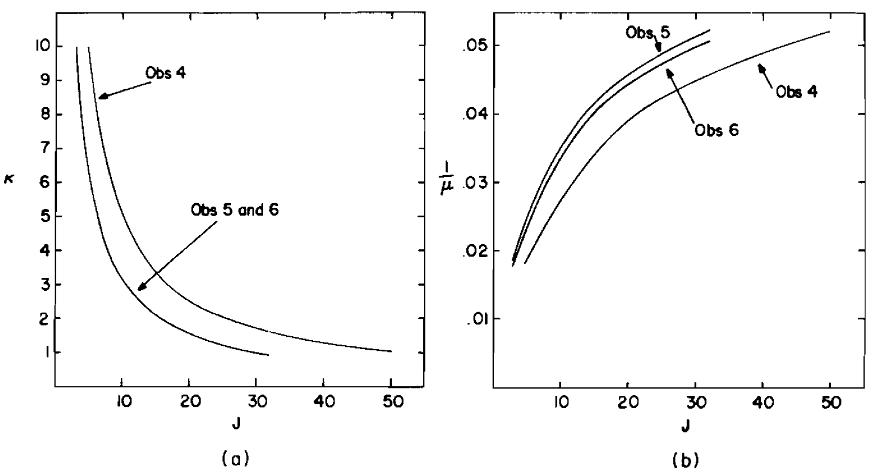

Fig. 10. Values of (a) κ and (b) $1/\mu$ for various J values. According to the timing model both κ and J are integers and the product κ times J is a constant. Its value as estimated from the data is shown in the left of the figure. Depending upon the value of J, a different $1/\mu$ estimate results, as shown in the right side of the figure. Thus, for Observer 4, if we assume κ is 3, then the number of parallel channels J is about 18, and the value $1/\mu$ is approximately .033 sec or about 30 pulses per second per channel.

VIII. The d' versus MRT Trade Off

Given the widespread use of d' to summarize sensitivity, a very natural way to represent the speed–accuracy trade off is by plotting d' versus MRT. Certainly this is an appropriate function for both the counting and timing

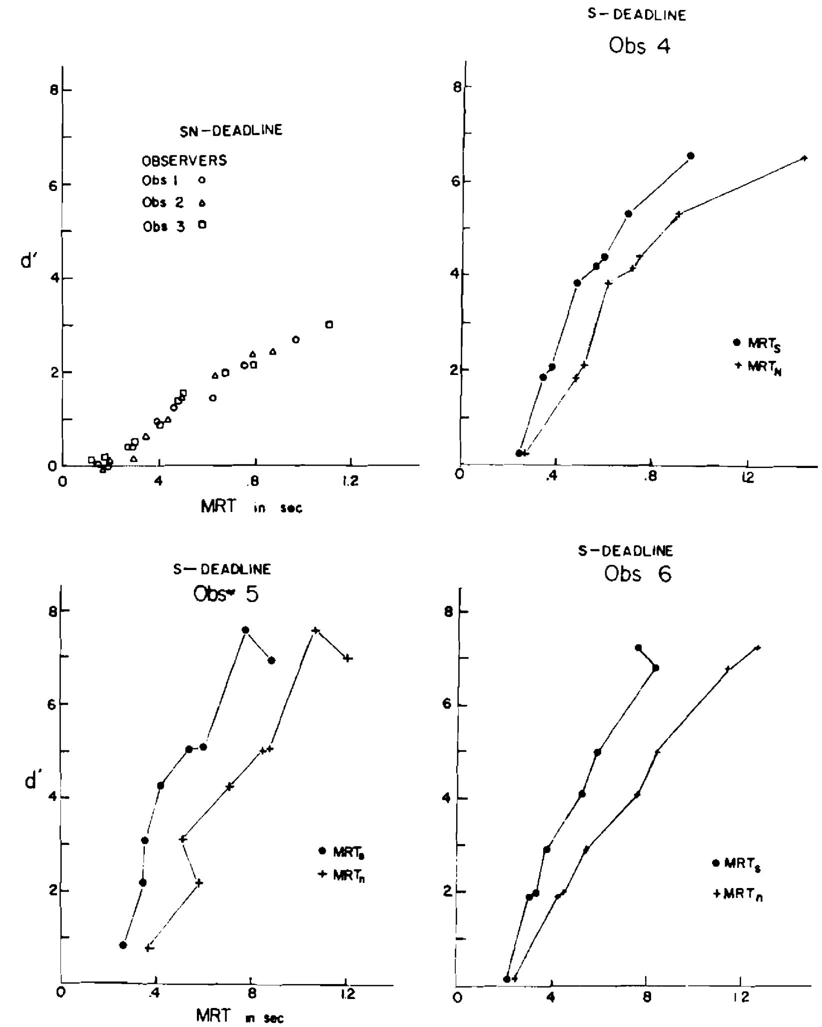

Fig. 11. Plots of d' versus MRT. The sn deadline has the three observers plotted together: (○) Observer 1; (△) Observer 2; (□) Observer 3. The s deadline shows d' plotted against both MRT_s (●) and MRT_n (+).

models. The only difficulty in constructing this function from our data is the fact that we do not have ROC curves for each deadline. However, the models say that the slope of the ROC curve is independent of either Δ in the counting model or κ in the timing model, so given our one estimate of that slope, we can estimate d' at each of the deadlines. This was done, and the resulting plots of d' versus MRT are shown in Fig. 11. Because the MRT to s and n are quite different in the s-deadline experiment, each of those three observers is described by two curves.

Qualitatively, these graphs exhibit two important features. First, the intercept corresponding to $d' = 0$ is distinctly less in the sn-deadline design than in the s-deadline one. Second, and far more spectacular, the slope of the s-deadline curves, even that corresponding to MRT_n, is considerably greater than that of the sn-deadline curves. Beyond 400 msec, the value of d' for the s-deadline is more than double that for the sn-deadline.

Is this predicted? For the counting model, Eq. (2) immediately yields

$$d' = \left(\frac{\mu - v}{\mu^{1/2}}\right)(\mathrm{MRT} - \bar{r})^{1/2}.$$

For the timing model, matters are slightly more complicated. From Eqs. (3) and (4) we may write

$$\mathrm{MRT}_s = \left[\bar{r} + \frac{1 + H(J)}{\mu}\right] + \left\{\frac{\kappa + [(\kappa + 1)^{1/2} - 1]H(J)}{\mu}\right\},$$

where we have grouped the terms on the right so that the second one is 0 when κ is 0, and hence the first describes the intercept when $d' = 0$. Obviously, the intercept of the timing model, which we believe applies to the s-deadline experiment, is greater than that of the counting model, which applies to the sn experiment. The difference is $[1 + H(J)]/\mu$. Observe that if we neglect $[(\kappa + 1)^{1/2} - 1]H(J)/\mu$,

$$\mathrm{MRT}_s \cong \bar{r} + \frac{1 + H(J)}{\mu} + \frac{\kappa}{\mu},$$

and so solving for κ and substituting in Eq. (8) we obtain

$$d' \cong \left(\frac{\mu}{v}\right)\left(\frac{\mu - v}{\mu^{1/2}}\right)\left(\mathrm{MRT}_s - \bar{r} - \frac{1 + H(J)}{\mu}\right)^{1/2}.$$

Similarly

$$d' \cong \left(\frac{\mu}{v}\right)^{1/2}\left(\frac{\mu - v}{\mu^{1/2}}\right)\left(\mathrm{MRT}_n - \bar{r} - \frac{1 + H(J)}{v}\right)^{1/2}.$$

We note that the rate of growth in the timing experiment for the s and n curves is μ/v and $(\mu/v)^{1/2}$, respectively, times that of the counting experiment.

A direct numerical comparison is not possible, however, for two reasons. First, the observers are different. Second, the predicted growth in the timing model is only approximate. A third point is the empirical curve in the sn-deadline experiment does not really agree with the predicted, square-root form; it is more nearly linear. This discrepancy could arise for either or both of two reasons. First, the estimates of values of d' near 0 are not very stable, and so those points may be misplaced somewhat. Second, for long deadlines it is to the subject's advantage to switch to timing behavior and so raise the value of d' above that predicted by the counting model. That there may have been a little of this is suggested by the slight departure from $\mathrm{MRT}_{sY} = \mathrm{MRT}_{nY}$ in Fig. 3.

Our conclusion, then, is that these data are about as one would expect if the sn deadline invoked counting behavior and the s-deadline, timing behavior. Independent of any theory, however, the expirical difference in the trade-off functions is striking.

Several relevant papers have come to our attention in connection with the d' versus MRT trade off. The key one is that of Taylor, Lindsay, and Forbes (1967) in which it is shown that in a 2×2 design d'^2 is approximately linear with MRT. They replotted data of Schouten and Bekker (1967) and confirmed that $d'^2 = A(\mathrm{MRT} - r)$. Lappin and Disch (1972), using easily discriminated visual stimuli and instructions to maintain a 25% error rate, grouped their data according to RT and then studied five different accuracy measures, including d' and d'^2 versus median RT. There was very little difference among them. Other data tending to support the d'^2 hypothesis can be found in papers by Fitts (1966), Pachella and Fisher (1972), Pachella, Fisher, and Karsh (1968), Pachella and Pew (1968), and Pew (1969).

Terminological inconsistency exists for the function relating a measure of accuracy to a measure of response time. We referred to it above as the speed–accuracy trade off, as have others; Pew suggested speed–accuracy operating characteristic (S–A OC), which seems compatible with the ROC terminology; and Lappin and Disch suggested latency-operating characteristic (LOC), which seems poor both because it has been used in at least one other way and because it emphasizes one aspect of the trade off at the expense of the other.

IX. Data from the Second sn-Deadline Experiment

After completing the above s-deadline experiment, we ran these observers in the sn-deadline procedure using a 500-msec deadline and obtained their ROC curves.

A. Mean Reaction Time

Table 1 shows the estimates of $MRT_n - MRT_s$ for these observers in both deadline procedures and the same thing for the first group of observers. We note that the sn deadline for these observers does not produce the negligible difference observed with the first three subjects run on this condition, as we would expect from the counting model, nor is the difference as large as that produced by the s-deadline procedure for the same observers. Rather the difference suggests some intermediate mode of behavior.

Table 1. *MRT_n-MRT_s (msec) Averaged over Five Different Conditions of the ROC Curve*

Observer	s-Deadline	sn-Deadline
1	—	4
2	—	3
3	—	27
4	101	35
5	185	44
6	144	42

B. The ROC Curve

Figure 12 presents the three ROC curves and, unlike those of Fig. 4, only one observer exhibits a slope of less than 1. The size of the slopes is noticeably smaller than those obtained with the s-deadline condition, but not nearly as small as those obtained from the first three observers using just the sn-deadline condition (Fig. 4).

One possible explanation of the change in slope of the ROC curve from the s deadline to the sn deadline is that the quality of detection is somewhat less in the sn deadline (lower d'). We therefore raised the signal level a small amount (3 dB) so that the slopes might be compared under more identical conditions. Fig. 13 shows the ROC for this condition. The slopes for two observers increase, whereas that of the third (Obs. 4) remains less than unity.

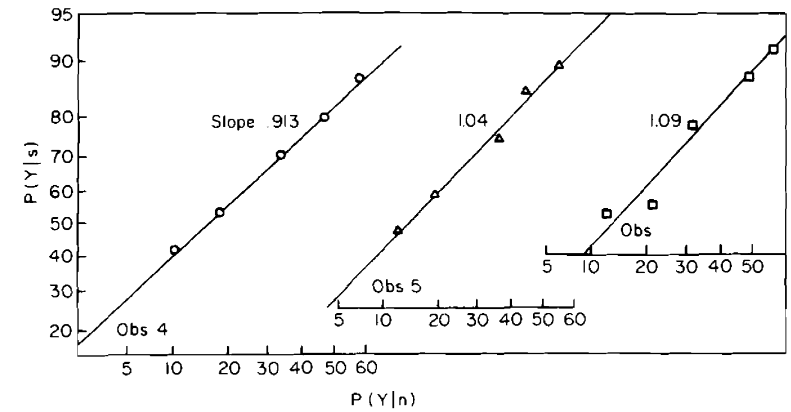

Fig. 12. ROC curves, plotted on double probability paper, for Observers 4, 5 and 6 using the sn-deadline procedure.

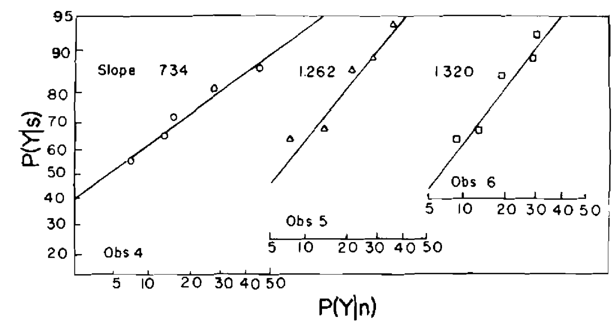

Fig. 13. ROC curves using the sn-deadline condition with signal 3 dB above those used in Figs. 9 and 12.

X. Discussion

There is little doubt that the s-deadline procedure results in unusual ROC data with slopes distinctly greater than 1. These data and the MRT data are consistent with the timing model and not with the other models. We had hoped, of course, that by changing to the sn deadline, the observers would have exhibited behavior similar to the first group of observers, which we would have interpreted as switching clearly from the timing mode to the counting mode. Both the MRT and ROC data are, roughly, half way between the two modes. It is as though the s-deadline training had created still another mode of behavior that we do not understand. One suspects that the order of experimental procedures has a strong effect and, at the least, we should run a group first on the sn deadline and follow it by an s-deadline procedure. In all likelihood, considerable instruction will be needed to shift the mode of behavior.

References

Fitts, P. M. Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, 1966, **71**, 849–857.

Laming, D. R. J. *Information theory of choice-reaction times*. New York: Academic Press, 1968.

Lappin, J. S., & Disch, K. The latency operating characteristic: I. Effects of stimulus probability on choice reaction time. *Journal of Experimental Psychology*, 1972, **92**, 419–427.

Link, S. W., & Tindall, A. D. Reaction times to comparative judgments of line length. Technical Report, No. 32.

Luce, R. D., & Green, D. M. A theory of psychophysics and response times based on inter-arrival times of "neural" pulses. *Psychological Review*, 1972, **79**, 14–57.

McGill, W. J. Neural counting mechanisms and energy detection in audition. *Journal of Mathematical Psychology*, 1967, **4**, 351–376.

Ollman, R. T. Fast guesses in choice reaction time. *Psychonomic Science*, 1966, **6**, 155–156.

Pachella, R. G., & Fisher, D. Hick's law and the speed–accuracy trade-off in absolute judgment. *Journal of Experimental Psychology*, 1972, **92**, 378–384.

Pachella, R. G., & Pew, R. W. Speed–accuracy tradeoff in reaction time: effect of discrete criterion times. *Journal of Experimental Psychology*, 1968, **76**, 19–24.

Pachella, R. G., Fisher, D. F., & Karsh, R. Absolute judgments in speeded tasks: quantification of the trade-off between speed and accuracy. *Psychonomic Science*, 1968, **12**, 225–226.

Pew, R. W. The speed–accuracy operating characteristic. *Acta Psychologica*, 1969, **30**, 16–26.

Schouten, J. F., & Bekker, J. A. M. Reaction time and accuracy. *Acta Psychologica*, 1967, **27**, 143–153.

Taylor, M. M., Lindsay, P. H., & Forbes, S. M. Quantification of shared capacity processing in auditory and visual discrimination. *Acta Psychologica*, 1967, **27**, 223–229.

Tippett, L. H. C. On the extreme individuals and the range of samples taken from a normal population. *Biometrika*, 1925, **17**, 364–387.

Yellott, J. I., Jr. Correction for guessing in choice reaction times. *Psychonomic Science*, 1967, **8**, 321–322.