

Speed and accuracy in comparative judgments of line length*

S. W. LINK and A. D. TINDALL
McMaster University, Hamilton, Ontario, Canada

The paper presents an extension of Henmon's (1906) finding that RT decreases as the difficulty of discriminating a difference between two line segments decreases. It is shown that, when an E RT deadline is imposed on the experimental task, RT remains constant with respect to changes in discrimination difficulty, but that correct response probability increases with increasing difference between two line segments. The data are examined in terms of current theories for the speed-accuracy trade-off.

This paper reports an extension of Henmon's (1906) observation that reaction time (RT) for judging if two line segments are of the same length varies as a function of stimulus difference. Henmon argued, "Differences in sensation should be equal if it takes equal time to perceive them. Again, if the differences are unequal, the greater the difference the shorter the time necessary to perceive it, and on the other hand, the smaller the difference the longer the time." In fact, Henmon supported his argument with results showing that RT did indeed become shorter as the difference between two line segments increased. This finding was investigated further by Birren and Botwinick (1955) and Botwinick, Brinley, and Robbin (1958). In both studies, results comparable to those of Henmon were obtained.

These previous results have fostered theories of RT performance which rest on assumptions similar to those introduced by Wald (1947) as the basis for sequential statistical analysis. In general, it is assumed that a segment of RT is occupied by a sequential sampling process which accrues stimulus information. When a S-controlled criterion on a function of the obtained stimulus information is exceeded, S emits a response. In addition, it is assumed that the rate of accrual is a property of S and is independent of the stimulus. Since two nearly identical stimuli presumably require more information to result in an accurate discrimination than do two quite distinct stimuli, decision process time increases as the two stimuli become increasingly similar.

From the standpoint of sequential sampling theories, S's contribution to RT comes from manipulation of decision criteria defined with respect to information obtained from the stimulus. Variations in decision criteria produce simultaneous changes in both RT and probability

performance. For a fixed discrimination, the concomitant variation in RT and probability performance has been characterized as the trade-off between speed and accuracy. By postulating a change in the S's decision criterion, it is possible to obtain either high accuracy and long RTs or low accuracy and short RTs. In this way, sequential sampling models provide a parsimonious account of numerous RT results.

The experiment reported here is the last of a series of five experiments (Link & Tindall, 1970) designed to investigate theoretical issues concerning the trade between speed and accuracy. These issues will not be treated in detail. Rather, a purely empirical result will be reported which suggests that the notion of a S-controlled decision criterion defined on sampled stimulus information can be profitably examined in a new experimental context.

The Ss participating in this experiment were not experimentally naive. In fact, Ss had participated in as many as 20,000 and no less than 8,000 previous trials that were similar to those in the present experiment. These prior trials each had the common characteristic that S was required to compare two successively presented line segments and determine if the second of the two stimuli was the same or different from the first. In addition, S was encouraged to respond before an E-imposed deadline, or critical time (Fitts, 1966; Pachella & Pew, 1968). During the minimum of 8,000 trials which preceded the present experiment, each S responded in 2,000 trials at each of four different RT deadlines, 300, 380, 460, and 540 msec, respectively. All Ss could maintain a mean RT substantially less than the imposed deadlines, but mean RT varied as a function of the deadline.

In the present experiment the range of the RT deadline was expanded so that Ss operated with instructions to be as accurate as possible, to respond before a deadline of 460 msec, and to respond

before a deadline of 260 msec. Within each of these conditions stimulus similarity was varied.

METHOD

Four right-handed college students were each paid \$22.50 to participate in 15 experimental sessions. A session, of approximately 40 min duration, consisted of 480 trials presented in two blocks of 240 trials each, with a rest break of 5 min between blocks. Each block of 240 trials was divided into four subblocks of 60 trials each with a 10-sec delay between subblocks. Within each of these subblocks, the level of stimulus similarity for nonidentical pairs of stimuli was fixed. Between subblocks stimulus similarity varied at random and was counterbalanced for the two blocks of each session. Ss were informed that there were as many identical as nonidentical pairs of stimuli within a subblock and that after each 10-sec delay following a subblock the level of difficulty of the nonidentical pair would be changed. On each day Ss were instructed to make responses within a fixed time criterion and encouraged to "beat the time" while being as accurate as possible. Finally, tests for sequential dependencies in presentation sequences within subblocks were performed. It was concluded that the sequential structure of the presentation sequences did not deviate from chance.

On each trial S initiated the sequence of trial events by depressing a trial-initiation key (TIK). During the depression of TIK a ready signal, the character R, was presented on a computer-controlled oscilloscope (Tektronix 602-P4 phosphor) for 500 msec. Immediately following the ready signal, the first of two line segments was presented for 200 msec and followed by an interstimulus interval of 200 msec, during which the oscilloscope display screen was blank. The second line segment was then presented until S made a same or different choice by releasing TIK and then depressing one of two (80-g) microswitches. The RTs reported here are measured from the onset of the second line segment to the depression of one of the two microswitches. After responding, S could be informed via the oscilloscope as to whether the response was correct (YES or NO) and whether the time deadline was exceeded (SPEED OK, TOO SLOW). Each feedback display lasted 500 msec and was followed by a new trial when S next depressed TIK. If S released TIK before the presentation of the second line segment, the trial was aborted and restarted with the next depression of TIK.

Five horizontal line segments, 2.0, 1.9, 1.8, 1.7, and 1.6 cm in length, were used throughout the experiment. For two Ss the

*This project was supported by Grant No. 214-1804 from the National Research Council of Canada.

Table 1
Results for Experiment 1

		Speed			460			Accuracy		
		P _c	M _c	M _e	P _c	M _c	M _e	P _c	M _c	M _e
B.T.	4Δs	.688	182	181	.953	336	309	.995	398	348
	3Δs	.708	185	190	.935	345	306	.978	393	368
	2Δs	.658	195	196	.855	348	338	.950	437	457
	Δs	.620	189	189	.765	345	338	.870	507	533
P.L.	4Δs	.735	198	194	.902	366	345	.973	467	460
	3Δs	.663	199	199	.864	363	334	.955	489	480
	2Δs	.620	191	196	.840	366	351	.893	479	511
	Δs	.650	202	197	.744	374	374	.795	507	511
S.D.	4Δs	.743	202	199	.950	356	319	.990	417	394
	3Δs	.675	201	192	.904	360	348	.963	434	433
	2Δs	.588	179	187	.782	358	358	.870	473	454
	Δs	.555	185	187	.686	349	335	.700	474	490
I.C.	4Δs	.693	191	173	.815	266	198	.995	471	503
	3Δs	.653	185	195	.803	271	197	.988	483	499
	2Δs	.655	168	168	.795	264	208	.963	506	526
	Δs	.613	169	169	.728	269	235	.865	571	614
Average	4Δs	.714	193	187	.905	331	293	.988	438	426
	3Δs	.676	192	194	.876	335	296	.971	450	445
	2Δs	.630	183	187	.818	334	314	.919	474	487
	Δs	.609	186	186	.731	334	321	.804	515	537
Standard Error	4Δs		1.6	2.0		1.6	5.1		1.9	11.5
	3Δs		2.1	2.0		1.7	5.2		1.9	11.3
	2Δs		1.6	1.9		2.0	3.5		2.4	10.0
	Δs		1.8	2.2		2.3	2.7		2.8	6.4

2.0-cm line segment was always used as a standard and presented as the first of the two line segments. For the remaining two Ss the 1.6-cm line was the standard. For ease of discussion, the stimulus conditions from most to least similar, will be referred to as Δs, 2Δs, 3Δs, and 4Δs.

Three RT deadlines (260, 460, and ∞ msec) were used during the experiment. During each experimental session only a single deadline was enforced. For those sessions with a 260-msec deadline, S was given feedback on only the speed of response. When no deadline was enforced, only accuracy feedback was provided. For the intermediate deadline, 460 msec, both speed and accuracy feedback followed each trial. Ss were first practiced for two sessions in the 460-msec deadline condition. Then four sessions of data were gathered for the 460-msec condition. Ss were then given the task of responding before a 260-msec deadline during four successive sessions. Another practice session with the 460-msec deadline was given and then four sessions without a deadline were completed. Each session yielded a total of 480 observations. From each subblock within a session, the first 10 trials were ignored, leaving a total of 400 observations per S per session. Ignoring practice sessions, each S contributed a total of 1,600 observations to each of the three deadline conditions. The total number of observations entering into the analysis is then 3 x 4 x 1,600 = 19,200.

All experimental events, measurements of time, recording of responses, and the

presentation of visual displays were controlled by a PDP-8I computer. The computer laboratory system has been described by Link (1969).

RESULTS

Table 1 provides estimates of the marginal probability of a correct response, P_c, the mean correct RT, M_c, and the mean error RT, M_e, for each S-deadline-similarity condition. The results in Table 1 were consistent across Ss and are summarized in Fig. 1. Values in Fig. 1 may be obtained by averaging the estimates for each S found in Table 1. For mean RTs, estimates of standard error based on the within-S variance are given at the bottom of Table 1. In the case of proportions, the

standard error of proportion, given 1,600 observations per point, is less than 0.0125.

In the left panel of Fig. 1, it can be seen that for the two RT deadline conditions, correct response RT remains relatively constant as stimulus similarity decreases. When prompted to be as accurate as possible, Ss displayed the well-known result of Henmon, that RT diminishes as stimulus difference increases. Mean error RTs for the 260-msec deadline condition can be seen to be nearly identical to the mean correct response RTs. In the case of the 460-msec deadline, mean error RT decreases as a function of increasing stimulus difference but is shorter than the correct response RT. For the accuracy condition there are few errors when the difference between the line segments is as large as 3Δs or 4Δs so that differences between mean correct and error RTs are at most suggestive. However, for the Δs and 2Δs differences the mean error RT is generally longer than mean correct RT.

Probability correct shown in the right panel of Fig. 1 continually increases as the difference between the two line segments increases. Of particular interest is the change in probability correct when Ss are required to respond before a deadline of 260 msec. It is apparent that with very fast response times probability performance improves across stimulus differences. Secondly, the form of change in probability performance as a function of stimulus difference clearly differs from the 260-msec to the 460-msec conditions. Finally, even when allowed to respond without a deadline, S's correct probability performance varies from .804 to .988 across stimulus differences.

A more important description of RTs is given in Fig. 2. Reaction time distributions were obtained for correct and erroneous judgments by constructing time intervals of 25 msec and computing the RT

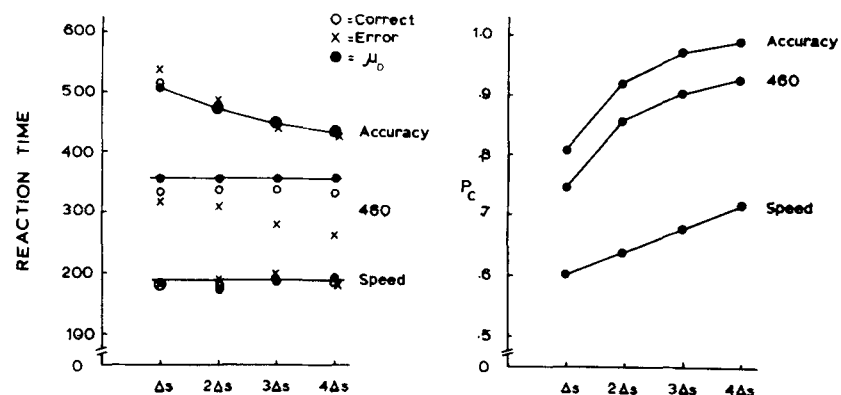


Fig. 1. In the left panel estimates of correct and error mean RTs together with μ_D , the estimate of mean RT corrected for guessing, as a function of differences in line length. The right panel gives estimates of correct response probability.

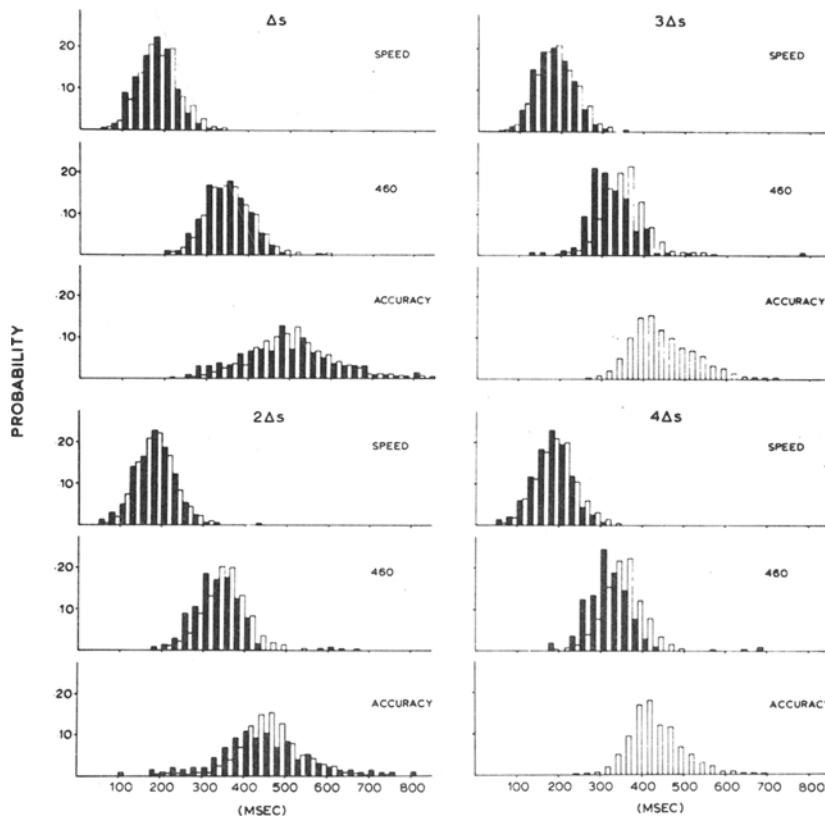


Fig. 2. Correct and error response RT histograms for each deadline similarity condition. Components of correct (open) and error (filled) distributions are placed side by side within each 25-msec interval.

histogram for each deadline-similarity condition. In Fig. 2 the correct and error-response probabilities, given a specified time interval, are placed side by side so that comparisons of these distributions may be made easily. It can be seen that, as the response-time deadline is relaxed, the entire RT distribution appears to shift and the variance tends to increase. This result was also observed for individual Ss. Secondly, for a fixed deadline, but looking across changes in similarity, there is no apparent shift in the RT distributions for the 260-msec deadline. For the 460-msec deadline, the correct response RT distribution is stable across similarity but the error distribution shifts. For the accuracy condition too few observations were available to yield sensible estimates of the 4Δs and 3Δs error RT distributions, and they have been omitted from the figure. In the remaining similarity conditions the errors are spread across the range of the correct-response RT distribution.

DISCUSSION

The results of the present experiment bring into question current assumptions in quantitative theories of RT performance.

These theories can be separated into two camps. First Ollman (1966) and Yellott (1967) have supposed that latency distributions can be associated with psychological states of the Atkinson (1963) or Luce (1963) forced-choice detection theories. In the simplest case, it is assumed that a stimulus presentation results in the activation of either a detection or a guessing state. From the detection state, a correct response results with probability one, while responses from the guessing state depend upon a bias process and not upon the stimulus. These two states have latency distribution functions L_D and L_G , respectively. An RT distribution is then a mixture (in the sense of Teicher, 1960),

$$L(t) = pL_D(t) + (1 - p)L_G(t)$$

where p and $(1 - p)$ are probabilities of entering the detection and guessing states, respectively. The probabilities p and $(1 - p)$ defining the mixing distribution are assumed to be under experimental manipulation. In particular, a trade-off of speed for accuracy can occur if it is assumed that the means associated with these distributions, μ_D and μ_G , satisfy

$\mu_D > \mu_G$. Presumably, by varying a time deadline from long to short, the probability of guessing can be increased and the mean RT, $\mu = p\mu_D + (1 - p)\mu_G$, decreased. The more guesses that occur, the more errors are made, and the shorter the observed mean RT.

Limiting our attention to a stimulus difference of 4Δs, where probability correct in the accuracy condition is nearly perfect, it is possible to subject the above model to a test based on a result due to Falmagne (1968). If, for each deadline similarity condition, the observed marginal RT distributions are assumed to result from a binary mixture of two underlying RT distributions and if two observed distributions, say ℓ_1 and ℓ_2 intersect, i.e., for some t $\ell_1(t) = \ell_2(t)$, then any two observed distributions must intersect at t regardless of the value of the mixing probabilities. From Fig. 2, the marginal distributions of RT for 4Δs, accuracy and 460, have a common point in the range of 375 to 400 msec. For a binary mixture to have given rise to these distributions, it is clear that the 4Δs, accuracy and speed, RT distributions must also have a common point in the same range, but do not.

The motivation for examining the theory for the 4Δs discrimination was to test the model under the most favorable circumstances. That is, it cannot be assumed easily that the probability of a correct response, given an activation of a detection state, is other than unity but for the experimental error of 1% or 2%. However, it can be argued that if response time depends upon the difference between the standard and comparison stimulus Falmagne's test must fail. That is, if no difference between the standard and comparison is presented, the mean RT may be unequal to the mean RT when a difference is presented. The overall RT distributions are then ternary, not binary, mixtures, and the test given above would be inapplicable. A simple calculation shows that Falmagne's criterion is still applicable.

Let S_K ($K = 0, 1, \dots, 4$) represent a presentation of a $K\Delta$ s discrimination, and let L_K be the RT distribution function for detections which occur when S_K is presented. Within all experimental subblocks, S_0 and $S_K \neq 0$, for a fixed $K\Delta$ s discrimination, were presented equally often. Thus the marginal RT distribution function for $K = 4$ may be written as

$$\begin{aligned} L(t) &= \frac{1}{2}[pL_0(t) + (1 - p)L_G(t)] \\ &\quad + \frac{1}{2}[pL_4(t) + (1 - p)L_G(t)] \\ &= p(\frac{1}{2})[L_0(t) + L_4(t)] \\ &\quad + (1 - p)L_G(t). \end{aligned}$$

Under the assumptions of the model, L_0 and L_4 are invariant with respect to experimental manipulations. Therefore a distribution $L_D = \frac{1}{2}[L_0 + L_4]$ exists, Falmagne's criterion is again valid, and the model for the speed-accuracy trade-off fails.

There are two probable causes for failure of the speed-accuracy model. First, it may be argued that the assumed underlying RT distributions, L_D and L_G , are not invariant with respect to changes in the RT deadlines. Secondly, it could be maintained that the assumption that an observed RT distribution results from a mixture of detection RTs and guessing RTs is simply not satisfied by the data. In order to isolate the reasons for failure of the speed-accuracy model, estimates of μ_D will be obtained and shown to vary as a function of the RT deadline. Thus, if the assumption of a mixture is appropriate, then the speed-accuracy model fails because the distribution function, L_D is not invariant with respect to the RT deadlines. To demonstrate that a mixture of detection RTs and guessing RTs produces the observed RT distributions, the data obtained within each RT deadline condition are examined. It is shown below that these data are consistent with the notion that an observed RT distribution consists of a mixture of underlying detection and guessing RT distributions.

An estimate of the mean RT for responses associated with detections, $\mu_D = \frac{1}{2}(\mu_0 + \mu_4)$ was shown by Yellott (1967) to equal $(P_c M_c - P_e M_e)/(P_c - P_e)$, where $P_c = 1 - P_e$ is the marginal correct response probability and M_c and M_e are the means for correct and error RTs. The estimate, μ_D , has the appealing property of being uninfluenced by either E-imposed or chance changes in bias probabilities, L_G , and p . Given the model, this estimate should not change as the RT deadline is experimentally varied. Estimates of μ_D , calculated by averaging estimates for each S, are represented by the filled circles in Fig. 1. It can be seen that in the case of a 4 Δ s difference the estimates of μ_D change systematically as the deadline varies from speed to accuracy.

For cases where there is some probability that the decision process produces errors, it has been shown (Link & Tindall, 1970) that μ_D may still be estimated by $(P_c M_c - P_e M_e)/(P_c - P_e)$. These estimates, also shown in Fig. 1, can be seen to increase as the deadline under which S operates is relaxed. There is little doubt that a model which assumes that the trade of speed for accuracy is produced by changes in the relative number of activations of detection and guessing states must fail.

On the other hand, assuming that within a deadline condition latency distributions for detection and guessing states are invariant with respect to stimulus differences, and that the observed RT distributions are mixtures of these underlying distributions, there is definite support for the assumption that an observed RT distribution results from a mixture of underlying RT distributions. Within the 260- and 460-msec deadline conditions there is no apparent change in the estimates of μ_D as difficulty of discrimination varies. Additional support is provided by an examination of the RT distributions. Suppose that within a deadline and across stimulus differences, responses resulting from a guessing state remain relatively stable with respect to probability and latency. For the detection state, error probability increases as the difference between the two line segments decreases, but the error latency distribution remains fixed, with the result that errors with long RTs are more likely as difficulty of discrimination increases. Thus, for the 4 Δ s condition, the error distribution should consist largely of fast responses, while for the Δ s condition the error distribution should consist of both short and also relatively long RTs. It would be expected that as ease of discrimination increased dissimilar correct and error distributions would separate and become distinct. Precisely this result is shown in Fig. 2 for the 260- and 460-msec deadlines. Finally, an examination of the marginal RT histograms revealed that within both deadlines the four histograms were quite similar, exhibited common points of intersection and, more importantly, showed a noticeable flattening of the peak of the marginal distribution as errors became more frequent.

Apparently, S is capable of two different modes of operation. When required to beat an E-imposed RT deadline, S's performance can be described by a two-state model which assumes that on each trial either a detection state or a guessing state is activated. Given an activation of a state, the probability of a particular response depends upon a decision process for the detection state and a response bias for the guessing state. The response produced by a state determines a unique latency distribution. These latency distributions vary as a function of the RT deadline, but within a deadline remain invariant with respect to changes in stimulus similarity. The S's second method of responding occurs when instructed to be as accurate as possible. The observed performance would be expected by a model based upon accuracy considerations. Such models (Stone, 1960; LaBerge, 1962;

McGill & Gibbon, 1965; and Laming, 1968) maintain that S responds when a S-controlled criterion, defined in terms of information obtainable from the stimulus, is exceeded. Collectively, these models predict an increase in RT as the difficulty of discrimination increases.

The problem with sequential sampling theories of RT when applied to the 460- and 260-msec deadline conditions is that accuracy changes as a function of changes in stimulus difference but RT does not. It may be argued that S adjusts a decision criterion to maintain an RT distribution compatible with the E-imposed deadline. But Ss may have experienced some difficulty in performing the criterion adjustment in the 260-msec condition since no accuracy feedback was provided. Moreover, any change in the decision criterion as a function of RT deadline forces S's behavior to depend upon time and therefore obviates the motivation for sequential sampling models. However, leaving motivation aside, there is definite support for the prediction of sequential sampling models that as the number of stimulus samples increases the variance of the RT distribution also increases. For the distributions shown in Fig. 2, variance increases as the RT deadline is relaxed.

In summary, the assumption that a trade of speed for accuracy results from a binary mixture of two RT distributions associated with detection and guessing performance states has been shown to fail. An extended version of the two-state model, which allows response errors to result from the detection state, appears to hold across variations in stimulus similarity if S is required to respond within an E-imposed RT deadline. The extended model apparently fails if no deadline is enforced.

The conclusion is quite clear. For a fixed deadline, RT distributions associated with the extended two-state model are invariant with respect to changes in stimulus difference while accuracy varies. Furthermore, since for two different deadlines the underlying RT distributions are invariant with respect to changes in stimulus difference, and since accuracy varies as a function of both deadlines and changes in stimulus difference, the S must be controlling the length of his response time. Presumably the temporal process (or processor) controlling RT imposes limitations on the amount of information fed to the decision process. In a sense, then, the temporal process dominates the decision process.

REFERENCES

- ATKINSON, R. C. A variable sensitivity theory of signal detection. *Psychological Review*, 1963, 70, 91-106.
 BIRREN, J. L., & BOTWINICK, J. Speed of

- response as a function of perceptual difficulty and age. *Journal of Gerontology*, 1955, 10, 433-436.
- BOTWINICK, J., BRINLEY, J. F., & ROBBIN, J. S. The interaction effects of perceptual difficulty and stimulus exposure time on age differences in speed and accuracy of response. *Gerontologia*, 1958, 2, 1-10.
- FALMAGNE, J. C. Note on a simple property of binary mixtures. *British Journal of Statistical and Mathematical Psychology*, 1968, 131-132.
- FITTS, P. M. Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, 1966, 71, 849-857.
- HENMON, V. A. C. The time of perception as a measure of differences in sensation. *Archives of Philosophy, Psychology, and Scientific Methods*, 1908, No. 8, 5-75.
- LA BERGE, D. L. A recruitment theory of simple behavior. *Psychometrika*, 1962, 27, 375-396.
- LAMING, D. *Information theory of choice reaction time*. New York: Academic Press, 1968.
- LINK, S. W. A computer controlled laboratory for visual perception and human learning. Third Canadian DECUS Proceedings, Toronto, 1969.
- LINK, S. W., & TINDALL, A. D. An investigation of speed versus accuracy: Comparative line length judgments. Technical Report No. 32, Department of Psychology, McMaster University, 1970.
- LUCE, R. D. A threshold theory for simple detection experiments. *Psychological Review*, 1963, 70, 61-79.
- McGILL, W. J., & GIBBON, J. The general gamma distribution and reaction time. *Journal of Mathematical Psychology*, 1965, 2, 1-18.
- OLLMAN, R. Fast guesses in choice reaction time. *Psychonomic Science*, 1966, 6, 155-156.
- 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.
- STONE, M. Models for choice reaction time. *Psychometrika*, 1960, 25, 251-260.
- TEICHER, H. On the mixture of distributions. *Annals of Mathematical Statistics*, 1960, 31, 55-73.
- WALD, A. *Sequential analysis*. New York: Wiley, 1947.
- YELLOTT, J. I. Correction for guessing in choice reaction time. *Psychonomic Science*, 1967, 8, 321-322.

(Accepted for publication July 27, 1970.)