

The elusive tradeoff: Speed vs accuracy in visual discrimination tasks*

RICHARD G. SWENSSON

Bell Telephone Laboratories, Holmdel, New Jersey 07733

Theoretical models for choice reaction time and discrimination under time pressure must account for Ss' ability to trade accuracy for increased speed. The fast guess model views these tradeoffs as different mixtures of "all-or-none" strategies, while incremental models assume they reflect different degrees of thoroughness in processing the stimulus. Three experiments sought tradeoffs for difficult visual discriminations, using explicit payoffs to control and manipulate pressures for speed and accuracy. Although guessing was pervasive, the simple fast guess model could be rejected; Experiments II and III obtained tradeoffs even when fast guesses were purged from Ss' data. Tradeoff functions fit by several formulations revealed: (1) slower rates of increase in accuracy for more similar stimuli, and (2) substantial "dead times" (80-100 msec slower than detection times) before discrimination responses could exceed chance accuracy. Errors were sometimes faster and sometimes slower than correct responses (depending on S's speed-accuracy trade); the latter effect may reflect a ceiling on S's achievable accuracy. A final discussion examines implications of the results for models of discrimination under time pressure; it suggests modifications in present models, focusing on the random walk model, and describes an alternative "deadline" model.

A choice reaction time (RT) task requires the S to make a discrimination response under pressures for both speed and accuracy. Most choice RT experiments use highly distinctive stimuli, discriminated under strong time pressure. Traditional discrimination experiments emphasize performance accuracy; they typically use highly similar stimuli and little or no time pressure.

Once significant time pressure is introduced into a discrimination task, an S's accuracy seems to depend on the speed of his choice response. If changes in the task requirements affect either performance measure, both mean RTs and error rates change in an inverse or *tradeoff* relation (Smith, 1968). The apparent ability of Ss to trade accuracy for increased speed poses not only an interesting problem for performance theories, but also an important methodological problem. Performance comparisons between tasks or stimuli are made difficult or impossible to interpret when S's mean

RTs and his error rates both differ, particularly if his faster performances are the less accurate. Even when error-rate and RT differences do agree in direction, meaningful quantitative comparisons have to rely on some theoretical model that describes the tradeoff function.

Models that propose to account for speed-accuracy tradeoffs must specify. (1) how S's ability to discriminate the stimuli deteriorates as he shortens his RTs, and (2) how the task constraints determine the particular performance trade that he will adopt. Current models divide naturally into two categories, based on their assumptions about the time course of S's discriminative judgment. Most assume an incremental growth over time in S's ability to discriminate the stimuli (Stone, 1960; La Berge, 1962; Edwards, 1965; Audley & Pike, 1965; Laming, 1968; Vickers, 1970). They represent S's classification of the stimulus as the outcome of a sampling process that operates on brief observations of the stimulus input. Each additional observation increases S's RT by some fixed amount (perhaps very small), but it also increases the probability that his response will be correct. For these *incremental models*, tradeoffs reflect partial processing of the stimulus information—variation in the amount of stimulus input S elects to process before responding.

The *fast guess model* (Ollman, 1966; Yellott, 1967) postulates only two internal states, and assumes that the discrimination occurs in an all-or-none step function. On any trial, S is considered to choose between two

response strategies; he can make a fast guess (with accuracy at chance level) when he detects that the stimulus has appeared or he can decide to process the stimulus information fully and make a slower (but accurate) discrimination response. Mixtures of these all-or-none strategies will map out a continuous speed-accuracy tradeoff in S's mean performance, as he varies the proportion of trials on which he guesses.

Although assumptions about S's stimulus discrimination and his available response strategies are generally well specified by tradeoff models, most make no attempt to represent the task constraints—the pressures for speed and accuracy—which presumably determine his actual performance. All models at least implicitly assume that these task constraints (usually communicated through instructions) affect the importance S attributes to obtaining a faster performance relative to a more accurate one. One way to quantify these constraints is to introduce explicit costs, associated both with S's errors and with increases in his RT (Edwards, 1961). Explicit costs and payoffs reduce the ambiguity in verbal instructions. In the form of large monetary incentives, they also make it reasonable to assume that S will follow his "instructions" *literally*, by attempting to maximize his net payoff. Swensson and Edwards (1971) showed that the incremental and fast guess models can yield highly divergent predictions if S tries to maximize his net income with certain types of explicit costs and payoffs.

Some notation will help to illustrate these divergent predictions. For any specified pair of equiprobable stimuli, let D be the payoff difference between a correct response and an error (equal for both stimuli), and let kRT be S's time cost (proportional to his RT) on each trial. Incremental models assume that any systematic changes in S's choice RT must reflect a change in the length of time he observed the stimulus before responding (i.e., the number of observations of the stimulus input that he sampled). In terms of these models, a time cost of kRT charges S a fixed cost for each sampled observation. For a particular incremental model, the statistical decision or random walk model with optional stopping, Edwards (1965) showed that changes in D (for any fixed k) prescribe performance at different unique points on S's tradeoff function, if his expected payoff is to be maximized. Because this result derives primarily from the model's assumption of an incremental growth in S's expected accuracy with each increase in his observing time, other incremental

*This research was supported in part by the Advanced Research Projects Agency, Department of Defense, monitored by the Air Force Office of Scientific Research, under Contract No. AF(638)-1235 with the Human Performance Center, Department of Psychology, University of Michigan, and in part by the National Aeronautics and Space Administration under Grant NGR 23-005-171 to the Engineering Psychology Laboratory, University of Michigan, monitored by the Ames Research Center, NASA. It also formed part of a dissertation submitted by the author in partial fulfillment of the PhD degree (psychology) in the University of Michigan, 1969.

models would yield similar prescriptions for maximization of expected payoff. Once S has learned how his accuracy and mean time cost covary (through the appropriate feedback), he can use D to determine his best speed-accuracy compromise.

Swensson and Edwards (1971) showed that if S can use only the two strategies assumed by the fast guess model, and if their expected payoffs differ for a given D and k , no tradeoff mixture of these strategies will ever maximize S's payoff. The expected payoff for any strategy mixture must always be some weighted average of those for his two pure strategies. Let D^* be the value of D (not necessarily an integer) that would yield equal expected payoffs for both S's guessing and his discrimination strategies. (D^* depends on k and can be estimated from S's performance; it is proportional to the difference in his mean RTs between these strategies.) To maximize his payoffs, S should guess consistently in conditions with $D < D^*$ and respond accurately on all trials when $D > D^*$.

Using these explicit costs and payoffs, combined with large monetary incentives, Swensson and Edwards found that seven of their eight Ss closely followed the fast guess model's prescriptions for maximizing payoffs. In most conditions, these Ss either responded accurately throughout or they obviously predetermined their responses by a simple rule on all 300-450 trials (e.g., "left on every trial") and performed faster. In virtually all of their remaining conditions, these seven Ss clearly switched between the two response strategies (usually only after many consecutive trials), apparently unsure about which strategy would yield the higher payoff. Dichotomous performance emerged even when the procedure deliberately tried to elicit intermediate tradeoffs (with values of D that were as close as possible to the estimated D^* for each S). These results make the fast guess model a more credible explanation of the tradeoffs reported by many choice RT experiments in which the relative costs of errors and time were ambiguous for Ss. Many Ss in choice RT experiments would consider it "cheating" to predetermine their responses very frequently (or very obviously). For most of Swensson and Edwards's Ss, the large monetary incentives presumably overrode such implicit "ethical" considerations.

Swensson and Edwards used easily discriminable stimuli of two types: (1) a bright square that appeared to the right or left of a vertical line, and (2) a diagonal bar rotated 45 deg clockwise or counterclockwise from

vertical. These easy stimuli produced only small differences in mean RT between the Ss' two strategies: 15-20 msec and 45-70 msec across Ss for the spatially coded and diagonal bar stimuli, respectively. Strong conclusions would have been impossible had not the large incentives made most Ss reluctant to switch strategies frequently. (In fact, one of their Ss did produce ambiguous data consistently: it was unclear whether his tradeoffs resulted from partial processing of the stimulus evidence or from varying amounts of intermittent guessing. Both the fast guess and random walk models fit his data equally well.) Larger differences in mean RT between an S's dichotomous guessing and discrimination strategies would make their presence easier to detect, even if he switched between strategies.

The three experiments reported here attempted to magnify differences in mean RTs by using stimuli that were hard to discriminate. Presumably, pairs of highly similar stimuli would require longer times for Ss to achieve accurate discrimination performance, but they should not affect stimulus detection times. If the fast guess model is correct, the mixtures of all-or-none discrimination strategies would be more easily observed with difficult stimuli; the appearance of such strategies would greatly increase the model's generality. If Ss can perform the partial processing assumed by the incremental models, difficult stimuli would provide a greater range of RTs over which these tradeoffs could be observed. These experiments used the most difficult pair of stimuli that each S could discriminate with no more than 1%-2% errors during his initial sessions. They used large monetary payoffs, and deliberately tried to produce speed-accuracy tradeoffs for each individual S.

EXPERIMENT I

Experiment I followed the procedure used by Swensson and Edwards (1971), holding the cost for time (proportional to S's RT on each trial) constant across conditions. The payoff for each correct response (D) varied between conditions in an attempt to converge on a small range of values around each S's estimated D^* for which he would either: (1) switch between a guessing and a discrimination (accurate) strategy, or (2) produce intermediate tradeoffs (e.g., 25% errors) without resort to guessing.

Method

Display and stimuli. Each S sat before a cathode ray tube display, controlled by a PDP-1 computer. He

depressed two microswitch response keys with his index fingers to begin each trial. This produced a faint fixation cross on the display, over which the stimulus appeared following a random foreperiod interval. To respond, S released only the left or only the right key. If S released the second key within 100 msec of the first, his response was treated as an error. If S responded before the stimulus appeared, the computer suppressed the stimulus and directed S to begin the trial again.

The purpose of the random foreperiod was to discourage S from attempting to estimate when the stimulus would occur, and then to produce a response coincident with it (thus obtaining spuriously fast RTs). The foreperiod distribution was uniform in 1-msec intervals over the range between 1 and 3 sec. A heavy penalty (\$1.25) for each premature response further discouraged S from any attempt to minimize his RT by anticipating the occurrence of the stimulus.

The stimulus appeared centered on the fixation cross; it was a rectangle in one of two equiprobable orientations, with its longer sides slanting northwest-southeast or northeast-southwest. S had to respond with his left finger if the rectangle's length lay along the diagonal slanting upward 45 deg to the left of vertical, and with his right finger if it lay along the diagonal that tilted 45 deg to the right. When the rectangle's length and width were very similar, it looked very much like the same tilted square in either orientation. The smaller the rectangle's length:width ratio, the more difficult was S's discrimination task. Initial informal sessions selected, for each S, the most difficult stimulus pair (smallest length:width ratio) that he could discriminate with no more than 1%-2% errors when told to disregard his RT completely. Table 1 presents the characteristics of the stimuli used for each S; their visual angle depended on S's choice of viewing distance, since his movements were unconstrained. Figure 1 illustrates the easiest pair used (length:width ratio of 55:53); Fig. 1a required a left-finger response, and Fig. 1b required a response with S's right finger.

Since the phosphor of the cathode ray tube faded fairly slowly, traces of the previous stimulus were removed prior to each trial by illuminating all points in the square area around the stimulus seen in Fig. 1. This insured that the stimulus for the next trial would appear on a faintly illuminated and homogeneous field. The location of the fixation point also varied slightly along the horizontal axis from

Table 1
Length:Width Ratios and Actual Dimensions in Inches of Stimulus Rectangles Used for Each S in Experiment I

S	Length:Width	Ratio	Actual Dimensions (Inches)	
			Length	Width
TM	55:53	1.038	1.520	1.464
EK	35:34	1.029	.966	.940
RP	40:39	1.026	1.104	1.078
GO	40:39	1.026	1.104	1.078
MC	40:39	1.026	1.104	1.078
MMcL	50:49	1.020	1.380	1.354
WL	63:62	1.016	1.740	1.712

trial to trial, as an additional precaution against effects due to traces of the previous stimulus.

The stimulus sequences had run length distributions that conformed as closely as possible to the expected distributions in a sequence of 350 independent trials from a binomial process with equal stimulus proportions. This restriction insured that each individual sequence was representative, and contained no fortuitous sequential dependencies. Each sequence was used for only a single session.

Procedure. The Ss were seven paid volunteer men, who received a fixed \$1.50 per hour rate for initial sessions used to select the stimuli. Thereafter, the time cost was introduced, and S's earnings depended entirely on the number of points he won. The net payoff on each trial was $D - kRT$ for a correct response and $-kRT$ for an error. The conditions (each an entire 350-trial session) varied the importance of fast relative to accurate performance by changing D, the payoff difference between a correct response and an error. When D was close to 0, S's net payoff was only slightly higher for a correct response than for an error; both depended primarily on the speed of his RT. As D increased, S's net payoffs depended more and more on how accurately he responded.

The first conditions provided a gradual transition between S's initial sessions (in which he ignored his RTs) and conditions that demanded very fast performances. Since no S had any previous opportunity to use or to discover a strategy of making fast preprogrammed guesses, the E carefully emphasized how speed became increasingly more important to his net payoff as D became smaller. The first of these conditions used a value of D large enough to require high accuracy. D decreased systematically between sessions in large steps until S used a pure guessing strategy (or made about 50% errors) for two successive conditions, and then increased again.

The remaining conditions began with a value of D as close as possible to each S's estimated D^* (based on his

performance in the previous 9-11 sessions). Thereafter, D changed between sessions according to the following rule. If S responded accurately, D decreased for his next condition; if he used a strategy of preprogrammed guesses, D increased. If S switched back and forth between the two strategies or used some unclassifiable strategy, D remained the same up to a limit of three sessions and then changed. Because long-term improvement in S's accurate-performance RTs caused D^* to decrease over sessions, the amount by which D changed for each new condition depended on how S's performance changed. Following each improvement in performance, the next condition tracked the change by approximating the new estimated value of D^* . Table 2 presents the values of D used for each S's real money conditions and shows that these values tended to converge over the later sessions.

Feedback and incentives. After each trial, S received a feedback display that showed, in both vector and numeric form: (1) the payoff for his response (D points if correct, 0 if an error), (2) the cost for his response delay (his RT in milliseconds multiplied by 0.55), and (3) his net gain or loss on that trial. To insure that these explicit costs and payoffs

were meaningful incentives, points were worth ¼ cent each. This meant that S was charged at the rate of \$1.38/sec for the time he took to respond. To compensate for the large variation in S's net payoff produced by varying D between conditions, he received a handicap in points at the beginning of each session. At the end of the session, his total net points either added to or subtracted from this handicap.

The relations between S's performance, payoffs, and monetary earnings (as presented here) were *thoroughly* explained and illustrated to each S, and were adhered to rigidly. The magnitude of his handicap for each condition adjusted S's net payoffs to yield an expected net payoff of \$1.50/h. Calculation of this handicap used S's most recent performance with either an accurate or a guessing strategy (whichever had the higher expected payoff for that value of D). This procedure provided an incentive for S to improve his speed and accuracy in order to increase his payoffs; it also enabled him to recoup the occasional large loss he incurred from a poor performance.

Results

It was possible to develop criteria that could separate, for each S, trials on which he preselected his response and guessed from other trials on which he apparently attempted to respond accurately. Both verbal reports and data from all seven Ss indicated that they easily discovered and used a guessing strategy in the early experimental conditions, when the payoff for a correct response decreased until their net payoff depended primarily on the speed of their RTs.

Two criteria were used to identify preprogrammed guesses among S's

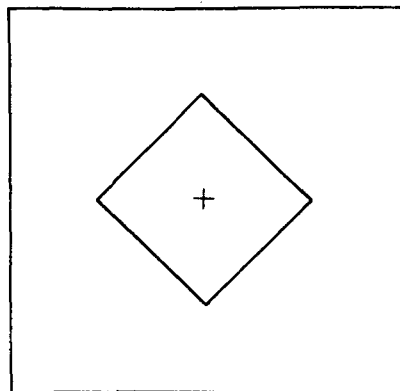


FIG. 1a

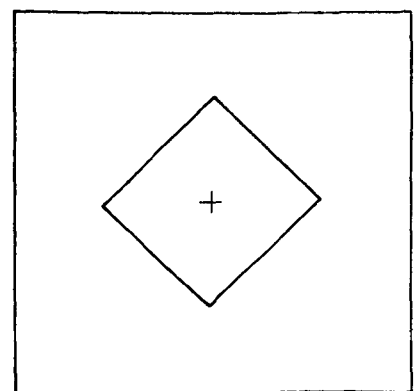


FIG. 1b

Fig. 1. The easiest pair of stimuli used (length:width ratio of 55:53). Figure 1a required a left-finger response and Fig. 1b required a response with the right finger.

Table 2
Difference in Payoff Between Correct Responses and Errors (D) Used
for Each S and Condition in Experiment I

Condition	S						
	TM	EK	RP	GO	MC	MMcL	WL
1	800	800	800	400	800	800	800
2	500	600	600	200	600	500	500
3	200	400	400	200	400	200	200
4	200	200	200	300	200	200	200
5	300	200	200	400	200	300	300
6	400	300	300	500	300	400	400
7	500	400	400	600	400	500	500
8	600	500	500	500	500	600	600
9	800	600	600	400	600	800	800
10	300	600	600	300	600	1000	1000
11	350	300	300	350	200	1000	250
12	300	250	200	350	200	300	225
13	275	250	225	400	150	250	275
14	290	225	225	350	150	275	260
15	280	225	250	300	175	260	280
16	270	250	250	275	200	260	300
17	260	275	235	290	225	240	310
18	260	250	230	305	200	220	325
19	240	250	200	300	185	220	275
20	240	250	225	310	185	210	290
21	225	250	215	350	195	225	280
22	250	235	220	350		215	270
23	250	225	225	310		215	260
24	240	225	220	305		220	230
25	230	210	225	280		225	220
26	235	200	222			222	225
27	235	190	220			224	230
28	232	180	221			224	235
29	230	170	221			220	240
30	225	170	200			210	250
31	220	170				210	260
32	225	180				210	270
33	235					200	270
34	230					205	265
35	235					215	255
36	240					215	245
37	250					225	

responses. Criterion 1 was identical to that used by Swensson and Edwards (1971). It separated all blocks of consecutive trials within which S selected his response by a consistent and mechanical rule that was unrelated to the identity of the stimulus on that trial. Three such rules were easily identified, each favored by at least one S: strict repetition of the same response, strict alternation between left and right responses, and making the response appropriate to the stimulus that occurred on the previous trial (stimulus following).

To be classified as preprogrammed by Criterion 1, a block of consecutive responses had to meet two conditions: (1) the mechanical rule for response selection had to be unbroken and (2) it had to contain at least three errors (i.e., S ignored at least three successive opportunities to abandon the rule and respond as directed by the stimulus). The invariable sudden change in level of S's RTs was used only to help resolve any remaining ambiguity about precisely where each block of preprogrammed responses began or ended. Individual blocks of preprogrammed trials ranged in length

from entire sessions (350 trials) to less than 20 consecutive trials, but were usually fairly long. Even excluding the entirely preprogrammed sessions, the mean length of these preprogrammed blocks ranged from 77 to 239 trials over the seven Ss.

Unlike Ss in the earlier experiments (Swensson & Edwards, 1971), however, only two of these seven Ss confined their use of a guessing strategy to long sequences of trials with a simple mechanical rule for determining which response to make. The other five Ss at least occasionally embedded very fast responses (with a high frequency of errors) among trials on which they apparently attempted to respond accurately, even in sessions having low overall error rates. This intermittent guessing required the use of a second criterion for identifying individual preprogrammed guesses.

Fortunately, because of the relatively long times necessary for Ss to achieve accurate performance with these difficult stimuli, the RT distributions for each S (pooling over all responses and conditions) were clearly bimodal, with few responses in the region between 240 and 280 msec.

Very few of any S's responses that were identified as preprogrammed by Criterion 1 were slower than 250 msec (1.5% for one S and less than 0.3% for the remaining six Ss after the first 10 experimental sessions). To identify the occasions on which S took a guess and made a preprogrammed response when he detected the stimulus, Criterion 2 separated and pooled all trials with RTs faster than 250 msec. It was applied in each condition, to all responses that were not classified as preprogrammed by Criterion 1.

Figures 2a-2f and 3a present mean RTs and error rates within each condition for both S's accurate and his preprogrammed modes of responding. Open points represent sessions in which S used either an accurate or a guessing (Criterion 1) strategy on all 350 trials. Solid points indicate mean RTs and error rates for each strategy in sessions that included both types of performance. For such mixed-strategy sessions, several points are plotted: solid triangles for preprogrammed responses identified by Criterion 1 (if any), solid square points for guesses identified by Criterion 2 (all responses with RT < 250 msec), and solid circles for S's remaining responses.

Because Criterion 2 was applied to all sessions, circular points (representing S's accurate performances) contain only responses that were 250 msec or slower. But the large number of conditions in which all 350 responses were slower than 250 msec (open circles) indicates that this procedure could not have had much effect on estimates of Ss' accuracy-strategy RTs. Mean guessing RTs identified by Criterion 2 (solid square points) tended to be variable because they frequently contained only a few observations. However, the general levels of Ss' mean RTs (about 200 msec) and error rates (about 50%) were comparable for preprogrammed responses isolated by both criteria.

Solid lines in Figs. 2a-2f and 3a represent weighted moving averages of mean RTs and error rates over each set of three adjacent conditions. Most of the deviant points in these figures simply reflect the greater variability of estimates based on only a small number of observations; they had little effect on the moving averages. Vertical bars in Figs. 2a-2f and 3a periodically indicate the variability of the RT distributions for each response strategy about its moving average. They represent plus and minus one "average" standard deviation, calculated as the square root of the weighted mean variance for the adjacent three conditions. Circled asterisks in Figs. 2d, 2e, 2f, and 3a indicate the part of a single condition for four of the seven Ss that could not

be clearly separated by the two criteria into accurate and preprogrammed-response performance. They were not included in the moving averages. Error rates and mean RTs for these trials fell in between those for the two types of performance. These intermediate performances occurred only on the first occasion that each S experienced a condition with a small value of D (demanding very fast responding) and may reflect the Ss' inefficient and sporadic first attempts to apply a guessing strategy.

Mean RTs for accurate performance decreased considerably over sessions. With the possible exception of the final conditions for S E.K. (Fig. 2f), these decreases reflect absolute improvement in Ss' discrimination times, since error rates showed little change. By the end of the 20-37 conditions, differences in mean RTs between preprogrammed-response and accurate performance ranged between 135 and 210 msec.

Error rates among responses with $RT \geq 250$ msec were about 0.10 for most Ss, substantially higher than each S's error rate in his initial sessions without time pressure. These increases in error rate (above 0.01-0.02) during the experimental conditions must be attributed either to a change in Ss' discrimination accuracy or to the presence of guesses. Since its all-or-none assumption requires S's discrimination performance to remain invariant, the fast guess model must postulate that some proportion of his intermittent guesses were slower than 250 msec. When these guesses were misclassified by Criterion 2, they would have inflated the error rate attributed to discrimination responses. Although this argument can be made to account for any set of data, it implies constraints on S's RT distribution for guess responses. The all-or-none assumption can be discredited if these constraints are sufficiently implausible.

Consider each S's responses with $RT \geq 250$ msec. The fast guess model's equations provide an estimate of q , the proportion of these responses that were guesses:

$$\frac{q}{1-q} = \frac{2p_e - 2\epsilon}{1 - 2\epsilon}, \quad (1)$$

where p_e is the observed error rate and ϵ is the error probability assumed for S's "true discrimination" responses. Multiplying this estimate of q by the number of responses with $RT \geq 250$ msec produces an estimate of the number of guesses that were 250 msec or slower. This leads to an estimate of the proportion of all guesses with $RT \geq 250$ msec, under

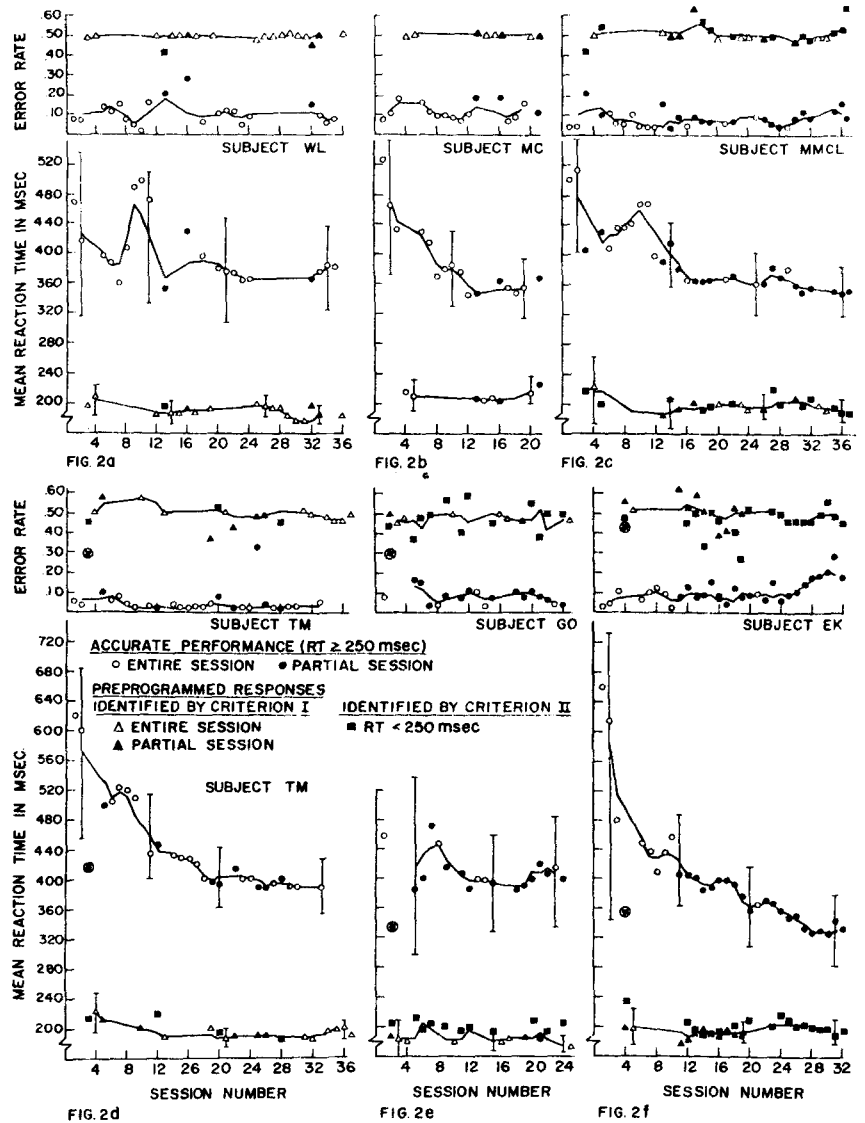


Fig. 2. Mean RTs and error rates for both accurate ($RT \geq 250$ msec) and guessing performance of six Ss in Experiment I. Open and filled points differentiate entire sessions under the same strategy from mixed sessions, in which S used both strategies. Triangular points represent preprogrammed guesses identified by the blocking procedure (Criterion 1); solid square points represent all responses faster than 250 msec (Criterion 2). Solid lines indicate moving averages over adjacent sessions, and vertical bars periodically indicate RT variability. The single performance indicated by a circled asterisk for three Ss did not separate clearly into the two strategies.

the assumption that responses faster than 250 msec were entirely guesses.

Table 3 presents for each S estimates of: q for responses slower than 250 msec, the number of these guess responses, and the implied proportion of guesses that had $RT \geq 250$ msec. These estimates are based on all of each S's data after Condition 10, and they assume that $\epsilon = 0.02$ (if ϵ were assumed to be less than 0.02, all estimates in Table 3 would be increased). To account for the observed error rates if each S could

have achieved an error rate of 0.02, the fast guess model with an all-or-none assumption must assume that Ss guessed rather frequently, and that between 26% and 96% of these intermittent guesses were 250 msec or slower.

Discussion

The data from this experiment seemed unambiguous; they contradicted the incremental models' prediction that Ss can (and will, when encouraged by payoffs) process only

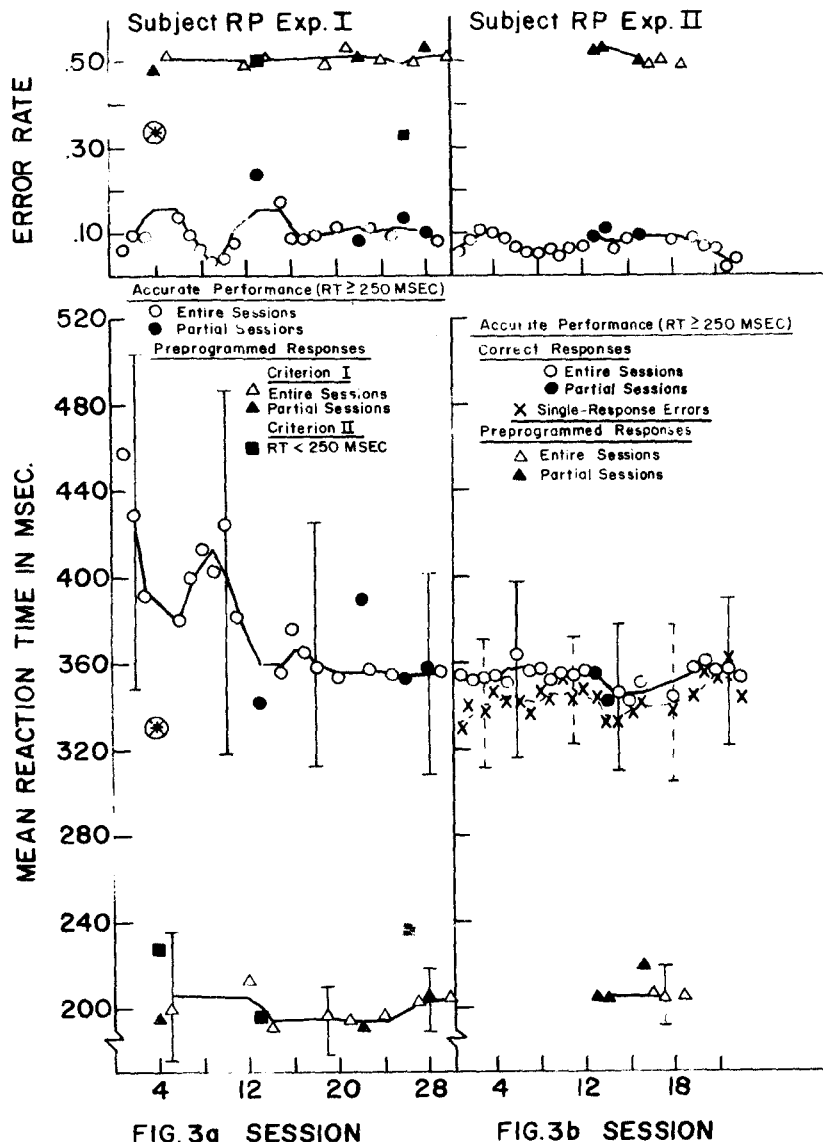


Fig. 3. Mean RTs and error rates for S.R.P.'s accurate performance and preprogrammed guesses in both Experiments I (Fig. 3a) and II (Fig. 3b). Open and filled points differentiate conditions in which he used the same strategy on all 350 trials from mixed sessions, in which he used both strategies. Figure 3b separately indicates mean RTs for correct responses (circles) and for single-response errors (crosses). Solid lines indicate moving averages over adjacent conditions; in Fig. 3b, solid and broken lines differentiate moving average RTs for correct responses and for single-response errors for R.P.'s accurate performance ($RT \geq 250$ msec). The single performance in Fig. 3a indicated by a circled asterisk did not separate clearly into the two distinct strategies.

intermediate amounts of evidence about the identity of the stimulus before responding. Each S used only the two types of strategies assumed by the fast guess model. Five of the seven Ss did not always group their preprogrammed responses into long sequences of consecutive trials, but the clear separation between RT distributions for the two strategies permitted identification of the preprogrammed guesses that Ss embedded among predominantly

accurate trials. Even after large amounts of practice with these difficult stimuli, mean RT differences between strategies were more than three times greater than those previously obtained with easily discriminable stimuli (Swenson & Edwards, 1971).

Only one feature of the data seemed inconsistent with the fast guess model: Ss' relatively high error rates (8%-12%), even when all responses faster than 250 msec had been

removed. This is embarrassing for the model's all-or-none assumption, since each S could and did achieve performance with no more than 1%-2% errors when he ignored his RT in the initial sessions used to select the stimuli. To account for error-rate increases of this magnitude, the model is forced to the unpalatable assumption that between 26% and 96% of Ss' intermittent guesses were slower than 250 msec, whereas obviously preprogrammed responses (identified by Criterion 1) were almost never this slow.

Thus, the fast guess model seems to be required to postulate a different RT distribution for each of S's two types of guess responses. Both distributions must be nearly identical in form below 250 msec (since the observed RT means and standard deviations were very similar), but the distribution for intermittent guesses would have to be bimodal. Although each S made relatively few responses whose RT was 240-280 msec, the model must assume a large proportion of intermittent guesses slower than 250 msec. The model must also assume that Ss made rather frequent use of this inefficient guessing strategy, despite their keen interest in maximizing their payoffs with these large incentives, and the ready availability of their preprogrammed guessing strategy.

An alternative explanation of the higher error rates in the experimental conditions would: (1) reject the assumption of all-or-none discrimination, (2) assume that few or no guesses were slower than 250 msec, and (3) suggest that Ss were able to compensate for their higher error rates by achieving generally faster discrimination RTs. Although incremental models cannot easily explain the absence of less accurate (but nonguessing) performance in these data, they may explain Ss' apparent refusal to respond as accurately as they could. As a consequence of the experimental procedure, the payoff for a correct response (D) varied over only a small range for each S's later conditions, about the value that would make him switch between a "guessing" and an "accurate" strategy. Perhaps achieving very low error rates with these difficult stimuli would have demanded increases in RT that were prohibitively expensive for these intermediate values of D. If so, larger values of D would increase the cost of making an error relative to Ss' cost for time and should force them to respond more accurately in order to maximize their payoffs.

EXPERIMENTS II AND III

Experiments II and III pursued the apparent contradiction between the fast guess model's assumption of

all-or-none discrimination and the change in Ss' performance accuracy that occurred when time pressure was introduced in Experiment I. Both experiments tested the possibility that because of the particular payoffs and time-cost function used in Experiment I, the Ss had refused to vary the amount of stimulus evidence they processed before responding—even though they could have performed such partial processing. Although the two studies were chronological, Experiment III being based upon the results of Experiment II, they will be reported and discussed together.

Experiment II used two experienced Ss from Experiment I, but varied over a large range the incentive for accurate relative to fast performance (D). The principal question was whether the Ss, when confronted with pressure from the payoffs to achieve very low error rates, would be forced to respond more slowly than they had in Experiment I. Following this, Experiment II pursued the effects of stimulus discriminability by giving one S extensive training on a new, extremely difficult pair of stimuli.

Experiment III tested a hypothesis (derived from Experiment II) that Ss had refused to exhibit any performances with high error rates that were free of guesses, because such performances were dominated in expected payoff by their guessing strategy. This *fixed-cost* hypothesis assumed that S had to pay a substantial fixed cost in time, with no improvement in his accuracy, for any attempt to base his response on the stimulus identity. Experiment III modified the time-cost function, intending to charge S only for his actual observing time, to make the hypothesized savings in time enabled by a guessing strategy irrelevant to his payoffs. In terms of the incremental models, removal of this large hypothesized cost (associated with taking the first stimulus observation) would make S's cost per observation independent of the number of observations he sampled before responding. In Experiment III, S's cost for time began only when his RT exceeded t msec (a free delay). If the fixed-cost hypothesis were correct, large variations in D should then have made Ss willing to exhibit formerly suboptimal tradeoffs with both high and very low error rates.

Method

Experiment II. The two Ss were R.P. and E.K., who had performed in 20-32 sessions during Experiment I. During his final 10-15 sessions in Experiment I, S R.P.'s accurate performance had remained fairly

Table 3
Fast Guess Model Estimates of the Frequency (\bar{q}) and Number of Guesses Among Responses With RT \geq 250 Msec, Number With RT $<$ 250 Msec, and the Estimated Proportion of Guesses Slower than 250 Msec

S	For RT \geq 250 Msec		N With RT $<$ 250 Msec	Estimated Proportion of Guesses With RT $>$ 250 Msec
	\bar{q}	Estimated Number of Guesses		
EK	.167	792	1341	.371
RP	.180	687	111	.861
TM	.012	69	191	.265
GO	.112	419	162	.721
MMcL	.110	606	1280	.321
MC	.187	397	18	.957
WL	.175	698	30	.959

stable, with about 10% errors and mean RTs of 350-360 msec. S E.K. displayed the greatest change in his accurate performance during Experiment I. Figure 2f shows that his early accurate performance had low error rates and the slowest mean RTs, while his accurate performance in the final sessions had the fastest mean RTs (325 msec) and highest error rates (up to 0.20) of all seven Ss.

Except for the variation in D over conditions, the apparatus, procedures, and stimuli were identical to those used in Experiment I. Over successive conditions, D increased and decreased systematically, from a value small enough to produce consistent preprogrammed guessing (or 50% errors) to values as high as 1½ times as large as Ss' cost per second of response time. Under these large values of D, an error cost S as much in his net payoff as an RT of 1,450 msec; one error more or less could change his net payoff by as much as 800 points (\$2.00). The 26 conditions for S R.P. and 32 conditions for S E.K. traversed this range of D in both ascending and descending order.

Following these 32 conditions, S E.K. performed in 56 conditions with a new, more difficult pair of stimuli. These had a length:width ratio of 85:84 (1.012) and measured 2.348 x 2.320 in. Between conditions, D changed systematically from values low enough to produce 50% errors to values almost 20 times as large as his cost per second of RT. Even though points were devalued from 1/4 to 1/10 cent per point, a single error with these extreme values of D changed S's net payoff by as much as \$10.00. Although S's income from individual sessions varied a great deal (and he frequently lost money), he averaged about \$2.25/h over the entire 56 sessions.

Experiment III. The three Ss, T.M., M.McL., and W.L., had produced performances that were stable during their final 10-15 sessions in Experiment I and devoid of any speed-accuracy tradeoffs. The apparatus, stimuli, and most

procedural details were identical to those used in Experiment I. Under the free delay modification, Ss received feedback about the duration of their RT only when it exceeded t msec (with $t = 250$ or 300 msec). If their RT was less than t msec, the feedback displayed only 0 cost for time. Since Ss were charged for time in excess of t at the same rate as before (0.55 points/msec), the cost function was $0.55(\text{RT} - t)$ for $\text{RT} > t$ and 0 for $\text{RT} < t$.

Ss W.L. and M.McL., whose accurate performance had error rates of about 0.10 during their final sessions in Experiment I, received conditions with both large and small values of D. The maximum value of D used was 5.5 times the cost per second for RT in excess of t (\$7.50 lost for each error). Because S T.M. had the lowest error rates in Experiment I when he responded accurately (0.01-0.02), he received only conditions with small values of D to encourage faster and less accurate performance.

All three Ss received one training session (with $t = 250$ msec), at an hourly rate of pay to accustom them to the new time-cost feedback. This session contained three blocks of 100 trials each under a high, moderate, and low value of D. During this practice session, it became obvious to all three Ss that their guessing strategy would always yield 0 cost for time and a guaranteed average net payoff of D/2 points per trial. To minimize the Ss' temptation to use this "safe" guessing strategy, D was reduced gradually, so that one of their previous "nonguessing" performances always predicted an expected payoff higher than D/2 points per trial. Ss W.L. and M.McL. received 18 and 19 test conditions of 350 trials, all with $t = 250$ msec. The free delay was 250 msec for S T.M.'s first five sessions, but it increased to 300 msec in Sessions 6-14.

Results

Response strategies. When they did not preprogram their responses, four

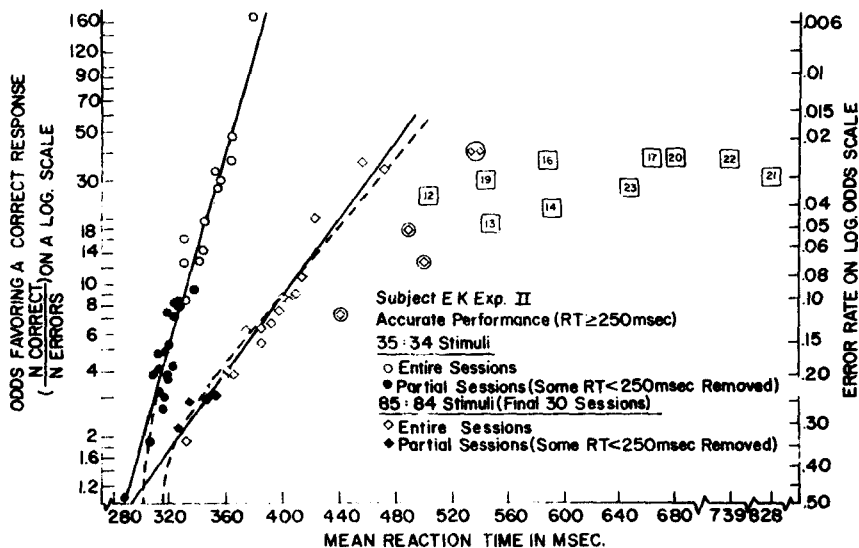


Fig. 4. S E.K.'s speed-accuracy tradeoff performance with both discrimination tasks in Experiment II. The figure plots odds transformations of S's accuracy (ratio of number of correct responses to number of errors) on a log scale against his mean RT in each condition for discrimination responses only (RTs \geq 250 msec). Open points represent conditions that contained no responses faster than 250 msec; filled points represent conditions from which some guesses (RT $<$ 250 msec) were removed. Solid lines and broken curves indicate two different tradeoff functions fitted to each set of data (see general discussion). Numbered square points show E.K.'s highly accurate performance (5% errors or less) in his first 26 conditions with the more difficult 85:84 stimuli. Neither these data nor the later conditions indicated by circled points were used in fitting the tradeoff functions.

of the five Ss in Experiments II and III produced data that showed clear evidence of speed-accuracy tradeoffs (presented in the following section). In both experiments, all Ss used a guessing strategy on at least some trials during conditions with small values of D. For both Ss in Experiment II, these guesses had mean RTs that were virtually identical to those in Experiment I. S E.K.'s guesses had to be identified by separating all responses faster than 250 msec (Criterion 2). With his easier pair of stimuli, guess responses had an error rate of .475 and a mean RT of 198 msec; in his more difficult discrimination task, these values were .502 and 196 msec.

In Experiment III, Ss' guesses were slower than they had been in Experiment I. S T.M. always repeated his left or right response for relatively long blocks of consecutive trials whenever he abandoned any attempt to respond accurately. Although these preprogrammed responses were obviously independent of the stimulus identity, and had an error rate of 0.481, their mean RT was 218 msec—more than 20 msec slower than his mean guessing RTs in Experiment I. The same phenomenon occurred in the data for Ss M.McL. and W.L., whose guesses had to be identified by their speed (using

Criterion 2). Of all responses faster than 250 msec, error rates were 0.512 and 0.493 for M.McL. and W.L., but their mean RTs were 212 and 224 msec. These mean times were about 20 and 40 msec slower than in Experiment I. The increased RTs for guesses in Experiment III could have resulted from the absence of any pressure to respond much more quickly than the free delay interval (250 or 300 msec), together with the lack of any feedback about how fast or slow these "zero-cost" responses actually were.

Of the five Ss, only S R.P. in Experiment II did not vary his accurate performance, even for the largest values of D used. Figure 3b plots R.P.'s data as done in Experiment I, except that it separately indicates his mean accurate performance RTs for correct responses and for single-response errors. As a comparison between Figs. 3a and 3b shows, his performance did not change appreciably between Experiments I and II. The decrease in error rates for his accurate performance during the final few sessions in Experiment II seemed to reflect only further improvement in performance, since it was not accompanied by any change in his mean RTs. Single-response errors (excluding trials when he made both responses less than 100 msec apart)

comprised 88% of R.P.'s total errors under the accurate strategy in Experiment II. Figure 3b shows that their mean RTs were consistently 10-15 msec faster than his correct responses.

Speed-accuracy tradeoffs. One S in Experiment II and all three Ss in Experiment III exhibited speed-accuracy tradeoffs as D changed between conditions. Figures 4 and 5a-5c plot each S's accuracy against his mean RT for all nonguessing responses in each condition (i.e., those with RT $>$ 250 msec, by Criterion 2). Accuracy is plotted as an odds transformation of S's error rate (proportion of correct responses divided by proportion of errors) on a log scale. This transformation reflects how much more likely S was to make a correct response than an error; it provides an essentially unbounded scale for accuracy, with chance performance (50% errors) at the baseline. The equivalent error rates are indicated at the right in each figure. Figure 4 presents S E.K.'s data from Experiment II for both the 32 conditions with the easier 35:34 stimuli (circles) and for his final 30 conditions with the difficult 85:84 stimuli (diamond-shaped points). Figures 5a, 5b, and 5c present the data for the three Ss in Experiment III. Open points refer to conditions in which all responses were slower than 250 msec; solid points represent conditions in which some guesses were identified (by Criterion 2) and removed. Triangular points in Fig. 5c indicate S T.M.'s performance in his final nine accurate conditions of Experiment I.

The curves in Figs. 4, 5a, 5b, and 5c represent functions fitted to the Ss' mean RTs, against different transformations of their accuracy. They will be discussed later. However, for all five sets of data, these fitted functions (and virtually any other reasonable extrapolations of these data points) intercept chance-level accuracy at RT values much slower than the times these Ss achieved when they made preprogrammed guesses. It was this "dead time" interval, noted for S E.K. in Experiment II (Fig. 4), that led to the "fixed cost" hypothesis and the procedure used in Experiment III. Figure 4 also indicates that the more difficult discrimination task greatly reduced the rate at which S E.K.'s accuracy increased as he slowed his mean RTs. However, extrapolations of E.K.'s data for both discrimination tasks in Fig. 4 appear to intercept chance-level accuracy at about the same RT value.

With the easier 35:34 stimuli, Fig. 4 shows that E.K. could reduce his error rate to less than 0.01, but there was

evidence of a ceiling on his accuracy for the more difficult 85:84 stimuli. The numbered square points in Fig. 4 indicate his performance in the most accurate of his conditions prior to Session 27 (those with error rates less than 0.05). Although in some of these earlier sessions E.K. drastically slowed his mean RTs (to 830 msec), he was unable to reduce his error rates below 0.03. The ordering of the session numbers shown in Fig. 4 indicates that these very slow performances cannot be dismissed simply as the result of insufficient practice. They occurred during his first experiences with payoffs that demanded extreme accuracy, when making an error was more costly than delaying his RT for 5-20 sec. Many of these individual RTs were longer than 2-3 sec. Figure 4 shows that E.K. could achieve error rates as low as 0.03 with much faster mean RTs than these; during later conditions, he rarely allowed his individual RTs to extend beyond 1 sec, for even the most extreme values of D.

Test of the all-or-none assumption. Although the data in Figs. 4 and 5 contain only responses with $RT \geq 250$ msec, the fast guess model might still be able to explain each S's tradeoff with its all-or-none assumption if some intermittent guesses were slower than 250 msec. However, in any condition, the number of these hypothesized guesses must account for both the observed decrease in mean RT and the observed

Table 4
Estimates of "True Discrimination" Performance (ϵ and RT_d), Tests of the Fast Guess Model, and Mean Estimates of the Guessing Frequency (q) With Lower and Upper Bounds [q (Min) and q (Max)]

S	Five Most Accurate Conditions			Remaining Conditions				
	ϵ	RT_d	N	$\bar{q} < \bar{q}$ (Min)	\bar{q} (Max) < \bar{q} (Min)	Mean Estimates		
						\bar{q} (Min)	\bar{q}	\bar{q} (Max)
EK (35:34)	.022	364	27	21	19	.300	.201	.236
EK (85:84)	.029	465	18	14	13	.311	.291	.332
TM*	.018	401	14	14	11	.236	.091	.120
MMcL	.027	371	14	13	12	.353	.237	.278
WL	.030	432	13	12	12	.414	.236	.282

*Estimates of ϵ and RT_d for S TM used the most accurate five of his final nine conditions in Experiment I.

increase in error rate. This section will reject this hypothesis by using the fast guess model's equations to obtain two irreconcilable estimates of q , the frequency of such guesses.

Two independent estimates of q can be obtained from S's performance in each condition, given certain assumed parameters of his "true discrimination" and guessing performance. Equation 1 estimates q from the error rate, given the error probability for a discrimination response (ϵ). Equation 2 estimates q from RT , the observed mean RT, given RT_d and RT_g , the mean time for discrimination and guess responses:

$$q = \frac{RT_d - \bar{RT}}{RT_d - RT_g} \quad (2)$$

The fast guess model with an

all-or-none assumption can be rejected if these two estimates of q differ excessively.

Notice that Eq. 2 estimates smaller values for q as RT_g decreases, because fewer guesses are required to account for the decrease in mean RT as guesses become faster. Consider only S's responses in each condition that have $RT \geq 250$ msec; obviously, 250 msec is the minimum value that can be assumed for RT_g among these responses. Consequently, with RT_g set at 250 msec, Eq. 2 estimates $q(\min)$, the smallest possible proportion of guesses that could account for this mean RT given RT_d . Equation 1 yields an estimate of $q(\max)$ under the assumption that all errors were produced by guesses ($\epsilon = 0$): $q(\max) = 2p_e$.

Estimates of ϵ and RT_d for each S's "true discrimination" responses were obtained by averaging performance over his five most accurate conditions (seen in Figs. 4 and 5), none of which had any responses with $RT < 250$ msec. These estimates of ϵ and RT_d (shown in Table 4) were applied in each of the S's remaining conditions to obtain q (from Eq. 1), $q(\max)$, and $q(\min)$. Table 4 presents the results of this analysis for each of the five Ss: mean values of these estimates, the number of conditions in which $q < q(\min)$, and the number in which $q(\max) < q(\min)$.

The estimated q from Eq. 1 is inconsistent with the all-or-none assumption if it is smaller than $q(\min)$. An even more conservative test of the fast guess model, less likely to reject the all-or-none assumption, considers the assumption to fail only if $q(\max) < q(\min)$. The assumption was rejected in a large majority of each S's individual conditions, even by the more conservative test. Examination of Eqs. 1 and 2 shows that these systematic discrepancies in estimates of q imply performances that were more efficient than could be accounted for by the fast guess model: Ss' mean RTs decreased more than

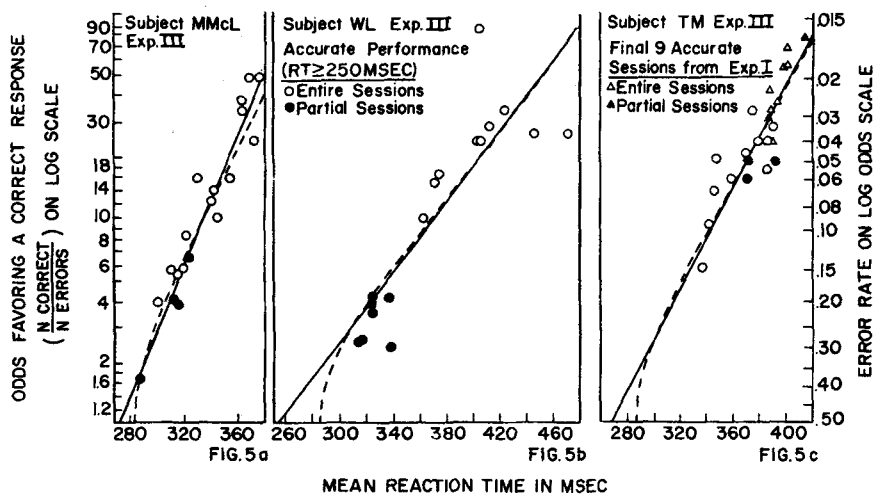


Fig. 5. Speed-accuracy tradeoff performance for each of the three Ss in Experiment III. The figures plot odds transformations of S's accuracy (ratio of number of correct responses to number of errors) on a log scale against his mean RT in each condition for discrimination performance only ($RTs \geq 250$ msec). Open points represent conditions with $RT \geq 250$ msec on all 350 trials, filled points represent conditions from which some preprogrammed guesses ($RT < 250$ msec) were removed. Triangular points in Fig. 5c indicate S T.M.'s final nine sessions in Experiment I. Solid lines and broken curves indicate two different tradeoff functions fitted to each set of data (see general discussion).

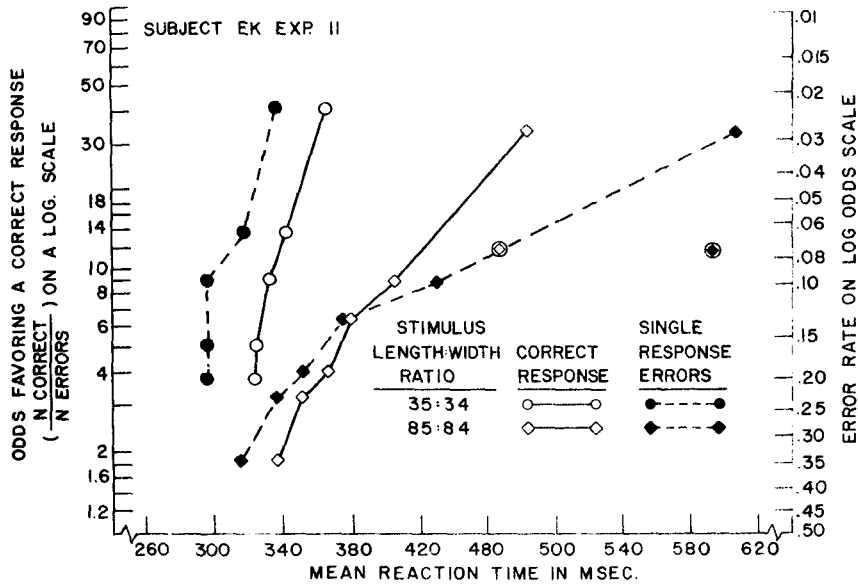


Fig. 6. S.E.K.'s mean RTs for both correct responses and single-response errors in grouped sessions within each discrimination task. The ordinate plots odds favoring a correct response (ratio of number of correct responses to number of all S's errors) in these pooled sessions. Circled points for the 85:84 stimuli represent the pooled data from the five circled conditions in Fig. 4.

their error rates should have increased, according to the model.

Correct response and error RTs. Two types of errors could occur when Ss did not preselect their response, but attempted to base it on the stimulus identity (performance with RTs > 250 msec). An S could make the inappropriate response only (single-response errors), or he could release the other response within 100 msec of his first (double-response errors). Single-response errors were by far the most frequent; they accounted for 92% and 98% of all S.E.K.'s errors with the two types of stimuli in Experiment II, and 98%, 88%, and 62% of all errors for Ss W.L., T.M., and M.McL. in Experiment III. Most of the double-response errors occurred in Ss' less accurate (faster) performances, and virtually all were occasions on which S made the inappropriate response first. In spite of the fact that the first of these double responses ostensibly differed from single-response errors only by the subsequent occurrence of the correct response within 100 msec, their mean RTs were consistently slower for all Ss. The frequency of these double responses was too low for any more detailed comparisons to be meaningful.

Figure 6 plots S.E.K.'s mean RTs for correct responses and for single-response errors, for both types of stimuli in Experiment II, against the odds favoring a correct response (counting all errors) on a log scale. Figure 7 presents similar plots of the

data from the three Ss in Experiment III. To obtain greater stability in the estimates of mean error RTs, conditions with similar total error rates were pooled. For the easier of S.E.K.'s discrimination tasks, Fig. 6 shows that single-response errors were substantially faster than correct responses at all levels of accuracy. For E.K.'s difficult discrimination task, and for those of all three Ss shown in Fig. 7, the direction and magnitude of

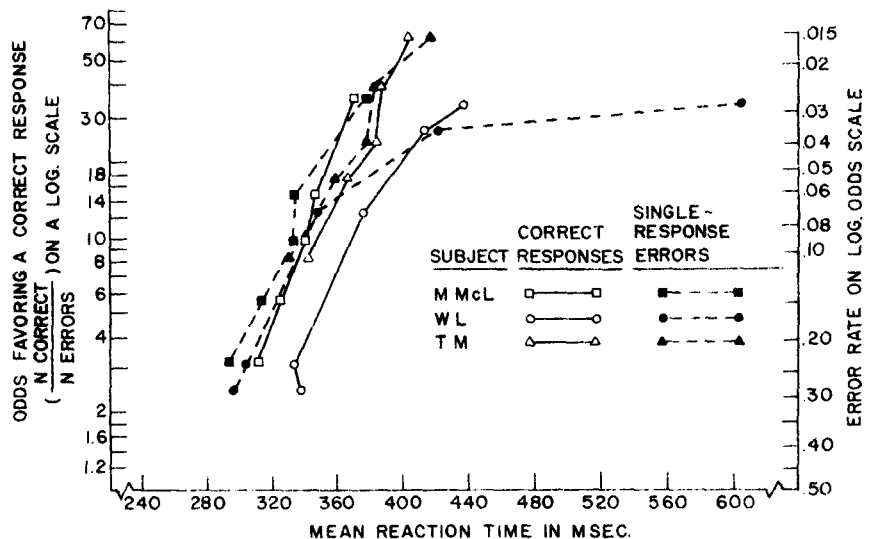


Fig. 7. Mean RTs for correct responses and for single-response errors in grouped sessions for each of the three Ss in Experiment III. The ordinate plots odds favoring a correct response (ratio of number of correct responses to number of all S's errors) in these pooled sessions.

the RT differences depended on how accurately S performed. Single-response errors were faster than correct responses *only* when both were fast and error rates were high. This difference decreased as the Ss responded more accurately, and actually reversed in their slowest and most accurate conditions (which were devoid of double-response errors). Similar, although much more variable, differences were observed within the individual conditions for each S.

In Experiment III, the largest difference in both directions occurred for S.W.L., who had the most difficult discrimination task (length:width ratio of 63:62) of the three Ss. His single-response errors were 41 msec faster than correct responses in his fastest conditions (with a combined error rate of 0.189), but his mean error RT was 167 msec slower than his mean RT for correct responses in conditions whose error rates varied about 0.029. For the most accurate of S.E.K.'s later conditions with the more difficult discrimination in Experiment II, his mean RT for errors became 125 msec slower than his mean RT for correct responses.

The considerably slower RTs for errors with S.E.K.'s more difficult (85:84) discrimination task appeared related to the apparent ceiling on his accuracy, noted for the data presented in Fig. 4. The data from E.K.'s earlier, very slow conditions (with error rates less than 0.05) were not included in Fig. 6, but as his overall mean RT continued to increase, so did the difference between mean RTs for errors and correct responses. In the

five slowest of these earlier conditions (not the earliest, but those exerting the greatest pressure for very accurate performance), E.K.'s error rate was 0.030 and his mean error RT was 775 msec slower than his mean time for correct responses.

GENERAL DISCUSSION

The Elusive Tradeoff

Experiments II and III clearly reject the fast guess model's assumption of all-or-none discrimination. Both demonstrated tradeoffs between discrimination speed and accuracy that could not be accounted for by a qualitative change in strategy (fast guesses). Experiment III also specified conditions in which Ss would exhibit such "true" tradeoffs. The Ss in Experiment I apparently used only two homogeneous response strategies, not because they *could not* trade accuracy for speed over a wide range without resort to fast guesses, but because the payoff structure *discouraged* such tradeoffs by making them suboptimal.

As the fast guess model assumed, a preprogrammed guessing strategy was an easily discovered option, and available to all Ss. It permitted a considerable increase in each S's speed, at the cost of chance-level error rates, by apparently allowing him to make a lower-level detection response to the simple presence of the stimulus. Since the time-cost function used in Experiments I and II was proportional to S's total RT, any reduction in RT increased his net payoff—if it did not also increase his error rate. Reasonable extrapolations of all five sets of tradeoff data in Experiments II and III (from which fast guesses had been removed) implied that S's accuracy would drop to chance level for performances considerably slower than he could obtain with his guessing strategy. This large fixed cost in time, for even the smallest improvement in accuracy, meant that the Ss in Experiments I and II earned a higher payoff by guessing until the loss in payoff for making an error (D) became fairly large.

The substantial dead time, apparently necessary before responses could be based on any evidence about the identity of the stimulus, can explain why Ss refused to perform with intermediate or high levels of errors in Experiment I (and perhaps also in Swensson and Edwards's experiments). Certainly, the free delay procedure of Experiment III, which reduced or eliminated the cost of this dead time, elicited considerably faster and less accurate performances than any of the three Ss had adopted in Experiment I. Their determination to avoid performing with high error rates,

together with values of D just sufficient to induce a switch from guessing to discriminating, probably accounts for the high degree of stability in Ss' mean RTs and error rates during the final conditions in Experiment I.

The explicit costs and payoffs did not completely control Ss' performance, however. In Experiment II, S.R.P. refused to slow his mean RT to attain the very low error rates he demonstrated during his initial training, even when the payoffs exerted considerable pressure for high accuracy. Also in Experiment II, S.E.K. produced fast but inaccurate nonguessing performances that were suboptimal—dominated in expected payoff by his guessing strategy. However, he did exhibit a considerable number of fast guesses during these sessions, as well.

Parameters of Speed-Accuracy Tradeoffs

Fitted tradeoff functions. To discuss tradeoff data—even on a strictly empirical level—the data points from individual performances must be summarized by a theoretical function that relates some measure of average RT to some measure of accuracy. Ideally, such a function would be implied by a particular model of S's real-time discrimination process. This section will discuss the tradeoff data from Experiments II and III in relation to four theoretical functions. All are monotonic tradeoff functions, in the sense that they associate an increase in S's expected accuracy with each increase in his mean RT. Any specific monotonic tradeoff function can be converted to a linear function of mean RT by defining "accuracy" as a particular monotonic transformation of error rate. Letting A represent the appropriate error-rate transformation, these functions have the form:

$$A = b(\text{RT} - C). \quad (3)$$

The constant C can be interpreted as the expected delay contributed by all processes that S must perform if his accuracy is to exceed chance level: stimulus input, detection, and motor times, plus any dead-time interval required for a choice response. The parameter b is the slope of S's tradeoff function; it reflects the rate of increase in A per unit increase in S's RT, once it exceeds C. For $\text{RT} < C$, A is assumed to be at chance level.

The five sets of tradeoff data obtained in Experiments II and III were fit by least-square approximations to Eq. 3 for A defined by the following four functions:

$$A = (p_c - p_e) \ln \left[\frac{p_c}{p_e} \right], \quad (4)$$

$$A = \ln \left[\frac{p_c}{p_e} \right], \quad (5)$$

$$A = (d')^2 = (2Z)^2, \quad (6)$$

$$A = H_t \\ = 1 + [p_c \log_2(p_c) + p_e \log_2(p_e)], \quad (7)$$

where p_c and p_e are the probabilities of a correct response and an error ($p_c + p_e = 1.0$), and d' in Eq. 6 is the familiar statistic of signal detection theory (Z is the distance value that separates the unit normal distribution into areas of p_c and p_e).

Only two of these accuracy measures derive from present tradeoff models. Equation 4 is the definition of expected accuracy that satisfies Eq. 3 for the optional stopping version of the random walk model (Edwards, 1965; Eq. 19). This version assumes that S follows his most efficient strategy in each condition, by setting a cutoff on his expected accuracy (A) and responding as soon as the evidence from his sampled observations satisfies this criterion. Equation 4 can also be obtained from other formulations of the optional stopping version (Stone, 1960; Laming, 1968), and holds only for choices between two equiprobable stimuli when error costs are equal for both. The function becomes much more complicated in less symmetric cases (Edwards, 1965).

A fixed stopping version of the random walk model assumes that S always preselects the number of observations he samples prior to each trial, and then responds with whatever accuracy he can achieve. Equation 6 provides an exact definition of A for this version when the diagnostic impact of a single observation, conditional on each stimulus, can be represented by two normal densities with equal variance (Green & Swets, 1966; Eq. 9.1). But in many other cases, Eq. 6 will provide a good approximation to A, whose adequacy increases with the size of S's fixed sample (increasing RT). Using this measure of A, Taylor, Lindsay, and Forbes (1967) produced very close fits to tradeoff data obtained by Schouten and Bekker (1967).

Unfortunately, other proposed incremental models do not present their predicted tradeoff functions in sufficiently explicit forms. The remaining two measures, Eqs. 5 and 7, were selected because they have been used successfully to fit empirical tradeoff data in other experiments.

Table 5

Slopes* of the Four Tradeoff Formulations Fitted to Data from Experiments II and III

Experiment	S	Random Walk Models		Empirical Functions	
		Optional Stop	Fixed Stop	H_t	$\ln\left(\frac{p_c}{p_e}\right)$
II	EK (35:34)	.541	2.854	.116	.517
	EK (85:84)	.213	.908	.053	.203
III	TM (55:53)	.316	1.502	.047	.283
	MMcL (50:49)	.375	1.912	.090	.366
	WL (63:62)	.226	1.026	.051	.198

*Estimated rate of increase in A (as defined by Equations 4-7) per 10-msec increase in S's mean RT.

Pew (1969) demonstrated good fits to both individual and group data using A defined by Eq. 5. This definition of A is proportional to the accuracy metric used for plotting Figs. 4-7; it approaches Eq. 4 asymptotically as p_c approaches 1.0. Equation 7 defines H_t , the number of bits of information about the identity of two equiprobable stimuli that are transmitted by S's responses (assuming equal error rates for both responses). As an empirical relation,¹ H_t has produced good linear fits to mean RT for tradeoff data in a number of experiments (Hick, 1952; Hale, 1969; Pachella & Fisher, 1969; Swanson & Briggs, 1969).

All four functions produced good fits to the five sets of tradeoff data shown in Figs. 4 and 5.² Even the poorest fit accounted for more than 72% of the total variance between speed and accuracy over conditions; the best fits accounted for more than 90%. No function was either decisively superior or obviously inferior for any set of data; the difference in percentage of variance accounted for by the best and worst fits ranged from 1.8% to 9.2%. Two of these four functions, those from Eqs. 4 and 6, respectively, are shown plotted as dotted and solid lines in Figs. 4 and 5. The four fitted functions differed substantially in shape only for performances with extremely high or extremely low error rates. Over the range of accuracy scores in these data (and those obtained by most other experiments), each of the four functions would provide an acceptable two-parameter fit to data generated by any other.

Tradeoff function slopes. Table 5 presents each function's estimated slope for the five sets of tradeoff data from Experiments II and III. The numbers represent the estimated rate of increase in the appropriate measure of accuracy for each 10-msec increase in S's mean RT. Increased stimulus difficulty (increasing similarity of the stimulus rectangles' length and width) clearly seemed to make Ss' tradeoff functions less steep. All tradeoff functions estimated their lowest rates

of increase in accuracy for Ss E.K. and W.L., who had the two most similar pairs of stimuli (length:width ratios of 85:84 and 63:62). Estimated slopes were also much lower for the tradeoff data from S.E.K.'s more difficult (85:84) pair of stimuli than for the data from his easier (35:34) pair.

Tradeoff function intercepts. Table 6 presents each function's estimate of C, S's RT intercept at chance-level accuracy, and also his estimated dead-time delay. These dead times are defined as the difference between each estimate of C and S's mean RT for preprogrammed guesses (from Experiments I or II); they represent each function's estimate of the delay (in addition to stimulus detection and response execution times) before S could respond with better-than-chance accuracy. Both versions of the random walk model estimated dead times close to 100 msec—about 20-30 msec longer than those from the two empirical tradeoff functions—when averaged over the five sets of tradeoff data. Unlike estimates of the tradeoff slopes, the intercept estimates showed little systematic change with stimulus similarity among these generally difficult pairs of stimuli. Apparently, the processes that determine the tradeoff function's slope seem to be sensitive to relatively small changes in discriminability, while those that contribute to the dead-time delay are

not. Table 6 shows a slight tendency toward slower intercepts for the more difficult of S.E.K.'s stimulus discriminations, but this did not occur with all of the four tradeoff functions.

Present incremental models that predict speed-accuracy tradeoffs devote scant attention to the function's RT intercept at chance accuracy. Without data to argue otherwise, these models have assumed that this intercept was at S's "simple" RT (Stone, 1960; Edwards, 1965; Vickers, 1970). The present data indicate that this assumption is certainly not true, at least for reasonably difficult discriminations. Dead-time estimates for these data were nearly as long as the range in RTs over Ss' tradeoff functions, from intercept estimates to the highest accuracy scores obtained.

The presence of these dead times could be interpreted in several ways. Two obvious possibilities arise from the distinction frequently made between two hypothetical serial stages of processing: stimulus identification and response selection (Smith, 1968; Welford, 1968). Differences between tradeoff-function intercepts and mean RTs for preprogrammed guesses could reflect: (1) some additional time S needed to wait (after he already had adequate evidence that the stimulus had appeared) before his processing yielded any evidence about the stimulus identity, or (2) some fixed choice time (after processing the stimulus information) required to select and prepare the appropriate physical response. The latter, response selection, interpretation of these dead times seems unsatisfactory, because it implies that choice RTs would always have to be at least 80-100 msec slower than preprogrammed guesses. Using the same physical responses, but easily discriminable stimuli, Swenson and Edwards (1971) obtained mean RT differences much smaller than these (15-20 msec and 45-70 msec), even when Ss' accurate performance was far

Table 6
Estimates in Milliseconds of RT Intercepts at Chance-Level Accuracy (C) by the Four Tradeoff Formulations, and Estimated Dead Time (DT)* for Each S

Experiment	S	Random Walk Models				Empirical Functions			
		Optional Stop		Fixed Stop		H_t		$\ln\left(\frac{p_c}{p_e}\right)$	
		C	DT*	C	DT	C	DT	C	DT
II	EK (35:34)	302	105	307	110	290	93	298	91
	EK (85:84)	317	120	324	127	306	109	292	95
III	TM	286	95	295	104	220	29	267	76
	MMcL	288	92	297	101	279	83	274	78
	WL	285	97	294	96	270	72	254	66
	Mean S	296	102	303	108	273	77	277	81

*Difference between each estimate of C and S's mean RT for preprogrammed guesses from Experiment I (Experiment II for S.E.K.).

better than chance level (3%-8% errors). It seems unlikely that the difficulty of Ss' discrimination should affect the amount of time they need to select or prepare their physical response, once the stimulus has been categorized.

Other Aspects of Discrimination Performance

Aside from the overall tradeoff functions, some of the more detailed data from Experiments II and III seem to have significance for theoretical models of discrimination. This section considers these effects and examines their generality.

Correct-response vs error RTs. Single-response errors, trials on which only the incorrect response occurred, comprised the great majority of all errors when Ss did not guess (RTs > 250 msec); double-response errors never occurred when Ss' error rates were less than 0.10. The difference in mean RTs between correct responses and these single-response errors seemed to depend on two factors: the difficulty of S's stimulus discrimination and the position on his speed-accuracy tradeoff at which he chose to perform. In earlier data with easily discriminable stimuli of similar type (diagonal bars), single-response errors were consistently 16-24 msec faster than Ss' correct responses (Swenson & Edwards, 1971). With the difficult discrimination tasks and time pressure in the present experiments, errors became increasingly faster than correct responses as D decreased and Ss' overall performance became faster and less accurate. This difference reversed for Ss' slowest and most accurate conditions in Experiments II and III (with the largest values of D), in which errors were slower than correct responses (except for S E.K., in his easier discrimination task). For the two most difficult discrimination tasks, mean error RTs were 125 and 167 msec slower than mean correct RTs when these Ss responded most slowly and accurately.

These data seem to resolve the apparently conflicting results reported by previous experiments. Faster RTs for errors seem to come from choice RT experiments (with relatively easy stimulus discriminations), which typically include strong time pressure (Lemmon, 1927; Weaver, 1942; Rabbitt, 1966; Egeth & Smith, 1967; Schouten & Bekker, 1967; Laming, 1968; Hale, 1969). Experiments that used more difficult stimuli generally also emphasized accuracy much more than speed of discrimination. These experiments typically report slower RTs for errors than for correct responses (Kellogg, 1931; Hecker,

Stevens, & Williams, 1966; Pickett, 1967; Audley & Mercer, 1968; Pike, 1968; Vickers, 1970). One early study (Henmon, 1911) had 10 Ss discriminate the longer of two lines that were very similar in length (20.3 and 20.0 mm), for many sessions with an emphasis on their response speed. Within the RT distributions for each S, both the fastest and the slowest responses contained higher frequencies of errors than those with intermediate RTs. Had the nonguessing (tradeoff) performance of each S in Experiments II and III been pooled over conditions, the resulting RT distributions would have resembled Henmon's (1911) data.

By the use of large variations in the pressure for accurate relative to fast performance, Experiments II and III demonstrate a continuous transition between the conditions that differentiate the traditional discrimination and choice RT experiments. Data from experimental conditions that imposed very large losses in payoff for errors resembled those of traditional discrimination tasks, in which time pressure is weak compared to the demand for accurate performance. As typically found in such experiments, Ss' performances were very accurate and their errors were generally slower than correct responses. Experimental conditions that reduced the cost of making errors made the time pressure more significant, as it is in the usual choice RT task. Even in the absence of fast guesses, when these Ss performed their difficult discriminations fast enough to begin trading accuracy for speed, their data began to resemble those typically obtained in choice RT experiments: errors became faster than correct responses.

An accuracy "ceiling." The existence of an upper limit on discrimination accuracy is not a novel concept. An accuracy ceiling is assumed implicitly in any experimental procedure that uses error rate as its sole measure of discrimination performance, without attempting either to make S respond quickly or to limit the time for which the stimulus is available to him. In Experiment II, S E.K. appeared unable to reduce his error rate below 0.03 for his more difficult task, however slowly he responded. Under extreme pressure to achieve high accuracy, he slowed his mean RT by 400 msec—almost triple the time he needed to reduce his error rate from 0.30 to 0.03—with no further reduction in errors.

Apparently after some point, additional observation time did not contribute further to the accuracy of S E.K.'s more difficult discrimination, but it did increase his cost for response time. Because he could have achieved

the same accuracy with faster RTs, all of these costly but futile attempts to eliminate errors represented suboptimal performances. In his later conditions, S E.K. refused to exhibit these attempts. None of the three Ss in Experiment III ever exhibited this kind of direct evidence of an accuracy ceiling, perhaps because their error costs were never as large and their discrimination tasks were somewhat easier. However, the slow error RTs noted in the most accurate performance of these Ss might indirectly reflect an accuracy ceiling. For S E.K., there was an intimate association between performance at maximum accuracy and his mean RT difference between errors and correct responses. When he slowed his overall performance most drastically, this difference grew to 775 msec.

Models for Discrimination Under Time Pressure

Performance in a discrimination task clearly depends on both of two types of factors: (1) characteristics of the stimuli to be discriminated and (2) the interacting pressures for speed and accuracy that are either explicit in S's costs and payoffs or are implicit in the instructions and procedure. The present experiments demonstrate how these latter procedural factors can determine many of the overall and detailed characteristics of discrimination performance. A model for discrimination under time pressure that is general enough to account for these data could subsume both easy and difficult discrimination tasks. The traditional discrimination and choice RT experiments might be regarded as special cases: highly similar stimuli discriminated under very weak time pressure, and dissimilar stimuli discriminated under strong time pressure.

The present results suggest a number of constraints upon any general model for discrimination under time pressure. Stimulus characteristics must determine limits on S's capability to trade discrimination speed for accuracy. The results suggest that they determine: (1) a minimum value of RT (slower than detection times) before accuracy can exceed chance level, (2) a specified growth in accuracy (assumed to be monotonic) as mean RT increases over the tradeoff region, and (3) an accuracy ceiling (perhaps measurable in practice only for difficult discriminations, whose ceiling is low). A prospective model would also have to account for how S's detailed performance seems to change over his range of possible tradeoffs: (1) slower mean RTs for errors than for correct responses when S apparently performs at or near the

maximum accuracy he can achieve, and (2) faster mean RTs for errors when S's accuracy is substantially less than its maximum level.

The most straightforward accounting of these phenomena is in terms of the incremental models, which provide for speed-accuracy tradeoffs in the discrimination process itself. The second of the following three sections attempts to do this; it points out present inadequacies in these models and discusses how they might be modified, focusing on the random walk model. The final section describes an alternative type of model, one that postulates a race between two independent processes: a resurrected all-or-none discrimination and an internal response "deadline." First, however, it seems appropriate to reconsider the fast guess model.

A modified fast guess model. Too sharp a focus on the failure of the all-or-none assumption for discrimination loses sight of the fast guess model's success in coping with a rather dramatic empirical fact. When an S performs a discrimination task under pressure to respond quickly, he may not (and frequently does not) always attempt to discriminate the stimulus. Yellott (1971) argues that a more sophisticated interpretation of the fast guess ideas would view them as a model of how different mixtures of S's available response strategies affect his mean performance. Whether the model should assume that stimulus discrimination is a single-stage (all-or-none) or a multiple-stage (incremental) process, Yellott feels, is a matter for empirical data to determine. In addition to simply deciding whether to guess or to respond from a discrimination state, S might also be able to predetermine which of several incremental discrimination stages would provide the output for his response.

Yellott (1971) shows how the fast guess model's equations enable statistical estimation of mean latencies for both guesses and "stimulus controlled" responses. If the estimated discrimination latency remains invariant over changes in S's accuracy, there is evidence that he used only a single stage of discrimination; lack of invariance would indicate the use of multiple stages. Yellott points out that a single-stage (all-or-none) assumption appears to hold quite well for the available data from easier discrimination tasks, such as Swenson and Edwards (1971).

Yellott's (1971) reinterpretation of the fast guess model, without a commitment to any particular model for stimulus discrimination, can make it consistent with the present results. However, the model now seems to

require the estimation of six parameters, which can assume different values in each condition. Three of these parameters relate to S's discrimination responses: (1) ϵ , the error probability, (2) RT_d for correct discrimination responses, and (3) RT_d for errors in discrimination. Both Parameters 2 and 3 seem necessary because the present results indicate a difference in RT between correct and erroneous discriminations, which apparently changes substantially as S changes his speed-accuracy trade (i.e., the number of stages in his discrimination process) without producing fast guesses. The three remaining parameters for each condition relate to characteristics of S's guesses: (4) q , the guessing probability, (5) the guess-response bias (between S's two responses, given that he guesses), and (6) RT_g , the mean RT for guesses in that condition.

With only two stimuli and responses, the data from each condition can provide only six independent quantities that can be used in estimating the six parameters of the expanded fast guess model: mean RTs for each of the four stimulus-response combinations and two error probabilities. Without some constraints on these parameters (e.g., some assumed invariance over conditions), the fast guess model becomes untestable, guaranteed to fit any set of data. Yellott (1971) seems to feel that the model is most useful as an estimation tool: e.g., to obtain statistical estimates of discrimination latencies that are uncontaminated by S's guessing times. But, although use of the model's equations avoids the difficult problem of identifying which of S's responses were guesses and which were discriminations, such a procedure assumes the model's validity. Without a means to check the fast guess model's assumptions, expanding the number of its parameters saves the model from rejection at the risk of making estimates of these parameters meaningless.

Incremental discrimination models. Present incremental models all assume that S has attempted to discriminate the stimuli. They differ primarily in their assumptions about how S accumulates information over observations, and the mechanisms by which he triggers a response. The essential predictions of most of these models are not affected by their assumptions about the kind of evidence these observations provide—i.e., either qualitative (binary) or quantitative (relative likelihood) evidence about which stimulus is present. *Random walk models*, in both the optional and fixed

stopping versions, assume that the cumulative impact of all S's stimulus observations is stored in a single "register." Each observation changes the current value in this register, adding or subtracting an amount assumed to depend on which stimulus was the more likely to have produced it (and usually also on the observation's likelihood ratio). The optional stopping version assumes that S responds as soon as the random walk of the register's value reaches either an upper or a lower cutoff, whose settings determine S's expected accuracy for each response (and indirectly, his expected RTs). The fixed stopping version assumes some timing mechanism or counter that terminates S's observing process, together with a single criterion that determines which response he will make, based on the sample of evidence in his register.

Versions of the accumulator model (LaBerge, 1962; Audley & Pike, 1965; Pike, 1968; Vickers, 1970) assume two registers, each accumulating evidence favoring one of S's two responses. Because these models assume that each observation changes only one accumulator without affecting the other, the process resembles a race; S's response depends upon which accumulator reaches its cutoff first. The *runs model* (Audley & Pike, 1965; Pike, 1968) is similar to the accumulator model for binary valued observations, except that an increase in one accumulator is assumed to "zero out" the other. Thus, the two registers accumulate only runs of consecutive observations that favor a particular response.

In their simplest forms, none of these models can explain how errors can be both faster and slower than correct responses. The optional stopping random walk model with fixed response cutoffs predicts identical mean RTs for errors and correct responses (Stone, 1960; Laming, 1968; Audley & Pike, 1968). The fixed stopping version makes the same prediction if the size of S's fixed sample does not change between trials; all responses based on the same number of observations will have the same mean RT (i.e., RT variability can result only from random variation in input and motor components of RT). Both the accumulator and the runs models predict that errors will be slower than correct responses (Audley & Pike, 1965; Pike, 1968), although some accumulator models may predict almost identical means for correct and error RTs when S's responses are fast and inaccurate (Vickers, 1970).

Several suggested modifications would allow an incremental model to predict faster mean RTs for errors than for correct responses. Each such

modification postulates some inherent variability from trial to trial, either: (1) in S's stopping criteria (Fitts, 1966; Pike, 1968), or (2) in the initial values assumed by S's internal register or accumulators when the stimulus appears (Laming, 1968). In any incremental model, S's stopping criteria simultaneously control both his expected accuracy and his mean RT by controlling the expected number of observations (expected sampling time) required to satisfy these criteria. Any variation in the stopping criteria between trials would associate S's less accurate responses with his shorter RTs (thus making the mean RT faster for errors). If, as suggested by Laming (1968), S began his sampling prematurely, then the initial values in his register or accumulators at the time the stimulus appeared might be different on each trial. By introducing variability in the expected number of observations S required to reach his (constant) stopping criteria from trial to trial, such premature sampling would produce effects similar to those of fluctuating criteria.

However, trial-to-trial variation of either type does not help the random walk models to account for *slower* mean error RTs in some conditions, nor does it help any of the incremental models to account for an upper limit on the accuracy S can attain with very difficult discriminations. None of these models presently provide for any explicit limit on the number of incremental discrimination stages. In particular, the random walk models directly imply that S could reduce his error rate to any arbitrarily low level by continued sampling—i.e., for sufficiently slow RTs.

One possible approach to modifying these models would follow the lead of signal detection theory (Green & Swets, 1966) and postulate that S uses an optimal policy, given certain assumed limitations on his information processing capabilities. The optional stopping random walk model, with response cutoffs that remain fixed at some particular expected accuracy, embodies the most efficient sampling procedure for basing a discrimination upon an unlimited number of independent and probabilistic observations (Stone, 1960; Edwards, 1965). The presence of an accuracy ceiling could reflect a limit on the number of useful observations S can sample. If S's observations were truncated, there would be a limit on the expected amount of evidence he could obtain before some "critical duration" elapsed.³ This would enforce a ceiling on his discrimination accuracy and change the sampling strategy that would be optimal for him to use.

With only a limited number of observations available, costs for errors, and a fixed cost per observation (proportional to time spent observing), an optimal sampling procedure would relax its decision criteria as the truncation point approached (Swets & Birdsall, 1967; Birdsall & Roberts, 1968; Rapoport & Burkheimer, 1971). By following this procedure, S would maximize his payoffs (minimize cost), because his sampling would terminate when additional observations could not provide evidence sufficient to change his decision. Such additional observations would be wasted; they would increase S's RT, but could not improve his accuracy. Thus, a modified optional stopping random walk model, with progressively relaxing cutoffs, would embody an optimal adjustment to an accuracy ceiling that resulted from a limit on S's effective observation interval. Pike (1968) showed that accuracy cutoffs that relax with the number of observations (observing time) cause the random walk model to predict slower mean RTs for errors than for correct responses. Cutoffs that become more lax over time will associate S's less accurate responses (more frequent errors) with his longer observation times (slower RTs). The greater frequency of errors among S's slowest RTs would produce a slower mean RT for errors—even though errors might represent only a small proportion of his responses.

The results of a large number of studies argue for a progressively relaxed cutoff in any discrimination model that postulates a stopping criterion based on S's expected accuracy. These experiments measured time and accuracy when the discrimination difficulty varied randomly from trial to trial (Henmon, 1906; Lemmon, 1927; Kellogg, 1931; Johnson, 1939; Festinger, 1943a, b; Birren & Botwinick, 1955; Botwinick, Brinley, & Robbin, 1948; Thurmond & Alluisi, 1963; Morgan & Alluisi, 1967; Pickett, 1964, 1967). For trials on which the more difficult stimuli occurred: (1) mean RTs were slower and (2) error rates were higher. The first of these results argues that Ss did not simply predetermine some observation interval and then respond with whatever accuracy they could achieve. The second result indicates that if these Ss established some kind of criterion on their expected accuracy for responding, they allowed this criterion to relax with increased observing time.

With the kinds of modifications considered here, either the optional stopping random walk model or the accumulator models might account for the present results. Both seem to require either: (1) trial-to-trial

variability in S's stopping criteria and some critical observation duration (to explain a ceiling on accuracy), or (2) two equivalent modifications that will produce the same effects. Although the modified random walk model seems more cumbersome than the accumulator models, this is primarily because of its greater original simplicity (assuming only one internal "register" rather than two "accumulators"). The two modified models may in fact be virtually indistinguishable by empirical data. Pike (1968) showed that for binary-valued observations, a random walk model whose optional stopping cutoffs relax by a fixed amount after each observation is formally identical to an accumulator model. Both types of models seem sufficiently unwieldy to motivate seeking an alternative type of model.

A deadline model. There is an alternative to the incremental models for the task of discrimination under time pressure. The *deadline model* begins with the original fast guess model, but adds another process that races with S's all-or-none discrimination to determine his response. It assumes that whenever S decides to attempt a discrimination (*not* to make a fast guess), his detection of the stimulus starts an internal timer. This timer triggers a response "deadline" after some variable interval, whose mean duration S can adjust. The S responds either when he reaches a discrimination state or when his deadline elapses, whichever comes first.

A change in his mean deadline time changes S's trade between speed and accuracy. Shorter deadlines are more likely to elapse before S reaches a discrimination state, thus they replace more of his slowest discrimination responses with chance-accuracy guesses triggered by the deadline. This deadline model resembles a model developed by Nickerson (1969), to account for RTs in "same"-"different" judgments. It has also been mentioned as a possible tradeoff mechanism by Yellott (1971) and R. T. Ollman (personal communication).

With no restrictions on the probability distributions assumed for discrimination or deadline latencies, the model is too flexible to be very useful. However, to account for the present results, the latency distribution for these difficult discriminations must be bounded from below and must assign substantial area to very long (perhaps infinite) latencies. The lower bound is required to account for the slow chance-accuracy intercepts of S's tradeoff functions; there must be some dead-time interval (e.g., 80-120 msec) after S detects the stimulus and before

he has any probability of entering a discrimination state. This lower bound can also explain faster mean RTs for errors than for correct responses when time pressure forces S to set very short deadlines. If S shifted his deadline distribution to extend partly below the minimum discrimination latency, then correct responses would outnumber errors only among S's slower RTs, which could include some discrimination responses.

The second constraint on S's distribution of discrimination latencies simply implies that sometimes S either would never reach a discrimination state, or would reach it only after some long delay. On these trials, only S's deadline insures that he will in fact respond within some reasonable amount of time. If the discrimination distribution were concentrated in some finite area, then S would be guaranteed to reach a discrimination state by setting his deadline sufficiently long; his mean RT and accuracy would increase monotonically, but only until all responses were based on discriminations. Thereafter, S's performance would not change, however long he set his deadline.

The presence of an accuracy ceiling suggests that the discrimination process is "blocked" and produces no output on some trials. After some point, longer deadlines continue to slow S's RTs, but do not further increase his proportion of discrimination responses. Mean RTs would increase much more rapidly for errors than for correct responses as S continued to extend his deadline, since these very slow deadline-produced responses would constitute a much higher proportion of his errors than of his correct responses.

Any serious test of the deadline model requires more complete specification (or independent estimation) of the probability distributions assumed for the deadline and discrimination delays. The model's credibility also requires a demonstration that Ss can actually carry out these two processes independently at the same time. In a simple RT task, Ollman and Billington (1972) have shown that Ss do have the ability to release a predetermined response, either when an internal deadline elapses (an anticipation response) or when an auditory signal occurs (a detection response), whichever comes first. In a later simple RT experiment (personal communication), Ollman was able to infer deadline time distributions that could be well approximated by standard distributions; some were best fit by normal distributions, others by the displaced exponential.

REFERENCES

- AUDLEY, R. J., & MERCER, A. The relation between decision time and relative response frequency in a blue-green discrimination. *British Journal of Mathematical & Statistical Psychology*, 1968, 21, 183-192.
- AUDLEY, R. J., & PIKE, A. R. Some alternative stochastic models of choice. *British Journal of Mathematical & Statistical Psychology*, 1965, 18, 207-225.
- BIRDSALL, T. G., & ROBERTS, R. A. Theory of signal detectability: Deferred decision theory. *Journal of the Acoustical Society of America*, 1965, 37, 1064-1074.
- BIRREN, J. E., & 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.
- EDWARDS, W. Costs and payoffs are instructions. *Psychological Review*, 1961, 68, 275-284.
- EDWARDS, W. Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology*, 1965, 2, 312-329.
- EGETH, H., & SMITH, E. E. On the nature of errors in a choice reaction task. *Psychonomic Science*, 1967, 8, 345-346.
- FESTINGER, L. Studies in decision: I. Decision time, relative frequency of judgment and subjective confidence as related to physical stimulus difference. *Journal of Experimental Psychology*, 1943a, 32, 291-306.
- FESTINGER, L. Studies in decision: II. An empirical test of a quantitative theory of decision. *Journal of Experimental Psychology*, 1943b, 32, 411-423.
- FITTS, P. M. Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, 1966, 77, 849-857.
- GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.
- HALE, D. J. Speed-error tradeoff in a three-choice serial reaction task. *Journal of Experimental Psychology*, 1969, 81, 428-435.
- HECKER, M. H., STEVENS, K. N., & WILLIAMS, C. E. Measurements of reaction time in intelligibility tests. *Journal of the Acoustical Society of America*, 1966, 39, 1188-1189.
- HENMON, V. A. C. The time of perception as a measure of differences in sensation. *Archives of Philosophical Psychology & Scientific Method*, 1906, No. 8.
- HENMON, V. A. C. The relation of time of a judgement to its accuracy. *Psychological Review*, 1911, 18, 186-201.
- HICK, W. E. On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 1952, 4, 11-26.
- JOHNSON, D. M. Confidence and speed in the two category judgment. *Archives of Psychology*, 1939, No. 241.
- KAHNEMAN, D., & NORMAN, J. The time-intensity relation in visual perception as a function of the observer's task. *Journal of Experimental Psychology*, 1964, 68, 215-220.
- KAHNEMAN, D., NORMAN, J., & KUBOVY, M. Critical duration for the resolution of form: Centrally or peripherally determined? *Journal of Experimental Psychology*, 1967, 73, 323-327.
- KELLOGG, W. N. The time of judgement in psychometric measures. *American Journal of Psychology*, 1931, 43, 65-86.
- LaBERGE, D. A recruitment theory of simple behavior. *Psychometrika*, 1962, 27, 375-396.
- LAMING, D. R. *J. Information theory of choice-reaction times*. New York: Academic Press, 1968.
- LEMMON, V. W. The relation of reaction time to measures of intelligence, memory, and learning. *Archives of Psychology*, 1927, No. 94.
- MORGAN, B. B., JR., & ALLUISI, E. A. Effects of discriminability and irrelevant information on absolute judgments. *Perception & Psychophysics*, 1967, 2, 54-58.
- NICKERSON, R. S. 'Same'-different' response times: A model and a preliminary test. In W. G. Koster (Ed.), *Attention and performance II: Acta Psychologica*, 1969, 30, 257-275.
- OLLMAN, R. T. Fast guesses in choice reaction time. *Psychonomic Science*, 1966, 6, 155-156.
- OLLMAN, R. T., & BILLINGTON, M. J. The deadline model for simple reaction times. *Journal of Cognitive Psychology*, 1972, in press.
- PACHELLA, R. C., & FISHER, D. F. Effect of stimulus degradation and similarity on the trade-off between speed and accuracy in absolute judgements. *Journal of Experimental Psychology*, 1969, 81, 7-9.
- PEW, R. W. The speed-accuracy operating characteristic. In W. G. Koster (Ed.), *Attention and performance II. Acta Psychologica*, 1969, 30, 16-26.
- PICKETT, R. M. The perception of a visual texture. *Journal of Experimental Psychology*, 1964, 68, 13-20.
- PICKETT, R. M. Response latency in a pattern perception situation. In A. F. Sanders (Ed.), *Attention and performance. Acta Psychologica*, 1967, 27, 160-169.
- PIKE, A. R. Latency and relative frequency of response in psychophysical discriminations. *British Journal of Mathematical & Statistical Psychology*, 1968, 21, 161-182.
- RABBITT, P. M. A. Errors and error correction in choice-response tasks. *Journal of Experimental Psychology*, 1966, 71, 264-272.
- RAPOPORT, A., & BURKHEIMER, G. J. Models for deferred decision making. *Journal of Mathematical Psychology*, 1971, 8, 508-538.
- SCHOUTEN, J. F., & BEKKER, J. A. M. Reaction time and accuracy. In A. F. Sanders (Ed.), *Attention and performance. Acta Psychologica*, 1967, 27, 143-153.
- SMITH, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, 69, 77-110.
- STONE, M. Models for choice-reaction times. *Psychometrika*, 1960, 25, 251-260.
- SWANSON, J. M., & BRIGGS, G. E. Information processing as a function of speed versus accuracy. *Journal of Experimental Psychology*, 1969, 81, 223-229.
- SWENSSON, R. G., & EDWARDS, W. Response strategies in a two-choice reaction task with a continuous cost for time. *Journal of Experimental Psychology*, 1971, 88, 67-81.
- SWETS, J. A., & BIRDSALL, T. G. Deferred decision in human signal detection: A preliminary experiment. *Perception & Psychophysics*, 1967, 2, 15-28.
- TAYLOR, M. M., LINDSAY, P. H., & FORBES, S. M. Quantification of shared capacity processing in auditory and visual discrimination. In A. F. Sanders (Ed.), *Attention and performance. Acta Psychologica*, 1967, 27, 223-229.
- THURMOND, J. B., & ALLUISI, E. A. Choice time as a function of stimulus

- dissimilarity and discriminability. *Canadian Journal of Psychology*, 1963, 17, 326-337.
- VICKERS, D. Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, 1970, 13, 37-58.
- WEAVER, H. R. A study of discriminative serial action: Manual response to color. *Journal of Experimental Psychology*, 1942, 31, 177-201.
- WELFORD, A. T. *Fundamentals of skill*. London: Methuen, 1968.
- YELLOTT, J. I., JR. Correction for guessing in choice reaction time. *Psychonomic Science*, 1967, 8, 321-322.
- YELLOTT, J. I., JR. Correction for guessing and the speed-accuracy tradeoff in choice reaction time. *Journal of Mathematical Psychology*, 1971, 8, 159-199.

NOTES

1. Hick (1952) suggested several conceptual models that would predict linear relationships between RT and H_t for errorless performance, when H_t reflected variations in the number of stimuli and responses and the uncertainty of the stimulus sequence. However, Ollman (1966) has pointed out that Hick's models cannot predict linearity between RT and H_t when these measures reflect only variations in S's speed-accuracy tradeoff. For the two-choice case, a corrected version of Hick's theories leads to the fast guess model (Ollman, 1966).
2. Data prior to Session 27, for S E.K.'s more difficult discrimination task, were not used in fitting these functions. As discussed later, these earlier accurate performances

seem to reflect a discontinuity in S's tradeoff function. The five unusually poor performances from his later conditions in the task (circled points in Fig. 4) were also arbitrarily omitted; correlations for all four functions were about equally degraded when these data points were included.

3. Critical durations, similar in magnitude to those considered here (i.e., 150-200 msec), have been found in other tasks that required identification of visual stimuli (Kahneman & Norman, 1964; Kahneman, Norman, & Kubovy, 1967). These studies measured the duration within which time-intensity reciprocity held for certain form discriminations.

(Accepted for publication February 25, 1972).