

A test of the deadline model for speed-accuracy tradeoffs

ERIC RUTHRUFF

NASA-Ames Research Center, Moffett Field, California

Two experiments were conducted to evaluate the deadline model for speed-accuracy tradeoffs. According to the deadline model, participants in speeded-response tasks terminate stimulus discrimination as soon as it has run to completion or as soon as a predetermined time deadline has arrived, whichever comes first. Speed is traded for accuracy by varying the time deadlines; short deadlines yield fast but sometimes inaccurate responses, whereas long deadlines allow for slow, accurate responses. A new prediction of this model, based on a comparison of reaction time distributions, was derived and tested in experiments involving the joint manipulation of speed stress and stimulus discriminability. Clear violations of this prediction were observed when participants made relative brightness judgments (Experiment 1) and when they made lexical decisions (Experiment 2), rejecting both the deadline model and the fast-guess model. Several alternative models for speed-accuracy tradeoffs, including random-walk and accumulator models, are compatible with the results.

The time required to perform simple cognitive tasks depends on many factors, not the least of which is the participant's set for speed versus accuracy. In fact, responses made under speed stress can be several hundred milliseconds faster than those observed under accuracy stress. This dramatic reduction in reaction time (RT) is accompanied by an equally dramatic increase in error rates, of course; speed is "traded" for accuracy. How is this speed-accuracy tradeoff achieved? More specifically, how is the human information processing system fine-tuned in order to produce fast/inaccurate responses rather than slow/accurate ones?

One view is that stimulus discrimination involves a gradual accumulation of information over time—continuously or in many discrete steps—and is set to terminate once a sufficient amount of progress has been made. The amount of progress deemed to be sufficient (known as the *criterion*) can be raised to yield more accurate responses or lowered to yield faster responses. Many different conceptualizations of the accumulation process have been offered, but two of the more popular versions are the *accumulator* model (LaBerge, 1962; Vickers, 1970) and the *random-walk* model (Edwards, 1965; Stone, 1960). According to accumulator models, separate "counters" are used to keep track of the amount of evidence favoring each possible response. It is typically assumed that once any of these counters has reached its criterion, the corresponding response

is executed. According to random-walk models, on the other hand, a single counter keeps track of the relative amount of evidence in favor of one response over another (e.g., with positive values indicating that response "A" is likely to be correct and negative values indicating that response "B" is likely to be correct). A response is executed once this counter has reached its positive or negative criterion. (See Heath, 1984, and Vickers & Smith, 1985, for critical evaluations of accumulator and random-walk models.)

The notion of an information accumulation process that runs until it has achieved the desired amount of progress has some intuitive appeal. On the other hand, it is certainly not the only possible mechanism for trading speed and accuracy. There are two very simple alternatives, the *fast-guess* model and the *deadline* model, in which participants are assumed to have no direct control over the time required to complete stimulus discrimination. As will be discussed next, the fast-guess model has been rejected, but the deadline model remains viable.

THE FAST-GUESS MODEL

According to the fast-guess model (Ollman, 1966; Yellott, 1971), participants decide before each trial to either complete the required discrimination or make a fast-guess response. Fast-guess responses are random guesses that are initiated as soon as the stimulus has been detected and therefore are very fast but inaccurate. So, although participants cannot control the duration of the discrimination process, they can trade speed for accuracy by varying the proportion of fast guesses.

Although there is some evidence that participants make fast guesses when placed under certain types of extreme speed stress (e.g., when the payoff for fast responses far outweighs the penalty for making errors; see Swensson, 1972), it is now clear that the fast-guess model is not gen-

During the course of this research, the author received support from Public Health Service Training Grant 2T32-MH14268 and National Research Service Award 1F31-MH10922. The research was also supported in part by National Institute of Mental Health Grant PHS-40733 to Jeff Miller. I would like to thank Ehtibar Dzhafarov, Allen Osman, Hal Pashler, Doug Rohrer, and especially Jeff Miller for many useful comments on earlier versions of the manuscript. Correspondence concerning this article should be addressed to Eric Ruthruff, NASA-Ames Research Center, MS 262-2, Moffett Field, CA 94035.

erally valid for choice RT tasks (Pachella, 1972; Reed, 1973; Swenson, 1972). For example, the model predicts that participants will make more fast guesses as speed stress is increased, resulting in roughly the same decrease in accuracy for all conditions within a block of trials. This is because the decrease in accuracy is determined solely by the number of fast guesses, and approximately equal numbers of fast guesses should be made in all conditions. However, it is rarely the case that the accuracy of all conditions suffer equally; participants under speed stress tend to make far more errors in the difficult conditions than in the easy conditions of an experiment (e.g., Pachella, 1972).

THE DEADLINE MODEL

The deadline model is similar to the fast-guess model in that participants are assumed to have no control over the time required for complete stimulus processing. Instead of making occasional random guesses, however, participants respond to speed stress by setting time deadlines; the greater the speed stress, the faster the time deadlines. Stimulus discrimination is terminated—and later (or “residual”) processes begin—once the discrimination has been completed or the deadline has been reached, whichever comes first. Thus, the RT on any given trial can be described as

$$RT = \min(\text{discrimination time, deadline time}) + \text{residual time.}$$

According to this model, the discrimination time and the deadline time depend on mutually exclusive sets of experimental factors. The discrimination time depends on the nature of the stimulus presented (e.g., its contrast), but it is not under the participant's control and therefore cannot be adjusted according to the relative demands on speed versus accuracy. Meanwhile, the deadline time will depend on the relative demands on speed versus accuracy; however, because it is set in advance of the trial, it will be independent of the stimulus presented.

How is a response chosen when the deadline arrives before the discrimination can be completed? When this model was first proposed (e.g., Ollman, 1977; Swenson, 1972; Yellott, 1971), it was assumed, for the sake of simplicity, that participants make a random guess in this situation. In the discussion that follows, however, I will allow for the possibility that participants use partial information accumulated prior to the arrival of the deadline to produce a “best guess,” which is then passed on to later processes (see Meyer, Irwin, Osman, & Kounios, 1988). I will assume only that the accuracy of this guess does not affect the duration of these later processes (see Sternberg, 1969).

The general version of the deadline model does not make any assumptions about the exact nature of the discrimination process—although presumably the discrimination involves some type of information accumulation. As noted above, however, the deadline model does assert that the time required to complete the discrimination process is not under the control of the participant—an assumption that seems most plausible in the case of relatively rapid mental operations that require little or no conscious interven-

tion. For such operations, it is conceivable that processing continues (unless interrupted by the arrival of a time deadline) until an asymptotic level of progress has been achieved and that participants have little or no access to the amount of progress at intermediate points in time. Or perhaps participants can, in principle, gain access to the amount of progress, but the costs of retrieving this information (in terms of time and effort) outweigh the potential gains.

The deadline model just described is attractive because it offers a very simple explanation of the speed-accuracy tradeoff, and it seems prudent to first consider simple models before moving on to more complex ones. Furthermore, participants appear to have little difficulty making responses before any given time deadline suggested by the experimenter (e.g., Pachella & Fisher, 1969, 1972), which is consistent with the notion that participants have access to a reasonably accurate internal timing mechanism. In addition, the deadline model can successfully explain many aspects of simple and choice RT data (e.g., Kornblum, 1973; Nickerson, 1969, 1971; Ollman & Billington, 1972; Swenson, 1972). In particular, it can explain the finding that participants under speed stress commit errors primarily in the difficult conditions of an experiment: difficult discriminations are the least likely to finish before the deadline arrives and so their accuracy will suffer the most from speed stress.

Wilding (1974), on the other hand, argued that the use of time deadlines should produce a pronounced peak in the right-hand tail of the RT distribution, contrary to his observations. In other words, there should have been a large number of slow responses with approximately the same RT, produced by trials on which a response was prompted by the arrival of a time deadline. The assumption that all responses initiated by a deadline will have roughly the same RT is unrealistic, however, given the likelihood of considerable trial to trial variability in the deadline times. This variability might be due to randomness inherent to the timing mechanism and/or to strategic changes in the deadline time across trials (e.g., participants might increase the deadline time after making an error). Furthermore, the durations of the processes that occur subsequent to the arrival of the deadline (e.g., response execution) may also contribute substantial variability to the RT. Therefore, deadline responses can produce highly variable RTs; in principle, these RTs could even be more variable than the RTs due to completion of the discrimination. The data reported by Wilding (1974), therefore, rule out only a narrow and implausible version of the deadline model.

In summary, the deadline model seems to be capable of explaining the existing data. Because the deadline model has not yet been ruled out and because it proposes a very simple mechanism for trading speed and accuracy, the present experiments were designed to provide a more decisive test of this model.

Analysis of RT Distributions

The deadline model is capable of predicting a very wide range of results, partly because the general version of the model does not specify the shapes of the distributions of

deadline times and discrimination times and because it does not specify how accurate participants should be as a function of the discrimination or deadline times. Nevertheless, by making a few seemingly reasonable assumptions, it is possible to derive a prediction, based on a comparison of RT distributions, that all versions of the deadline model outlined above should obey. Essentially, the prediction is that there will be a limit on the speed of responses under high-speed stress and high stimulus discriminability relative to other conditions of an experiment.

Consider an experiment with two levels of speed stress (speed stress and accuracy stress) factorially combined with two levels of stimulus discriminability (easy and hard). Assume that stimulus discriminability is manipulated within blocks, so that participants cannot anticipate discrimination difficulty and therefore would be constrained to set the same time deadlines in the easy and hard conditions. This experiment consists of four conditions: easy discrimination/speed stress (ES), easy discrimination/accuracy stress (EA), hard discrimination/speed stress (HS), and hard discrimination/accuracy stress (HA).

A prediction of the deadline model for such an experiment, which will henceforth be referred to as the *deadline model inequality*, is that

$$F_{ES}(t) \leq F_{EA}(t) + F_{HS}(t), \text{ for all } t \quad (1)$$

where F_{ES} , F_{EA} , and F_{HS} are the cumulative distribution functions (CDFs) of RTs in Conditions ES, EA, and HS, respectively. Essentially, this expression states that the proportion of responses faster than t msec in Condition ES should always be less than or equal to the sum of the proportions of responses faster than t msec in Conditions EA and HS. So, when considered alone, Conditions EA and HS will of course produce fewer fast responses than will Condition ES; however, when their responses are "pooled" together, they must yield at least as many fast responses as does Condition ES. As will be noted in the General Discussion section, the deadline model inequality is also implied by the fast-guess model but is not implied by random-walk or accumulator models.

The logic behind this prediction of the deadline model, which will be formalized in the next section, is as follows. RTs in Condition ES must come from one of two kinds of trials: (1) those on which the easy discrimination has been completed, and (2) those on which a deadline (set under speed stress) has arrived. Trials of type 1 above also occur in Condition EA. If anything, they will occur more frequently in Condition EA because the easy discrimination is more likely to beat a deadline set under accuracy stress than it is to beat a deadline set under speed stress. Furthermore, trials of type 2 above occur in Condition HS. If anything, they will occur more frequently in Condition HS because the deadline will be reached more often. Therefore, when the trials from Conditions HS and EA are pooled together, they must be at least as fast as the trials from Condition ES alone.

Proof of the Deadline Model Inequality

Before deriving the deadline model inequality, it is necessary to establish some notation. Let P_E and P_H be random

variables representing the time required to complete the discrimination process in the easy and hard discrimination conditions, respectively. As noted earlier, these random variables should be independent of speed stress, because, according to the deadline model, participants have no control over the time required to complete the discrimination.

Also let D_S and D_A be random variables representing the deadline times under speed and accuracy stress, respectively. The deadline times should be independent of the discriminability condition, because the deadlines are set in advance of each trial (before discriminability condition is known) and because the timing mechanism is assumed to operate independently of the discrimination process.

A random variable R is also needed to represent the time required for the residual processes (e.g., response execution) that take place subsequent to the "race" between the discrimination process and the time deadline. As an approximation, one could assume that R does not depend on the experimental condition. In the interest of generality, however, I will allow for the possibility that the duration of the residual processes depends on discriminability when the discrimination process wins the race and depends on the level of speed stress when the deadline wins the race. In other words, experimental manipulations that affect the loser of the race should have no influence on what takes place after the race (note that if this were not the case, then the deadline model inequality would not necessarily hold). Let R_S and R_A represent the duration of residual processes under speed and accuracy stress, respectively, when the deadline wins the race, and let R_E and R_H represent the duration of residual processes in the easy and hard discrimination conditions, respectively, when the discrimination process wins the race.

The eight random variables just described (P_E , P_H , D_S , D_A , R_E , R_H , R_S , and R_A) need not be independent, but their various joint distributions must be the same in each experimental condition in which they are defined.¹

Using this notation, the RT in Condition ES, denoted by the random variable RT_{ES} , will be equal to $P_E + R_E$ on any trial in which the discrimination is completed prior to the deadline (i.e., $P_E < D_S$) and will be equal to $D_S + R_S$ on any trial in which the discrimination is not completed prior to the deadline (i.e., $P_E \geq D_S$). Thus,

$$RT_{ES} = \begin{cases} P_E + R_E, & \text{if } P_E < D_S \\ D_S + R_S, & \text{if } P_E \geq D_S \end{cases}$$

These two events are mutually exclusive, so

$$\begin{aligned} F_{ES}(t) &= p(RT_{ES}(t) \leq t) \\ &= p[(P_E < D_S) \cap (P_E + R_E \leq t)] \\ &\quad + p[(P_E \geq D_S) \cap (D_S + R_S \leq t)]. \end{aligned} \quad (2)$$

This expression says that $F_{ES}(t)$ has two additive components: one due to trials on which the discrimination is completed prior to the deadline, and one due to trials on which it is not.

Next, it will be shown that, according to the deadline model, $F_{EA}(t)$ should be greater than or equal to this first

component, $p[(P_E < D_S) \cap (P_E + R_E \leq t)]$, at each time t . It will then be shown that $F_{HS}(t)$ should be greater than or equal to the second component, $p[(P_E \geq D_S) \cap (D_S + R_S \leq t)]$, at each time t . These two facts lead directly to the deadline model inequality, $F_{ES}(t) \leq F_{EA}(t) + F_{HS}(t)$.

The probability of observing an RT less than or equal to t msec in Condition EA can be expressed as a sum of two components:

$$F_{EA}(t) = p[(P_E < D_A) \cap (P_E + R_E \leq t)] \\ + p[(P_E \geq D_A) \cap (D_A + R_A \leq t)].$$

Because participants are expected to set much shorter deadlines under speed stress than under accuracy stress, it seems quite reasonable to assume that for every possible combination of values of P_E and R_E , D_S stochastically dominates D_A —that is, $p(D_S \leq t | P_E = p, R_E = r) \geq p(D_A \leq t | P_E = p, R_E = r)$ for all t, p , and r .² This simply means that the frequency of deadline times less than or equal to t msec will be greater under speed stress than under accuracy stress, regardless of the values of P_E and R_E . This assumption implies that

$$p[(P_E < D_A) \cap (P_E + R_E \leq t)] \\ \geq p[(P_E < D_S) \cap (P_E + R_E \leq t)]. \quad (3)$$

$F_{EA}(t)$ must be greater than or equal to the left side of Inequality 3; therefore, it must also be greater than or equal to the right side of Inequality 3. So,

$$F_{EA}(t) \geq p[(P_E < D_S) \cap (P_E + R_E \leq t)]. \quad (4)$$

The probability of observing an RT less than or equal to t msec in Condition HS is also composed of two components:

$$F_{HS}(t) = p[(P_H \geq D_S) \cap (D_S + R_S \leq t)] \\ + p[(P_H < D_S) \cap (P_H + R_H \leq t)].$$

Because the hard discrimination takes much longer to complete than does the easy discrimination, it seems reasonable to assume that for every possible combination of values of D_S and R_S , P_E stochastically dominates P_H . This implies that

$$p[(P_H \geq D_S) \cap (D_S + R_S \leq t)] \\ \geq p[(P_E \geq D_S) \cap (D_S + R_S \leq t)]. \quad (5)$$

$F_{HS}(t)$ must be greater than or equal to the left side of Inequality 5; therefore, it must also be greater than or equal to the right side of Inequality 5. So,

$$F_{HS}(t) \geq p[(P_E \geq D_S) \cap (D_S + R_S \leq t)]. \quad (6)$$

Substituting $F_{EA}(t)$ for $p[(P_E < D_S) \cap (P_E + R_E \leq t)]$ (see Inequality 4) and substituting $F_{HS}(t)$ for $p[(P_E \geq D_S) \cap (D_S + R_S \leq t)]$ (see Inequality 6) into Equation 2 yields the deadline model inequality, $F_{ES}(t) \leq F_{EA}(t) + F_{HS}(t)$. Note that this proof does not require any assumptions regarding the shape of the distributions of any of the random variables, nor does it require any assumptions regarding the accuracy of responses when the deadline arrives or regarding

the accuracy of responses when the discrimination is completed.

It should be emphasized that the proposed test states an *upper limit* on the speed of responses in Condition ES relative to Conditions EA and HS. This test is therefore conservative in the sense that small deviations from a deadline model would probably not produce violations of the inequality. On the other hand, if violations are observed, then one can be confident that the model is wrong.

Also note that the right side of the deadline model inequality approaches a value of 2 when t becomes large, whereas the left side approaches a value of 1. It follows that there will always be a value of t beyond which the deadline model inequality holds, whether the deadline model is true or not. So when the inequality does hold for large t , this should not be construed as partial support for the deadline model.

The following example may help to illustrate why data that do not obey the deadline model inequality are difficult to reconcile with the deadline model. Figure 1 shows F_{ES} , F_{EA} , F_{HS} , and the sum $F_{EA} + F_{HS}$ as a function of time for a set of hypothetical data that clearly violate the deadline model inequality. In this plot, an RT of less than 350 msec occurs about 60 times per 100 trials in Condition ES. Can many of these fast RTs be due to completion of the easy discrimination? This seems very unlikely, because, when participants are under accuracy stress, they almost never complete the easy discrimination (Condition EA) early enough to respond in less than 350 msec. Can the fast RTs in Condition ES be due to the arrival of early deadlines? This also seems very unlikely, because, under speed stress, these early deadlines should be triggering fast responses when paired with the hard discrimination condition (i.e., Condition HS) as well, yet there are

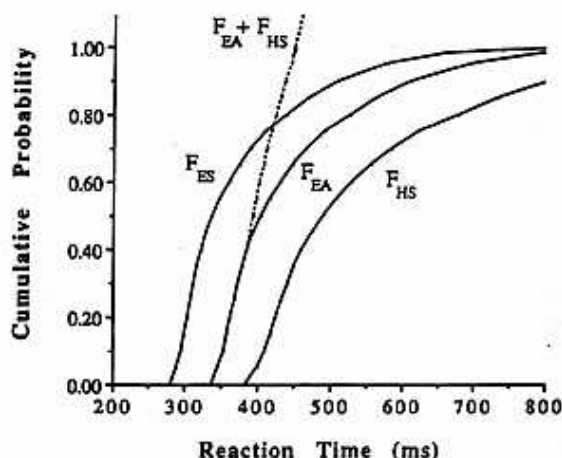


Figure 1. Plot of the cumulative distribution function of RTs in the easy discrimination/speed stress (ES), easy discrimination/accuracy stress (EA), and hard discrimination/speed stress (HS) conditions, along with the sum of the cumulative distribution function of RTs in Conditions EA and HS, for a hypothetical set of data that violate the deadline model inequality.

essentially no responses faster than 350 msec in Condition HS. Therefore, there is no way for the deadline model to explain where the large number of fast responses in Condition ES came from, because they could not have been due to completion of the easy discrimination and they could not have been due to the arrival of a deadline.

EXPERIMENT 1

To determine if the deadline model inequality holds, Experiment 1 was conducted using a task in which participants judged the relative brightness of two visual stimuli under accuracy stress and speed stress.

Method

Participants. Six male and 4 female undergraduates (between the ages of 18 and 24) at the University of California, San Diego, participated to fulfill a class requirement. Each had normal or corrected-to-normal visual acuity and participated in a single session lasting about 40 min.

Stimuli. Stimulus displays consisted of two white, filled rectangles of roughly uniform intensity, displayed on an NEC Multisynch video monitor. Each rectangle subtended approximately 0.6° horizontally and 1.0° vertically from a typical viewing distance of 60 cm. The centers of the rectangles were placed approximately 1.0° to the left and right of fixation.

Procedure. The participants were instructed to press the "z" key with the left index finger if they judged the left stimulus to be brighter or to press the "f" key with the right index finger if they judged right stimulus to be brighter. Within each block of the experiment, the difficulty of the brightness discrimination was manipulated by varying the disparity between the intensities of the two stimuli. There were four levels of discriminability, referred to as the *easy*, *medium*, *moderate*, and *hard* discrimination conditions. Note that the key analyses reported below utilized only the easy and hard conditions; the intermediate conditions were included primarily to discourage the participants from preparing for the easy or hard discrimination. In the easy condition, the luminance of the less intense stimulus was approximately 1.4 cd/m^2 , whereas the luminance of the more intense stimulus was approximately 3.2 cd/m^2 . The corresponding luminance values were roughly 1.7 and 2.4 cd/m^2 in the medium condition, 1.8 and 2.3 cd/m^2 in the moderate condition, and 1.9 and 2.2 cd/m^2 in the hard condition. These luminance values were chosen so that the average brightness of the two stimuli remained approximately constant across conditions. Discriminability condition was chosen randomly on each trial, as was the side (left vs. right) on which the more intense stimulus was displayed, with the restriction that each discriminability/side combination occur exactly 10 times in every block.³

The relative emphasis on speed versus accuracy was manipulated between blocks. In the speed stress blocks, the participants were instructed to respond very quickly, so that they were making approximately 10% errors. In the accuracy stress blocks, the participants were instructed to make less than 3% errors. Half of the participants were initially given speed stress instructions, and half were given accuracy stress instructions. The participants completed 100 practice trials under these speed/accuracy instructions, followed by three blocks of 80 trials. The participants were then given 20 practice trials under the opposite speed/accuracy instructions, followed by three more blocks of 80 trials. Each block was preceded by three warm-up trials. Feedback on mean RT and percent correct was provided at the end of each block.

The sequence of events within a trial was as follows. Each trial began with the presentation of a fixation cross for 800 msec, followed by a blank screen for 300 msec. The stimuli were then displayed until the participant made a response. If the response was in-

correct, an error message was displayed for 800 msec. The next trial began after an intertrial interval of 1 sec.

Results and Discussion

Mean RTs and error rates for Experiment 1 are shown in Table 1 as a function of speed versus accuracy stress and discrimination difficulty. An analysis of variance on mean RTs revealed reliable effects of discriminability [$F(3,27) = 27.66, p < .001$] and speed versus accuracy stress [$F(1,9) = 19.21, p < .001$], as well as a reliable discriminability \times speed versus accuracy stress interaction [$F(3,27) = 16.17, p < .01$]. An analysis of error rates revealed reliable main effects of discriminability [$F(3,27) = 17.2, p < .001$] and speed versus accuracy stress [$F(1,9) = 12.15, p < .01$].

The CDF of RTs in Conditions EA, ES, HA, and HS are shown in Figure 2. To obtain these CDFs, the 10 RTs in the easy and hard conditions for a given participant, block, and side of the more intense stimulus (left vs. right) were rank ordered. These 10 ordered numbers estimate the RTs at the 5th, 15th, 25th, ..., 95th percentiles of the true CDF in that condition for a given participant and block. The RTs at each percentile (5th, 15th, etc.) were then averaged across participants, blocks, and the side of the more intense stimulus to produce the composite 10-point CDFs shown in Figure 2. This procedure for obtaining an average RT distribution, which is essentially a form of "Vincenzizing" (Ratcliff, 1979; Vincent, 1912), serves to eliminate unwanted variance due to individual differences and practice effects.

Figure 3 shows the CDF from Condition ES and the sum of the CDFs from Conditions EA and HS;⁴ standard error bars are drawn around the mean RTs at each of the 10 percentile points.⁵ To be consistent with the deadline model inequality, the CDF for Condition ES should be below and to the right of the sum of the CDFs for Conditions EA and HS. Violations of this prediction were observed for RTs less than about 420 msec. To determine if these violations were consistent across participants, the mean RTs were compared using paired *t* tests at each of the 10 percentile points. This analysis revealed that the violations were reliable ($p < .05$) at the 5th, 15th, 25th, 35th, 45th, and 55th percentiles; therefore, the present data provide clear evidence against the deadline model.

The analyses described in the previous paragraph were based on a comparison of the two most extreme discriminability conditions (easy and hard), because this seemed likely to provide the most sensitive test for violations of the deadline model inequality. However, analogous tests can also be conducted using the other possible pairings of the four discriminability conditions (e.g., easy and medium,

Table 1
Mean Reaction Time (in Milliseconds) and Percent Error (PE)
in Experiment 1 for the Speed and Accuracy Stress Conditions
as a Function of Stimulus Discriminability

| Condition | Discriminability | | | | | | | |
|-----------------|------------------|-----|--------|-----|----------|-----|------|------|
| | Easy | | Medium | | Moderate | | Hard | |
| | RT | PE | RT | PE | RT | PE | RT | PE |
| Speed stress | 371 | 2.7 | 404 | 4.8 | 417 | 7.5 | 464 | 14.3 |
| Accuracy stress | 446 | 0.2 | 503 | 1.0 | 574 | 1.0 | 749 | 6.7 |

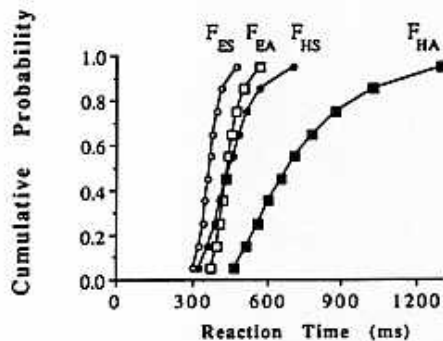


Figure 2. Plot of the cumulative distribution function of RTs in the easy discrimination/speed stress (ES), easy discrimination/accuracy stress (EA), hard discrimination/speed stress (HS), and hard discrimination/accuracy stress (HA) conditions in Experiment 1.

easy and moderate, etc.). All but one of these five alternative analyses revealed significant violations of the deadline model inequality. The analysis that failed to reveal significant violations was the one involving the medium and moderate conditions—presumably because of the relatively small discriminability effect between these two conditions. However, even this comparison revealed a trend in the direction of violation of the deadline model inequality at 3 of the 10 percentile points. These results suggest that the significant violations reported above were not due to any peculiarities of the easy and hard discrimination conditions.

The present data indicate that participants do not naturally set time deadlines—at least not in the manner suggested by the deadline model. It is still possible, however, that the participants set time deadlines in some situations, but the present procedure inadvertently discouraged use of this strategy. To evaluate this hypothesis, a control experiment ($N = 10$) was conducted in which the participants were instructed to respond before a time deadline of 390 msec⁶ in the speed stress blocks. In addition to explicit instructions to set time deadlines, the participants received immediate feedback, in the form of a 100-Hz tone, whenever they failed to respond before the deadline. In all other respects, the method of this control experiment was identical to Experiment 1.

Figure 4 shows the results from this control experiment. The participants appeared to comply with the instructions to try to respond before the imposed deadline, but the results again violated the deadline model inequality. How, if not by setting time deadlines, were the participants able to consistently respond before the 390-msec deadline? One obvious answer is that, through trial and error, the participants found the lowest evidence criterion that consistently resulted in responses faster than 390 msec (i.e., such that their responses were not often followed by the 100-Hz tone). Similar results were obtained in a further control experiment ($N = 10$) in which the participants were asked to respond as close as possible to 390 msec after stimulus onset and were given feedback on their actual RT after each response. These results do not necessarily indicate that partic-

ipants are incapable of setting time deadlines as described by the deadline model, because it is always possible that these experiments did not offer the most effective inducements to do so. Nevertheless, the results certainly argue that the participants in the present experiments were quite reluctant to use this strategy.

EXPERIMENT 2

The purpose of Experiment 2 was to test the generality of these results using a lexical decision task, in which participants decide whether or not letter strings form known English words. This particular task was chosen because it seems very different from the brightness judgments of Experiment 1 and because it is frequently studied. The key manipulation was the frequency (high vs. low) of the stimulus words. For the sake of continuity, the high- and low-frequency conditions will be referred to as the *easy* and *hard* conditions, respectively.

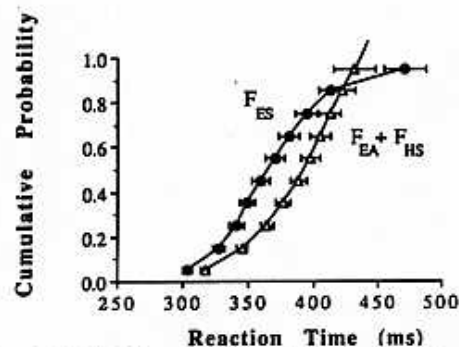


Figure 3. Plot of the cumulative distribution function of RTs in the easy discrimination/speed stress (ES) condition plus the sum of the cumulative distribution function of RTs in the easy discrimination/accuracy stress (EA) and hard discrimination/speed stress (HS) conditions of Experiment 1. Standard error bars are drawn around the mean RTs at each percentile.

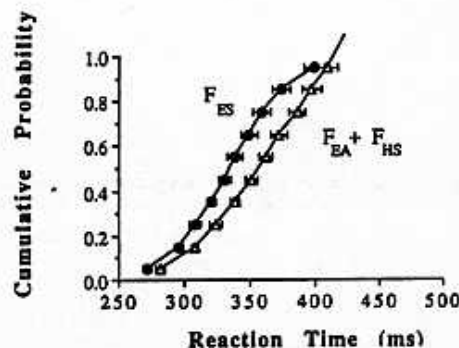


Figure 4. Plot of the cumulative distribution function of RTs in the easy discrimination/speed stress (ES) condition plus the sum of the cumulative distribution function of RTs in the easy discrimination/accuracy stress (EA) and hard discrimination/speed stress (HS) conditions of the control experiment. Standard error bars are drawn around the mean RTs at each percentile.

Table 2
Mean Reaction Time (in Milliseconds) and Percent Error (PE)
in Experiment 2 for the Speed and Accuracy Stress Conditions
as a Function of the Lexical Status of the Letter Strings
(High Frequency, Low Frequency, and Nonword)

| Condition | Lexical Status | | | | | |
|-----------------|----------------|-----|---------------|------|---------|-----|
| | High Frequency | | Low Frequency | | Nonword | |
| | RT | PE | RT | PE | RT | PE |
| Speed stress | 542 | 4.0 | 635 | 17.1 | 618 | 8.3 |
| Accuracy stress | 760 | 0.2 | 870 | 1.0 | 881 | 1.0 |

Method

Except where noted, the method was identical to that of Experiment 1.

Participants. Four male and 4 female undergraduates (between the ages of 18 and 23) at the University of California, San Diego, participated to fulfill a class requirement. The participants were native English speakers. All had normal or corrected-to-normal vision.

Stimuli. The letter strings subtended approximately 1.6° to 2.0° horizontally (depending on the number of letters) and 0.6° vertically from a typical viewing distance of 60 cm. A total of 720 English words were used in the experiment; half were high-frequency words and half were low-frequency words, drawn from Kućera and Francis (1967). The high- and low-frequency word lists each contained 120 four-letter words, 120 five-letter words, and 120 six-letter words. One nonword was generated from each of the 720 words by replacing one of the consonants with a different consonant, such that the overall frequency of the different consonant letters remained approximately the same, and so that the nonwords were orthographically and phonetically regular.

From the total pool of 720 words, 360 were randomly selected and placed into List A and 360 were placed into List B, with the restriction that each list contain equal numbers of high- and low-frequency words of every length. The 360 nonwords derived from the List A words were then placed into List B, and vice versa, so that no participant responded to a word and the nonword derived from that word. Half of the participants saw letter strings drawn from List A, and half saw letter strings drawn from List B.

Procedure. On each trial, a string of four to six letters was presented at fixation. The participants were instructed to press the "z" key with the left index finger if the letter string spelled an English word and to press the "/" key with the right index finger if it did not. Half of the participants were initially given speed stress instructions, and half were given accuracy stress instructions. The participants completed 120 practice trials under these instructions, followed by two blocks of 120 trials. The participants then completed 20 practice trials under the opposite speed/accuracy instructions, followed by two more blocks of 120 trials.

Results and Discussion

Mean RTs and error rates for Experiment 2 are shown in Table 2. The 95% confidence interval for the effect of word frequency (low vs. high) was 110 ± 21 msec under accuracy stress and 93 ± 21 msec under speed stress. Figure 5 shows the CDF from Condition ES (high frequency/speed stress) and the sum of the CDFs from Conditions EA (high frequency/accuracy stress) and HS (low frequency/speed stress). Significant violations of the deadline model inequality were observed at the 15th, 25th, 35th, 45th, and 55th percentiles. Thus, Experiment 2 provides further evidence against the deadline model.

GENERAL DISCUSSION

Two experiments were conducted to test the deadline model, according to which stimulus discrimination is terminated once (1) the discrimination process has been completed or (2) a predetermined time deadline has arrived, whichever comes first. Participants are assumed to have no control over the time required to complete the discrimination process, so they respond to speed stress only by adjusting their time deadlines. A prediction of this model, the deadline model inequality, was tested in a task involving a relative brightness judgment (Experiment 1) and in a lexical decision task (Experiment 2). Both experiments revealed clear violations of the inequality, and therefore it appears the deadline model does not provide an adequate explanation of how participants traded speed for accuracy.

These data also rule out the fast-guess model, which is simply a special case of the deadline model in which participants set either extremely short deadlines (when they decide to make a fast guess) or very long deadlines (when they decide to complete the discrimination).⁷ Using this same line of reasoning, these data also argue against a model in which participants sometimes set deadlines and sometimes make fast guesses; this is essentially a version of the deadline model in which participants often set extremely short deadlines.

On the other hand, the results present no great difficulties for models in which evidence is accumulated until a criterion amount of progress—which can be adjusted up or down by the participant—has been made (e.g., the variable criterion model of Grice & Spiker, 1979, random-walk models, accumulator models). A simple example of such a model is shown in Figure 6. Evidence accumulates stochastically over time but tends to fall within the shaded region on the left when the discrimination is easy and tends to fall within the shaded region on the right when the discrimination is hard. The accumulation of evidence terminates as soon as the amount of evidence has equaled or

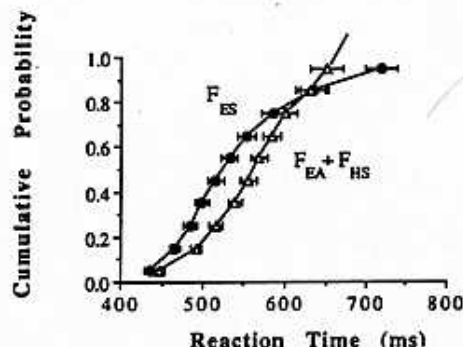


Figure 5. Plot of the cumulative distribution function of RTs in the easy (high-frequency)/speed stress (ES) condition plus the sum of the cumulative distribution function of RTs in the easy (high-frequency)/accuracy stress (EA) and hard (low-frequency)/speed stress (HS) conditions of Experiment 2. Standard error bars are drawn around the mean RTs at each percentile.

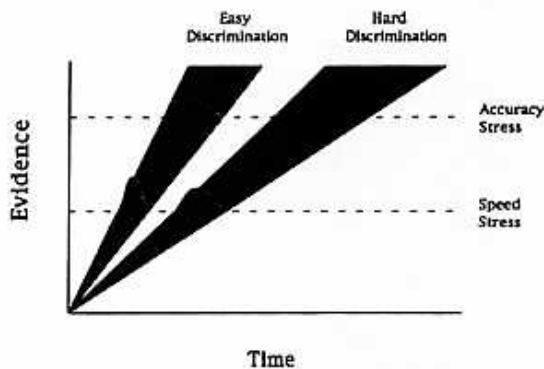


Figure 6. Hypothetical plot of evidence accumulation over time in the easy and hard discrimination conditions. Evidence accumulates stochastically over time but tends to fall within the shaded region. Accumulation terminates once the amount of evidence has reached the criterion (indicated by the dashed line), which is set to a low value under speed stress and a high value under accuracy stress. The dark polygons represent the resulting distributions of accumulation times across trials in the various experimental conditions. The accumulation times in the easy discrimination/speed stress condition are much shorter than those of the other experimental conditions, which would lead to violations of the deadline model inequality.

exceeded the criterion (the dashed line), which is set to a relatively low value under speed stress and is set to a relatively high value under accuracy stress. The dark polygons represent the resulting distributions of accumulation times across trials. As can be seen from Figure 6, there are more short accumulation times in the easy-speeded condition than there are in all the other conditions combined, which would result in violations of the deadline model inequality. Figure 6, of course, illustrates just one possible model of evidence accumulation; however, there is no obvious reason why other members of this class of models could not also predict especially fast accumulation times in the easy-speeded condition.

In conclusion, the deadline model and the fast-guess model, by themselves or in combination, cannot explain how the participants traded speed for accuracy in these experiments. On the other hand, the results are compatible with models in which participants respond to speed stress by raising or lowering an evidence criterion. The possibility remains, however, that deadlines do play some role—perhaps a minor one—in speed-accuracy tradeoff. For example, it is conceivable that participants respond to increasing speed stress both by lowering an evidence criterion and by setting shorter time deadlines. Further research will be needed to evaluate such possibilities.

REFERENCES

- EDWARDS, W. (1965). Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology*, 2, 312-329.
- GRICE, G. R., & SPIKER, V. A. (1979). Speed-accuracy tradeoff in choice reaction time: Within conditions, between conditions, and between subjects. *Perception & Psychophysics*, 26, 118-126.
- HEATH, R. A. (1984). Random-walk and accumulator models of psychophysical discrimination: A critical evaluation. *Perception*, 13, 57-65.
- KORNBLUM, S. (1973). Simple reaction time as a race between signal detection and time estimation: A paradigm and model. *Perception & Psychophysics*, 13, 108-112.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analyses of present-day American English*. Providence, RI: Brown University Press.
- LABERGE, D. (1962). A recruitment theory of simple behavior. *Psychometrika*, 27, 375-396.
- LOFTUS, G. R., & MASSON, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1, 476-490.
- MEYER, D. E., IRWIN, D. E., OSMAN, A. M., & KOUNIOS, J. (1988). The dynamics of cognition and action: Mental processes inferred from speed-accuracy decomposition. *Psychological Bulletin*, 95, 183-237.
- NICKERSON, R. S. (1969). "Same"-"different" response times: A model and a preliminary test. *Acta Psychologica*, 30, 257-275.
- NICKERSON, R. S. (1971). "Same"-"different" response times: A further test of a "counter and clock" model. *Acta Psychologica*, 35, 112-127.
- OLLMAN, R. T. (1966). Fast guesses in choice reaction time. *Psychonomic Science*, 6, 155-156.
- OLLMAN, R. T. (1977). Choice reaction time and the problem of distinguishing task effects from strategy effects. In S. Dornic (Ed.), *Attention and performance VI* (pp. 99-113). Hillsdale, NJ: Erlbaum.
- OLLMAN, R. T., & BILLINGTON, M. J. (1972). The deadline model for simple reaction times. *Cognitive Psychology*, 3, 311-336.
- PACHELLA, R. G. (1972, May). *Memory scanning under speed stress*. Paper presented at the meeting of the MPA, Cleveland.
- PACHELLA, R. G., & FISHER, D. F. (1969). Effect of stimulus degradation and similarity on the trade-off between speed and accuracy in absolute judgments. *Journal of Experimental Psychology*, 81, 7-9.
- PACHELLA, R. G., & FISHER, D. F. (1972). Hick's law and the speed-accuracy trade-off in absolute judgment. *Journal of Experimental Psychology*, 92, 378-384.
- RATCLIFF, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, 86, 446-461.
- REED, A. V. (1973, August). Speed-accuracy trade-off in recognition memory. *Science*, 181, 574-576.
- STERNBERG, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276-315.
- STONE, M. (1960). Models for reaction time. *Psychometrika*, 25, 251-260.
- SWENSSON, R. G. (1972). The elusive tradeoff: Speed vs accuracy in visual discrimination tasks. *Perception & Psychophysics*, 12, 16-32.
- TOWNSEND, J. T. (1990). Truth and consequences of ordinal differences in statistical distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 3, 551-567.
- VICKERS, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, 13, 37-58.
- VICKERS, D., & SMITH, P. (1985). Random-walk and accumulator models of psychophysical discrimination: A critical evaluation. *Perception*, 14, 471-497.
- VINCENT, S. B. (1912). The function of the vibrissae in the behavior of the white rat. *Behavioral Monographs*, 1 (No. 5).
- WILDING, J. M. (1974). Effects of stimulus discriminability on the latency distribution of identification responses. *Acta Psychologica*, 38, 483-500.
- YELLOTT, J. I., JR. (1971). Correction for fast guessing and the speed-accuracy tradeoff in choice reaction time. *Journal of Mathematical Psychology*, 8, 159-199.

NOTES

- Note that some of the possible joint distributions are undefined (e.g., the bivariate distribution of P_E and P_{H1}) because the corresponding mental processes never operate on the same trial. Also note that many of the possible joint distributions are defined only in one of the four experimental conditions because the corresponding mental processes operate on the same trial only in that condition (e.g., the bivariate distribution of P_E and D_S is only defined in Condition ES).
- Note that this form of stochastic dominance of D_S over D_A implies that the corresponding CDFs do not cross, but it does not necessarily imply that the density functions of D_S and D_A do not cross (see Townsend, 1990).

3. The trial was rerun later in the block if the participant failed to respond within 2.5 sec (<0.2% of all trials).

4. The summed CDF was obtained by first pooling together the RTs from Conditions EA and HS and then calculating an average CDF for the pooled RTs in the manner described in the previous paragraph.

5. The standard errors were calculated separately at each percentile, using the corresponding CDF-type (ES vs. EA + HS) \times participants interaction error term (see Loftus & Masson, 1994).

6. Pilot data indicated that this time deadline was short enough to force even the fastest participants to respond more quickly than they would under

accuracy stress yet long enough so that even the slowest participants were able to respond well above chance in all conditions.

7. Also note that fast guesses should produce RTs in the 150- to 250-msec range, yet there were very few responses (<1%) this fast for any of the participants.

(Manuscript received January 6, 1995;
revision accepted for publication May 14, 1995.)