Perception & Psychophysics
1976, Vol. 19 (1), 92-102

Speed-accuracy tradeoff functions in choice reaction time: Experimental designs and computational procedures

CHARLES C. WOOD and J. RICHARD JENNINGS
Walter Reed Army Institute of Research, Washington, D.C. 20012

The existence of tradeoffs between speed and accuracy is an important interpretative problem in choice reaction time (RT) experiments. A recently suggested solution to this problem is the use of complete speed-accuracy tradeoff functions as the primary dependent variable in choice RT experiments instead of a single mean RT and error rate. This paper reviews and compares existing procedures for generating empirical speed-accuracy tradeoff functions for use as dependent variables in choice RT experiments. Two major types of tradeoff function are identified, and their experimental designs and computational procedures are discussed and evaluated. Systematic disparities are demonstrated between the two tradeoff functions in both empirical and computer-simulated data. Although all existing procedures for generating speed-accuracy tradeoff functions involve empirically untested assumptions, one procedure requires less stringent assumptions and is less sensitive to sources of experimental and statistical error. This procedure involves plotting accuracy against RT over a set of experimental conditions in which subjects' criteria for speed vs. accuracy are systematically varied.

Despite the widespread use of choice reaction time (RT) as a dependent variable in contemporary information processing research, there remain a number of methodological and interpretative problems associated with RT experiments (Pachella, 1974). The present paper concerns a potential solution to one of the more important interpretative problems in choice RT data, the possibility of tradeoffs between speed and accuracy.

The term "speed-accuracy tradeoff" refers to the observation that subjects are capable of trading increases in speed for decreases in accuracy and vice versa over a substantial range (e.g., Fitts, 1966; Lappin & Disch, 1972a; Pachella & Fisher, 1969, 1972; Swensson, 1972a). In a typical RT experiment, subjects are instructed to respond as rapidly and accurately as possible. However, instead of responding at a rate that will produce zero errors, subjects virtually always perform at somewhat less than perfect accuracy. Thus, the RT and error rate values obtained in typical RT experiments represent a compromise between the incompatible demands for maximum accuracy and minimum RT. The particular compromise between speed and accuracy adopted in a given situation (i.e., the speed-accuracy *criterion*) can be manipulated by a number of variables, including instructions or explicit payoffs favoring speed or accuracy (e.g., Fitts, 1966; Lappin & Disch, 1972a; Swensson, 1972a).

The authors wish to thank R. T. Ollman for making available a copy of his unpublished manuscript (Note 1). Requests for reprints should be sent to Charles C. Wood, Division of Neuropsychiatry Walter Reed Army Institute of Research, Washington, D.C. 20012.

Two important limitations on the interpretation of RT data imposed by the possibility of speed-accuracy tradeoffs have been discussed by Pachella (1974). First, subjects' criteria for speed vs. accuracy may vary across experimental conditions, rendering impossible a comparison of RT between conditions unconfounded with changes in speed-accuracy criteria. The second and more subtle limitation is caused by the typical form of the tradeoff function relating average proportion correct to average RT. Because such functions are negatively accelerated, very small changes in error rate at high levels of accuracy may produce large changes in RT (see, for example, Figure 1 and Pachella, 1974, Figure 4). This is precisely the region of the speed-accuracy tradeoff function (i.e., error rates less than 10%) in which reported data from many choice RT experiments fall.[1]

The difficulties created by variable error rates in RT experiments have led to suggestions that the complete tradeoff function between speed and accuracy be used as a measure of RT performance in which changes in bias for speed vs. accuracy can be directly assessed (e.g., Lappin & Disch, 1972a, b; Pachella, 1974; Pew, 1969). Lappin and Disch (1972a) summarize the logic and rationale for such suggestions in the following way: "This technique is based upon the proposition that a decision process determines the point in time at which perceptual processing is terminated and an overt response is selected. Thus, a given pair of RT and error rate values reflects performance under a specific decision criterion; but the complete tradeoff function between RT and error rate is presumed to reflect performance under varying decision criteria applied to a constant perceptual process. The functional relation between two dependent variables is thereby used to obtain a decision-free measure of the perceptual process. The same approach is used in signal detection theory: The functional relation between hit rate and false-alarm rate, the receiver operating characteristic (ROC), is used as the decision-free measure of detection, rather than any specific value of hit rate or false-alarm rate" (p.420).

Pursuing the analogy with signal detection theory, the rationale for computing speed-accuracy tradeoff functions provides a means of distinguishing three different sources of a change in mean RT from one experimental condition to another: (a) a change in subjects' speed-accuracy criteria; (b) a change in processing efficiency independent of changes in speed-accuracy criteria; or (c) some combination of the two.[2] Identical tradeoff functions across experimental conditions would imply that differences in RT or accuracy were produced by changes in speed-accuracy criterion. In contrast, tradeoff functions that differ significantly across conditions would imply differences in processing efficiency independent of changes in criterion for speed vs. accuracy. This rationale has proven useful in a number of recent experiments. For example, variables whose RT effects have been associated with changes in the complete tradeoff function include stimulus discriminability (Pachella & Fisher, 1972), stimulus intensity (Lappin & Disch, 1972b; Rabbitt & Vyas, 1970), amount of practice in the task (Rabbitt & Vyas, 1970), and graded doses of alcohol (Jennings, Wood, & Lawrence, 1976). In contrast, variables whose RT effects have been primarily associated with changes along a single tradeoff function include stimulus probability (Harm & Lappin, 1973; Lappin & Disch, 1972a) and "repetition effects" in serial choice reactions (Swensson, 1972b).

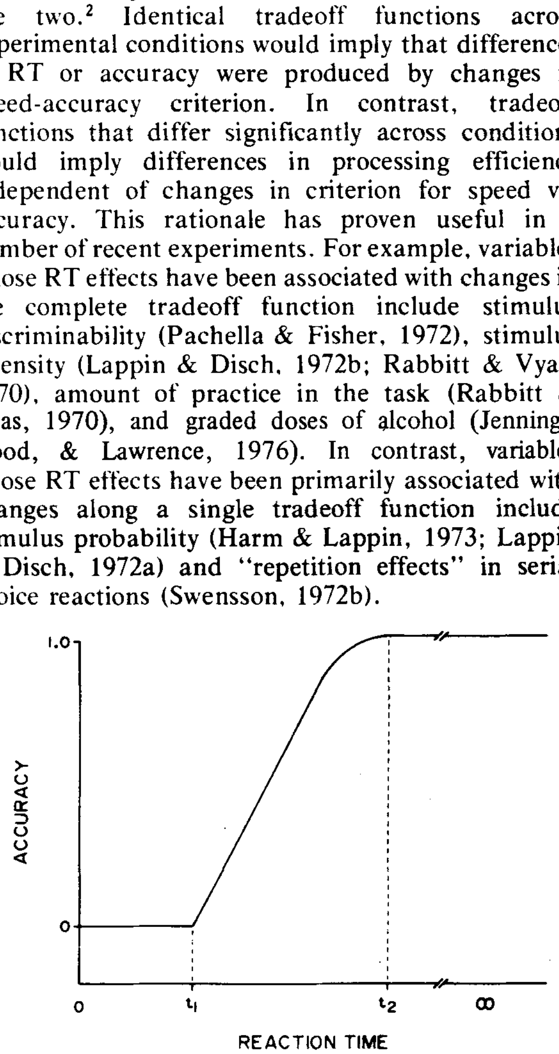

Figure 1. An idealized speed-accuracy tradeoff function in which accuracy is plotted as a function of reaction time. The value t_1 is the RT at which accuracy begins to exceed chance, while t_2 is the RT at which accuracy reaches maximum.

The purpose of the present paper is to review and compare existing procedures for generating empirical speed-accuracy tradeoff functions. First some basic characteristics of empirical tradeoff functions are outlined. Second, the major experimental designs and computational procedures used to generate such functions are reviewed and their assumptions are discussed and evaluated. Third, empirical comparisons of the two major types of tradeoff function are presented based on both experimental and computer-simulated data. Throughout this paper, the primary emphasis will be upon the use of tradeoff functions as dependent variables which do not depend upon any particular model of the tradeoff process. However, a number of the empirical issues to be considered may have implications for tradeoff models.

Basic Characteristics of Empirical Speed-Accuracy Tradeoff Functions

An idealized version of an empirical speed-accuracy tradeoff function is presented in Figure 1, in which a generalized measure of accuracy, A ($0 \leq A \leq 1$), is plotted as a function of RT. This idealized function will be used to illustrate some basic properties common to all empirical speed-accuracy tradeoff functions, regardless of the procedures by which they are derived. Specific procedures for deriving such functions are considered in detail in subsequent sections.

The idealized function in Figure 1 may be divided into three distinct regions: (a) a region in which $RT < t_1$ and $A \cong 0$; (b) a region in which $t_1 \leq RT \leq t_2$ and A is some increasing function of RT; and (c) a region in which $RT > t_2$ and $A \cong 1$.

The first region represents the average interval necessary for accuracy to exceed levels expected by chance. Some models of the tradeoff process assume that the upper bound of this region, t_1, approximates the mean RT required for simple detection of the stimuli involved, where no choice response is required (e.g., Edwards, 1965; Stone, 1960). However, data presented by Swensson (1972a) suggest that the duration of the first region may be significantly longer than the "simple RT" for the same stimuli. Pending additional data, a strictly operational interpretation of the region $RT < t_1$ appears preferable to interpretations in terms of specific processing stages.

The second region of the idealized function in Figure 1 is the region that has received most empirical and theoretical attention. In this region, A is an increasing function of RT. Although the exact form of this increasing function remains in question, it has been typically approximated by the linear equation

$$A = m(RT - c), \qquad (1)$$

in which A and RT represent measures of accuracy and reaction time, respectively; c corresponds to t_1 in Figure 1, the intercept of the function with the RT axis at $A = 0$; and m is the slope of the function reflecting the amount of increase in A for each unit increase in RT. The linear relationship between A and RT expressed by Equation 1 has typically accounted for greater than 80% of the empirical variance in these two variables (e.g., Lappin & Disch, 1972a).

Under conditions of adequate linear fit, the parameters m and c may be used as summary statistics for an analysis of the effects of various independent variables on the tradeoff function. For example, Lappin and Disch (1972b) reported that increases in stimulus intensity produced both an increase in the slope and a decrease in the intercept of empirical tradeoff functions. At the present stage of development, it is unclear whether the slope and intercept parameters represent completely independent or partially correlated aspects of performance. Therefore, in addition to separate comparisons of the slope and intercept parameters across experimental conditions, it may be useful to combine the slope and intercept into a single measure representing the overall level of performance in each condition. One way of combining the slope and intercept parameters is to insert fixed values of accuracy into Equation 1 for each condition and solve for the corresponding values of RT. In this way, an "equal-accuracy contour" may be derived which represents the effect of the different experimental conditions on RT at a constant level of accuracy. Similarly, "equal-RT contours" may be derived by inserting fixed values of RT into Equation 1 and solving for corresponding values of accuracy. A second way to combine the slope and intercept parameters is to compute the integrated area under Equation 1 between the intercept and some arbitrary value of RT near the upper end of the range of empirically obtained RTs, resulting in a measure reflecting the total area under the tradeoff function. This procedure is analogous to computing the area under an ROC function in a signal detection experiment (Green & Swets, 1966).

The parameter estimates for m and c in Equation 1 as well as the other measures just discussed all depend upon the particular scale of measurement selected for A. Although different transformations of proportion correct, P(C), are specified as linearly related to RT by different models and theories, to date no single transformation has been shown to provide consistently better fits to empirical data than various alternatives. For example, Lappin and Disch (1972a) reported no consistent differences in least squares goodness of fit among measures of A including: (a) d', the detectability measure of signal detection theory; (b) $(d')^2$ discussed by Taylor, Lindsay, and Forbes (1967); (c) H_t, the information transmitted between stimuli and responses; and (d) $-\ln\eta$, the sensitivity parameter of Luce's (1963) choice theory. The average proportion of variance accounted for by Equation 1 ranged between .86 and .91 for the different A measures. Roughly equivalent fits have been reported by Swensson (1972a) for $(d')^2$, H_t, and "odds" transformations derived from stimulus sampling models. At the present time, therefore, a choice among alternative scales for A need not imply acceptance of a particular model or theory over other alternatives.

The asymptotic third region of the idealized function in Figure 1, in which $RT > t_2$ and $A \cong 1$, has received only limited empirical attention. This region deserves additional study, however, since it may provide an important link between the study of discrimination under time limitations and the more traditional approaches to discrimination in which accuracy is the primary performance limitation (see Swensson, 1972a).

Experimental Designs and Computational Procedures

Two major methods have been used to generate speed-accuracy tradeoff functions from empirical data. Ollman (Note 1) has recently discussed theoretical aspects of these two types of tradeoff function, and his definitions and terminology are employed in the present paper. The first function is termed the *speed-accuracy tradeoff function* (SATF). This function is defined as the relationship between accuracy and RT across a set of experimental conditions designed to manipulate subjects' speed-accuracy criteria systematically over a wide range. For example, consider an experiment in which subjects are rewarded for correct responses which occur before a specified RT "deadline" on each trial. By systematically varying the length of the deadline interval in different experimental conditions, a wide range of speed and accuracy of performance may be obtained. The SATF for such an experiment would be computed simply by plotting P(C) (or some appropriate transform) against mean RT (or other measure of central tendency) for each of the different deadline conditions.

The type of design in which a number of different experimental conditions are used to vary subjects' speed-accuracy criteria will be referred to as a *multiple-condition* design. At a minimum, two such conditions are necessary to generate data for an SATF, although several conditions are typically used in practice to provide a more fine-grained estimate of the function. In addition to the RT deadline procedure described above (cf. Fitts, 1966; Green & Luce, 1973; Link, 1971; Pachella & Fisher, 1969, 1972; Pachella & Pew, 1968), a second type of multiple-condition design involves no explicit deadline intervals but relies instead upon a continuous payoff scheme that systematically rewards different speed-accuracy criteria in different experimental conditions (e.g., Swensson, 1972a). A third procedure which combines aspects of both the deadline and continous payoff methods is the "band payoff" procedure introduced by Snodgrass, Luce, and Galanter (1967). In this procedure, payoffs decrease according to some specified function as the obtained RT deviates further in either direction from a specified target value. Thus, a variety of different procedures can be used to manipulate subjects' speed-accuracy criteria in a multiple-condition design. For convenience, such conditions will be referred to as speed-emphasis conditions, regardless of the specific form of the criterion manipulation. While there are a number of interesting psychological questions concerning different speed-emphasis manipulations (e.g., Swensson & Thomas, 1974), such questions are beyond the scope of the present paper and will not be considered further.

The second major type of speed-accuracy tradeoff function is termed the *conditional accuracy function* (**CAF**).[3] This function is formally defined as the conditional probability of a correct response given that RT equals a particular value t, computed at all values of t. In practice, the CAF is computed by ranking the RTs in a given experiment from low to high, partitioning the data into equal-N categories, and plotting accuracy against mean RT for each RT category (e.g., Lappin & Disch, 1972a; Rabbitt & Vyas, 1970).

In contrast to the SATF, the CAF does not require a multiple-condition design and can be computed from a single RT distribution. In principle, therefore, the CAF can be computed from any choice RT experiment in which both speed and accuracy of performance are measured. In practice, the CAF has often been computed from an experimental design which shall be termed a *single-condition* design to distinguish it from the multiple-condition design required for the SATF. Instead of using different speed-emphasis conditions to manipulate subjects' speed-accuracy criteria, the single-condition design instructs subjects to vary their criteria over a wide range in a single experimental condition. An example of this design is the experiment of Lappin and Disch (1972a), in which "well-practiced subjects were asked to respond at a speed that would result in an error rate of about 25%. They were instructed to avoid guessing, but to adjust their RTs from trial to trial in order to maintain the appropriate speed and error rate." (p. 422). Other examples of CAFs derived from single-condition designs have been reported by Harm and Lappin (1973), Lappin and Disch (1972b, 1973), and Rabbit and Vyas (1970).

Despite their superficial similarities, the SATF and CAF differ in a number of important respects. Some differences are associated with the experimental designs typically used to generate each type of function, while other differences are associated with the computational procedures used to derive the two functions. Since the CAF is not limited to single-condition designs, it is particularly important in comparing the SATF and CAF to distinguish between factors associated with the experimental designs and factors associated with the computational procedures per se.

The multiple-condition design required for the SATF and the single-condition design typically used for the CAF differ in three major ways. The first difference concerns the specific means used to vary subjects' speed-accuracy criteria in the two designs. The multiple-condition design employs explicit experimental manipulations to induce systematic variations in subjects' criteria, while the single-condition design relies on subjects to vary their criteria appropriately over the course of a single experimental condition. These procedural differences may significantly influence both the range and stability of criterion variations actually obtained in a given experiment. The second difference between multiple- and single-condition designs concerns the experimenter's ability to determine whether subjects' criteria have varied appropriately in a given experiment. Although there are no means presently available for determining the exact range of criteria employed in a given experiment, the multiple-condition design at least permits the experimenter to determine whether RT and accuracy have varied systematically across speed-emphasis conditions. In contrast, the single-condition design provides no means other than the single RT distribution and error rate for empirically assessing whether subjects varied their criteria according to instructions. The third major difference between multiple- and single-condition designs concerns the interaction of experimental variables with subjects' speed-accuracy criteria. Both designs assume that the variations in criteria do not interact with performance in other ways. Although no systematic investigations of this assumption have been reported, the single-condition design seems less likely to produce interactions between criterion variations and other aspects of performance than the multiple-condition design in which subjects' criteria are explicitly manipulated.

In addition to the differences in experimental designs just described, the computational procedures used to derive the SATF and CAF also differ significantly. This difference concerns the way in which the data are partitioned into categories prior to computing RT and accuracy values, and has significant implications for the assumptions required by the two procedures. The SATF partitions the data into categories based on the different speed-emphasis conditions in a multiple-condition experiment. The mean RT and accuracy values for each speed-emphasis condition are then used as summary statistics for performance over the range of speed-accuracy criteria which actually occurred in that condition. Note that this procedure requires only relatively weak assumptions concerning variations in subjects' speed-accuracy criteria. First, the SATF requires that subjects' criteria vary systematically across the different speed-emphasis conditions. As noted above, this assumption can be assessed in the RT and accuracy data from any given experiment. Second, it is not necessary to assume that a single criterion or even a small range of criteria are employed within a given speed-emphasis condition. Rather, the mean RT and accuracy values for each condition are assumed to be reasonable summary statistics for the bivariate speed-accuracy distribution associated with the *average* speed-accuracy criterion employed in that condition.[4]

In contrast, the CAF partitions the data into categories based on *obtained RT* instead of on some external basis as does the SATF. This fact makes the CAF dependent upon very specific assumptions concerning variability in subjects' speed-accuracy criteria. By computing accuracy conditional upon obtained RT, the CAF implicitly assumes either: (a) that the CAF is invariant across changes in speed-accuracy criteria; or alternatively (b) that a given subject's speed-accuracy criterion is constant in the data from which a CAF is computed. If neither of these conditions is met, then computing accuracy conditional upon obtained RT would mix together trials which have similar RTs but which arise from high-accuracy and low-accuracy criteria.

The assumption of a constant speed-accuracy criterion in the data contributing to a given CAF is difficult to justify on logical grounds and cannot be tested empirically. In the single-condition design typically used to generate a CAF, subjects are instructed to vary their speed-accuracy criteria from trial to trial in order to maintain a specified average RT and error rate as noted above. Indeed, Lappin and Disch (1972a) explicitly noted that their design "virtually guarantees that the decision criteria did not remain fixed" (p. 423). A possible way to reduce criterion variability is to compute a CAF from data obtained in a single speed-emphasis condition of a multiple-condition design. However, the a priori likelihood that a given empirical RT distribution results from a constant criterion is very low, even in the data from a single speed-emphasis condition. More importantly, the degree of criterion variability in a single RT distribution cannot be empirically determined regardless of the experimental design used to generate the data. That is, it is impossible to distinguish RT variability associated with variations in speed-accuracy criteria from RT variability arising from a single speed-accuracy criterion. Thus, regardless of whether a CAF is computed from a single-condition design like that of Lappin and Disch (1972a) or from a single speed-emphasis condition in a multiple-condition design, the constant criterion assumption is probably unjustified and, in any case, cannot be empirically tested.

The alternative assumption, that the CAF is invariant across changes in speed-accuracy criterion, may be more reasonable a priori than the constant criterion assumption. If the CAF were invariant across changes in criterion, variability in subjects' criteria would be immaterial since the CAFs for different criteria would be the same. Unlike the assumption of a constant speed-accuracy criterion, the question of CAF invariance is amenable to direct empirical test. The nature of this empirical test, together with theoretical aspects of the CAF invariance question, have been recently discussed by Ollman (Note 1) in the context of the data reported by Schouten and Bekker (1967).

Schouten and Bekker (1967) employed a multiple-condition design with different pacing signals ("intended RTs") to manipulate subjects' speed-accuracy criteria. Subjects were instructed to respond "in coincidence" with the pacing signal on each trial. The interval between the stimuli and the pacing signals was constant within a block of trials and varied systematically across blocks from 100 to 800 msec. Separate CAFs were computed for each individual speed-emphasis condition. Based on their results, Schouten and Bekker (1967) made the following important observation: "The fractions of errors obtained from different enforced τs [i.e., the different speed emphasis conditions] reasonably fit *one* standard curve. An actual reaction time of, say, 250 msec can result from both an induced reaction time of $\tau = 200$ msec and $\tau = 300$ msec. The experimental fact that all fractions of errors fit one standard curve then means that this fraction is determined by the actual rather than the intended reaction time" (p. 146).

Based on the Schouten and Bekker (1967) data, Ollman (Note 1) introduced an important theorem concerning the relation between the CAF and SATF. In effect, this theorem states that if the CAF is invariant across changes in speed-accuracy criteria, then the CAF is a close approximation to the SATF for the same set of data. However, as Ollman (Note 1) points out, the empirical utility of this theorem depends entirely upon the assumption of CAF invariance across changes in speed-accuracy criterion. At present, the only available data in which this assumption may be tested are those of Schouten and Bekker (1967). Moreover, although Schouten and Bekker (1967) concluded that their functions fit "one standard curve," statistical support for this conclusion was not presented and close inspection of their data reveals a small but systematic departure from invariance (see especially their Figure 4). These

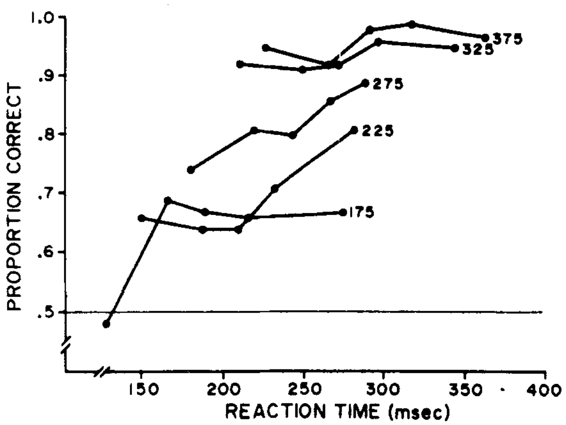

Figure 2. Conditional accuracy functions derived from different speed-emphasis (RT deadline) conditions. The nominal deadlines in milliseconds for each condition are shown at the right of each function. The line at P(C) = .5 represents chance performance.

observations raise the possibility that the apparent support for CAF invariance in the data of Schouten and Bekker (1967) resulted not from an invariant CAF across speed-emphasis conditions but from significant variability in subjects' speed-accuracy criteria *within* each condition.[5] In light of this possibility, an independent test of CAF invariance was deemed highly desirable. The results of this test are presented in the following section, which compares empirical CAFs from different speed-emphasis conditions in a manner analogous to that described by Schouten and Bekker (1967).

Comparison of Empirical Conditional Accuracy Functions from Different Speed-Emphasis Conditions

The present comparison of CAFs across different speed-emphasis conditions was derived from a recent experiment concerning the effects of graded doses of alcohol on speed-accuracy tradeoff functions in choice RT (Jennings, Wood, & Lawrence, 1976). A multiple-condition design was employed in which five signaled deadline intervals were used to manipulate subjects' speed-accuracy criteria. The deadline intervals were 175, 225, 275, 325, and 375 msec, each presented in a separate block of 100 trials. A daily session consisted of one block of trials for each deadline condition, preceded by a block with 450-msec deadline presented for practice at the beginning of each session.

Two pure tone stimuli (1,000 and 1,100 Hz, 70 dB SL) were assigned to two alternative response buttons. On each trial, the acoustic stimulus was followed by a visual stimulus which signaled the end of the deadline interval. Five subjects received at least two sessions of practice before beginning a five-dose alcohol schedule. The data presented here were derived from the last practice session for each subject prior to any experience with the alcohol manipulation.

Separate CAFs were computed for each deadline condition for each subject in the following manner. The RTs for each condition were ranked from low to high, divided into five equal-N categories, and P(C) and mean RT for each category were computed. This procedure resulted in five separate CAFs for each subject, each CAF consisting of five joint speed-accuracy values (i.e., mean RT and P(C) for each quintile of each RT distribution). Mean CAFs across the five subjects are shown in Figure 2. In this figure, P(C) is plotted as a function of mean RT for each RT quintile in each of the five deadline conditions.

In marked contrast to the apparent invariance in the Schouten and Bekker (1967) data, the results shown in Figure 2 reveal systematic differences in the CAFs for the different speed-emphasis conditions. For example, in the RT range between 225 and 275 msec, mean P(C) across the five subjects ranged from approximately .67 in the 175-msec deadline condition to .94 in the 375-msec condition. For intermediate deadline conditions, P(C) was an increasing function of the nominal deadline interval.

Statistical comparison of the CAFs for different deadline conditions is complicated by the varying degree of CAF overlap along the RT axis. To evaluate the statistical reliability of the results shown in Figure 2, the RT interval between 225 and 275 msec was selected as having a sufficient number of trials in all CAFs for each subject to permit reliable comparison. Accordingly, the P(C) values for this interval were computed for each subject and deadline condition and were compared using the Friedman nonparametric analysis of variance procedure (Siegel, 1956). The results of this analysis indicated that the average P(C) values for $225 < RT < 275$ differed significantly as a function of deadline condition $\chi^2_r = 16.20$, df = 4, $p < .01$.

These results clearly demonstrate that the CAF was not invariant across changes in speed-accuracy criteria in the present experiment. That is, accuracy depended not only upon obtained RT but also upon the speed-emphasis condition in which a given trial occurred. This conclusion should not, however, be overgeneralized. The statistical considerations discussed by Ollman (Note 1) indicate that a sensitive test of CAF invariance requires far more trials in each CAF than were available for the present comparison. Nevertheless, most of the sources of statistical and experimental error in such a comparison are biased in favor of accepting the hypothesis of CAF invariance. Therefore, the clear failure of CAF invariance under the relatively insensitive conditions of the present comparison suggests that CAF invariance should not be assumed in a given experiment without direct empirical support.[6]

Comparison of Conditional Accuracy Functions and Speed-Accuracy Tradeoff Functions in Empirical and Computer-Simulated Data

The results presented in the preceding section raise significant questions concerning the suitability of the CAF for use as a bias-controlled dependent variable in choice RT experiments. Together with the evidence against CAF invariance just presented, the inability to determine the degree of criterion variability in a given RT distribution suggests that empirical CAFs may be influenced significantly by variations in subjects' speed-accuracy criteria. In order to assess the expected magnitude of criterion variability effects on the CAF, this section first compares CAFs and SATFs derived from empirical data containing variability in subjects' speed-accuracy criteria. Second, computer-simulated data are presented in which criterion variability effects are directly investigated by systematically varying the RT distributions representing different speed-accuracy criteria.

It should be emphasized that both the empirical and computer-simulated comparisons in this section are based on data corresponding to a multiple-condition design. Thus, these comparisons illustrate the consequences of computing CAFs over data which are presumed to contain systematic criterion variability based on the demonstrated effectiveness of the speed-emphasis conditions. The inference is made that similar effects can occur when CAFs are computed from data in which the magnitude of criterion variability cannot be directly assessed.

Figures 3a and 3b compare SATFs (filled circles) and CAFs (open circles) for two typical subjects from the experiment discussed in the preceding section. The solid and dotted lines are the least-squares solution to Equation 1 with A = P(C) for the SATF and CAF, respectively. The consequence of mixing data from different speed-accuracy criteria in the CAFs for these data is that both the slope and the intercept of the CAFs are lower than those of the corresponding SATFs. The reasons for this disparity may be seen intuitively by reexamination of Figure 2. An SATF for the data in Figure 2 would plot P(C) against mean RT for each of the five deadline conditions, resulting in a function connecting the "centroids" of the within-condition CAFs representing differing average criteria. In contrast, the overall CAF for these data would first rank the RTs across all deadline conditions, then compute P(C) and mean RT based on equal-N categories across the entire range of RT. The net result of mixing data representing different criteria is to increase the accuracy reflected by the overall CAF relative to the SATF at short RTs, while this effect is reversed at long RTs. In effect, then, a CAF computed on data having substantial criterion variability yields a function which reflects neither the SATF nor the individual CAFs representing each speed-accuracy criterion, but some combination of the two.

A related consequence of the ranking procedure required by the CAF is that the magnitude of criterion variability is directly related to the amount of overlap between the RT distributions representing different speed-accuracy criteria. This relationship was first noted in empirical comparisons of CAFs and SATFs like those shown in Figure 3. Subjects showing smaller disparities between CAFs and SATFs also had less variable RT distributions in the different speed-emphasis conditions. For example, the mean standard deviation (SD) over the five deadline conditions for the subject represented in Figure 3a was 62.49 msec. The corresponding average SD for the subject with the smaller disparity between CAF and SATF (Figure 3b) was 52.42 msec.

In order to illustrate the range of disparities between SATF and CAF which can be produced by different amounts of overlap, computer-simulated data are presented below in which two variables related to such overlap were systematically varied. The first variable is simply the standard deviation (SD) of the RT distributions associated with each speed-accuracy criterion (SD = 10, 40, and 70 msec, respectively). The second variable is the difference in mean RT between correct and error responses associated with each criterion ($RT_c - RT_e$ = 25, 0, and -25 msec, respectively).[7] Thomas (1974) and Ollman (Note 1) have independently demonstrated that $RT_c - RT_e$ for a given distribution is directly related to the form of the CAF for that distribution:

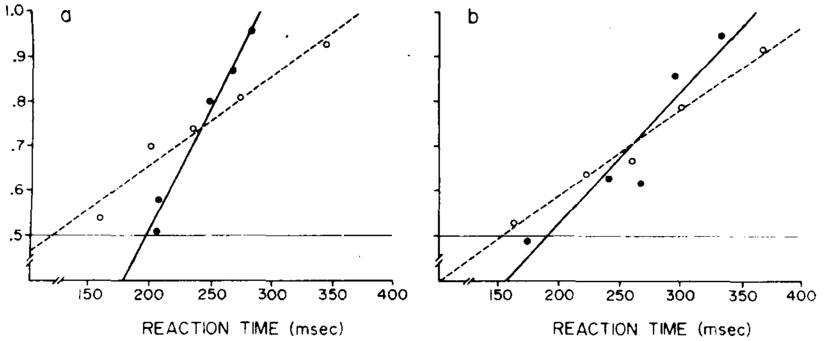

Figure 3. Speed-accuracy tradeoff functions (filled circles and solid lines) and conditional accuracy functions (open circles and dotted lines) derived from the same sets of data for each of two subjects. The line at P(C) = .5 represents chance performance.

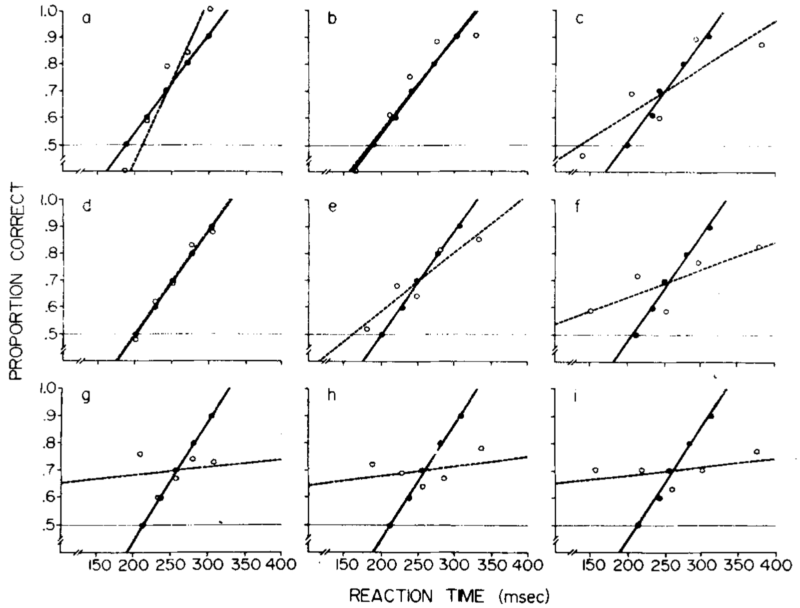

Figure 4. Speed-accuracy tradeoff functions (filled circles and solid lines) and conditional accuracy functions (open circles and dotted lines) derived from computer-simulated data. The three columns differ in the variability of RT distributions for each speed-accuracy criterion (SD = 10, 40, and 70 msec from left to right). The three rows differ in the relationship between the mean RTs for correct and error responses ($RT_c - RT_e$ = 25, 0, and -25 msec from top to bottom). The line at P(C) = .5 represents chance performance.

positive values of $RT_c - RT_e$ indicate CAFs with positive slopes, negative values of $RT_c - RT_e$ indicate CAFs with negative slopes, and RT distributions with $RT_c - RT_e = 0$ indicate CAFs with zero slope. Thus, in addition to influencing the amount of overlap between RT distributions from different criteria, the variations in $RT_c - RT_e$ also illustrate the effects of variation in the form of the CAFs for individual criteria on the overall CAF computed across variations in criteria.

The three values of SD and the three values of $RT_c - RT_e$ were combined factorially to produce the nine pairs of functions presented in Figure 4. Functions in different columns differ in SD (10, 40, and 70 msec from left to right), while functions in different rows differ in $RT_c - RT_e$ (25, 0, and -25 msec from top to bottom). Figure 4f, for example, shows the functions with SD = 70 msec and equal mean RTs for correct and error responses.

The pairs of functions shown in Figure 4 illustrate how the magnitude of the disparity between a CAF and a SATF derived from the same data varies systematically as a function of the variability of the individual RT distributions and the relation between correct and error distributions. Consider first Figure 4d, in which the mean RTs for correct and error responses are equal ($RT_c = RT_e$) and the RT variability is relatively low (SD = 10 msec). In this case, there is very little overlap in the RT distributions for different speed-emphasis conditions and, consequently, very little effect of plotting accuracy contingent upon obtained RT in the CAF. However, increases in the overlap between adjacent RT distributions caused by increasing SD (center and right columns of Figure 4) are associated with increasing disparity between CAFs and SATFs (cf. Figures 4d, 4e, and 4f).

The effect of variations in the relationship between RT distributions for correct and error responses (and therefore the slope of the individual CAFs) may be illustrated by comparing the different rows in Figure 4. When errors are faster on average than correct responses (top row, $RT_c - RT_e$ = 25 msec), the CAFs are steeper than the CAFs in the middle row which have the same SD but equal correct and error RTs. In contrast, when errors are slower on average than correct responses (bottom row, $RT_c - RT_e$ = -25 msec), CAFs are much less steep than the corresponding CAFs with equal correct and error RTs. The CAFs in the bottom row are nonmonotonic functions, with accuracy first decreasing and then increasing with increasing RT. The least squares linear solutions for these CAFs account for little of the total variance and are included only for consistency with the remainder of the figure. Like the effect of increases in SD, these effects of the relative speed of correct and error responses are also produced by mixing responses from different average speed-accuracy criteria.

The comparisons of CAFs and SATFs presented in this section should serve to increase suspicion of the CAF as a suitable function for general use as a dependent variable in choice RT experiments. The computer simulations demonstrate that the effects of criterion variability are highly variable and depend upon the RT distributions for individual speed-accuracy criteria which cannot be empirically measured. Both the degree of variability associated with each criterion and the form of the CAFs for individual criteria can exert substantial influences on a CAF computed over different speed-accuracy criteria. In contrast, the SATF is by definition less sensitive to criterion variability than the CAF, and the results of the computer simulations verify this difference between the two functions. While the SATF remained relatively stable over large changes in the variability of RT distributions representing different speed-accuracy criteria, the CAFs ranged from essentially flat when $RT_c - RT_e < 0$ to very steep when $RT_c - RT_e > 0$, and varied systematically with changes in SD. Thus, unless it becomes possible to determine the degree and form of criterion variability in the RT distribution from a single condition, it appears to be essentially impossible to estimate the magnitude of criterion variability effects in a given empirical CAF.

Since a number of variables other than those illustrated here can potentially affect the form and magnitude of criterion variability effects on the CAF, the results of this section should be viewed as illustrative rather than definitive in nature. Nevertheless, the disparities between CAFs and SATFs illustrated here are not atypical, since those shown in Figures 4c, 4e, and 4f conform closely to those obtained in empirical data (cf. Figure 3).

Conclusions

The purpose of this paper has been to review and evaluate alternative procedures for generating empirical speed-accuracy tradeoff functions for use as bias-controlled dependent variables in choice RT experiments. Although the primary focus has been upon empirical functions which do not depend upon any particular model of the tradeoff process, the results also have implications for the use of tradeoff functions in testing alternative models.

From an empirical perspective, an analysis of the experimental designs and computational procedures for the SATF and CAF suggests that the two functions differ both in terms of complexity of experimental design and also in the stringency of their underlying assumptions. In terms of experimental complexity, the CAF would appear to be far superior. The procedure of computing accuracy conditional upon obtained RT allows the CAF to be computed from any RT distribution in which both RT and accuracy are recorded on each trial. Thus, the CAF requires only minimal experimental effort over that required in a traditional choice RT experiment. In contrast, the SATF requires a specialized experimental design involving multiple speed-emphasis conditions.

The simplicity of experimental design required for the CAF is not without significant cost, however, since computing accuracy conditional upon obtained RT requires stringent assumptions which are either less justifiable or less testable than those of the SATF. Specifically, the CAF requires either: (a) that the CAF is invariant across changes in subjects' speed-accuracy criteria; or (b) that a given subject's criterion is constant in the data from which a CAF is computed. An empirical test of the CAF invariance assumption was presented in which this assumption clearly failed. The constant-criterion assumption is difficult to justify on logical grounds and cannot be tested empirically. In contrast, the SATF requires relatively weak assumptions concerning variability in subjects' speed-accuracy criteria. First, subjects' criteria must vary systematically across the different speed-emphasis conditions. This requirement can be assessed in the RT and accuracy data from any given experiment. Second, the mean RT and accuracy values for each speed-emphasis condition are assumed to be reasonable summary statistics for the bivariate speed-accuracy distribution associated with the average criterion in that condition. This assumption is reasonable a priori and is not contradicted by available data.

The empirical consequences of computing CAFs over data with variability in subjects' speed-accuracy criteria were examined by comparing CAFs and SATFs in: (a) empirical data containing criterion variability, and (b) computer-simulated data in which the form and magnitude of criterion variability were systematically varied. If the consequences of computing CAFs over data with criterion variability were either small or predictable, then the experimental simplicity of the CAF might outweigh its more restrictive assumptions. However, the results of the computer simulations demonstrated that the effects of criterion variability on the CAF were both large and variable, and were dependent upon variables which cannot be directly measured in empirical RT distributions. In contrast, the SATF is by definition relatively unaffected by the same changes in criterion variability which produced large alterations in the CAF.

Despite these reservations concerning the general use of the CAF as a dependent variable, the CAF is an important theoretical tool and may have empirical utility in certain circumstances. Although both the CAF and SATF have been used to test various theories and models of the speed-accuracy tradeoff process (e.g., Green & Luce, 1973; Lappin & Disch, 1972a; Swensson, 1972a), the CAF appears to be more discriminating among alternative models than the SATF (cf. Pachella, 1974; Thomas, 1974). However, theoretical predictions about the CAF are limited in that they have typically been derived under the assumption of a constant speed-accuracy criterion. The incorporation of criterion variability effects into the predictions of existing models would appear to be a particularly useful theoretical advance. The CAF may also be useful in a strictly empirical sense when a multiple-condition design is not experimentally feasible or when the potential effects of criterion variability are inconsequential for the specific application involved. In addition, if criterion variability can be assumed to be equal in different conditions of a choice RT experiment, then the CAF might be a reasonable means of comparing performance across conditions.

In summary, the function relating accuracy to RT over multiple speed-emphasis conditions (the SATF) would appear to be the best available procedure for generating empirical speed-accuracy tradeoff functions. This function is simple to compute, is dependent upon less stringent assumptions than the CAF, and is less sensitive to variability in subjects' speed-accuracy criteria. Although the multiple-condition design required to compute the SATF is clearly more complex than the design required for the CAF, its additional benefits appear to outweigh the additional cost. This conclusion must be tempered by the as yet untested possibility that the payoff or deadline manipulations associated with the SATF may interact with the choice RT task in ways other than simply manipulating subjects' speed-accuracy criteria. For example, the multiple-condition design may be thought of as a "secondary task" paradigm of the type discussed by Kerr (1973) in the sense that adhering to experimenter-imposed speed-emphasis conditions may represent an additional demand on processing capacity over and above the choice RT task being investigated. In view of the potential utility of the SATF in a wide variety of experimental contexts, this possibility merits systematic investigation.

REFERENCE NOTE

1. Ollman, R. T. *On the similarity between two types of speed/accuracy tradeoff.* Unpublished manuscript, 1974.

REFERENCES

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.

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

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

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

Grice, G. R., Hunt, R. L., Kushner, B. A., & Morrow, C. Stimulus intensity, catch trial effects, and the speed-accuracy tradeoff in reaction time: A variable criterion theory interpretation. *Memory & Cognition*, 1974, **2**, 758-770.

Harm, O. J., & Lappin, J. S. Probability, compatibility, speed, and accuracy. *Journal of Experimental Psychology*, 1973, **100**, 416-418.

Jennings, J. R., Wood, C. C., & Lawrence, B. E. Effects of graded doses of alcohol on speed-accuracy tradeoff in choice reaction time. *Perception & Psychophysics*, 1976, **19**, 85-91.

Kerr, B. Processing demands during mental operations. *Memory & Cognition*, 1973, **1**, 401-412.

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

Lappin, J. S., & Disch, K. The latency operating characteristic: II. Effects of visual stimulus intensity on choice reaction time. *Journal of Experimental Psychology*, 1972, **93**, 367-372. (b)

Lappin, J. S., & Disch, K. Latency operating characteristic: III. Temporal uncertainty effects. *Journal of Experimental Psychology*, 1973, **98**, 279-285.

Link, S. W. Applying RT deadlines to discrimination reaction time. *Psychonomic Science*, 1971, **25**, 355-358.

Luce, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. I). New York: Wiley, 1963.

Pachella, R. G. The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition.* Potomac, Md: Erlbaum Associates, 1974.

Pachella, R. G., & Fisher, D. F. Effect of stimulus degradation and similarity in the trade-off between speed and accuracy in absolute judgments. *Journal of Experimental Psychology*, 1969, **81**, 7-9.

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

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

Pew, R. W. The speed-accuracy operating characteristic. In W. G. Koster (Ed.), *Attention and performance II.* Amsterdam: North Holland, 1969.

Rabbitt, P. M. A., & Vyas, S. M. An elementary preliminary taxonomy for some errors in laboratory choice RT tasks. In A. F. Sanders (Ed.), *Attention and performance III.* Amsterdam: North Holland, 1970.

Schouten, J. F., & Bekker, J. A. M. Reaction time and accuracy. In A. F. Sanders (Ed.), *Attention and performance.* Amsterdam: North Holland, 1967.

Siegel, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

Snodgrass, J. G., Luce, R. D., & Galanter, E. Some experiments on simple and choice reaction time. *Journal of Experimental Psychology*, 1967, **27**, 143-153.

Stone, M. Models for choice-reaction time. *Psychometrika*, 1960, **25**, 251-260.

Swensson, R. G. The elusive tradeoff: Speed versus accuracy in visual discrimination tasks. *Perception & Psychophysics*, 1972, **12**, 16-32. (a)

Swensson, R. G. Trade-off bias and efficiency effects in serial choice reactions. *Journal of Experimental Psychology*, 1972, **95**, 397-407. (b)

Swensson, R. G., & Thomas, R. E. Fixed and optional stopping models for two-choice discrimination times. *Journal of Mathematical Psychology*, 1974, **11**, 213-236.

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.* Amsterdam: North Holland, 1967.

Thomas, E. A. C. The selectivity of preparation. *Psychological Review*, 1974, **81**, 442-464.

Wickelgren, W. A. Dynamics of retrieval. In D. Deutsch and J. A. Deutsch (Eds.), *Short-term memory.* New York: Academic Press, in press.

NOTES

1. Even further, Pachella (1974) notes: "In the extreme, if subjects actually produced zero errors in all conditions (as the general reaction time instructions ask of them), the reaction times would be essentially uninterpretable. This is because an infinite number of average reaction times can result in zero errors . . . Thus, very low error rates, while often the mark of a careful experiment, may also result in artifactual differences in reaction time" (p. 63).

2. The term "efficiency" used to refer to the bias-free component of RT performance in the present paper was introduced by Swensson (1972b). This component has been referred to as "capacity" by Thomas (1974) and "perceptual processing" by Lappin and Disch (1972a). Other terms for "speed-accuracy criterion" used in the present paper have been "trade-off bias" (Swensson, 1972b) and "decision criterion" (Lappin & Disch, 1972a).

3. A number of different terms have been used to refer to empirical functions relating speed and accuracy of performance. Functions analogous to the CAF have been termed "latency operating characteristics" by Lappin and Disch (1972a), "T functions" by Rabbit and Vyas (1970), and "micro-tradeoff" functions by Pachella (1974) and Thomas (1974). Functions analogous to the SATF have been termed "speed-accuracy operating characteristics" by Pew (1969) and "macro-tradeoff functions" by Pachella (1974) and Thomas (1974). The use of Ollman's (Note 1) terminology for the two computational procedures in the present paper is intended to be both a meaningful description of how each function is derived but also theoretically neutral with respect to alternative models of the tradeoff process.

4. This assumption is quite reasonable when the data from a given speed-emphasis condition represent the region of the speed-accuracy tradeoff in which accuracy is an increasing function of RT (i.e., the region between t_1 and t_2 in Figure 1). However, when performance *outside* the region between t_1 and t_2 is heavily represented in the data from a given speed-emphasis condition, the resulting mean RT may be significantly influenced by RTs outside the tradeoff region. Therefore, when generating an empirical SATF: (a) the manipulations of subjects' speed-accuracy criteria should be designed with a view toward minimizing the range of RTs in each speed-emphasis condition; and (b) speed-emphasis conditions with accuracies very near chance or perfect performance may be omitted from the process of fitting the data with linear (or alternative) equations. A very similar problem occurs in fitting empirical CAFs with linear equations. Lappin and Disch (1972a) suggest a method for minimizing the effect of data outside the tradeoff region for the CAF.

5. Ollman (Note 1) raises this possibility explicitly, and a number of other investigators have expressed similar concerns concerning the possibility of distortion in CAFs computed over trials with varying speed-accuracy criteria (e.g., Pachella, 1974, p. 79; Wickelgren, in press, pp. 15-17; and from a slightly different perspective by Thomas, 1974, pp. 447-449).

6. Two additional sources of evidence are also inconsistent with the invariance assumption. A recent experiment from this laboratory has replicated and extended the test of CAF invariance reported here. Within-condition CAFs very similar to those in Figure 2 were obtained using a band payoff instead of an RT-deadline design. Second, although their experiment was not explicitly addressed to the invariance assumption, Grice, Hunt, Kushner, and Morrow (1974) showed that functions analogous to the CAF from an "accuracy emphasis" condition differed substantially from those from a "speed emphasis" condition (cf. their Figures 11 and 12).

7. In order to be comparable to the empirical functions presented in Figure 3, the CAFs and SATFs derived from the computer-simulated data were also based on five speed-emphasis conditions with 100 trials per condition. For simplicity, normal distributions were used for both the correct and error response RT distributions in all comparisons. The means of the correct response distributions were held constant at 200, 225, 250, 275, and 300 msec, respectively. The means of the error response distributions were varied relative to these values to produce the variations in RT_c - RT_e.

(Received for publication July 7, 1975; revision received October 6, 1975.)