Categorization response time
with multidimensional stimuli

F. GREGORY ASHBY, GEOFFREY BOYNTON, and W. WILLIAM LEE
University of California, Santa Barbara, California

Categorization response time (RT) was examined in three separate experiments, in each of which exemplars varied on two physical dimensions. Three different types of stimuli were used: (1) horizontal and vertical line segments of varying length that were joined at an upper left corner, (2) rectangles of varying width and height, and (3) circles or semicircles of varying size with a radial arm of varying orientation. No evidence was found that stimulus familiarity or the category prototypes played any special role in determining categorization RT. Instead, RT decreased with distance from the stimulus to the categorization decision bound.

The categorization of objects is fundamental to human information processing and is required at some level by nearly every perceptual or cognitive task. Categorization has long been studied, and historically the dependent variable of primary interest has been response accuracy. A few studies have examined response time (RT; see, e.g., Hyman & Frost, 1975), but in most of these, the exemplars varied on only a single physical dimension (see, e.g., Bornstein & Monroe, 1980; Cartwright, 1941). In this article, we examine categorization RT when the exemplars vary on more than one physical dimension. Including RT with accuracy in the analysis of categorization experiments can only increase the observability of the categorization process as well as improve the testability of theories and models.

Few specific hypotheses can be found in the literature to explain how categorization RT changes with the physical characteristics of the stimuli. A notable exception is the *RT-distance* hypothesis, according to which RT decreases with the distance in psychological space from the stimulus representation to the decision bound that separates the exemplars of the contrasting categories (Ashby & Maddox, 1991, 1992b). The idea is that a stimulus that falls near a category bound has ambiguous category membership, hence categorization is slow, whereas a stimulus that is far from the category bound is easy to classify and therefore its RT is short.

In addition, a number of other hypotheses have been proposed implicitly. For example, an obvious prediction is that RT is faster to more familiar stimuli. Also, a natu-

ral prediction of prototype theory is that RT decreases with the similarity between the stimulus and the category prototype (see, e.g., Rips, Shoben, & Smith, 1973; Rosch, 1973a). This article reports RTs from three categorization experiments in which stimuli varied on two physical dimensions. Three different types of stimuli were used: (1) horizontal and vertical line segments of varying length that were joined at an upper left corner, (2) rectangles of varying width and height, and (3) circles or semicircles of varying size with a radial arm of varying orientation. In each experiment, the RT-distance hypothesis was supported. In addition, we found no evidence that stimulus familiarity or the category prototypes played any special role in determining categorization RT.

The RT-Distance Hypothesis

The RT-distance hypothesis is motivated by decision-bound theories of categorization, which assume that the perceptual effect of each presentation of a category exemplar can be represented as a point in a multidimensional perceptual space and that repeated presentations of the same exemplar do not always lead to the same perceptual effect (Ashby, 1992; Ashby & Gott, 1988; Ashby & Lee, 1991, 1992; Ashby & Maddox, 1990, 1992a). Decision-bound theory assumes that a practiced subject divides the perceptual space into regions and associates a category label with each region. On each trial, the subject categorizes an object by determining in which region the stimulus representation falls. The partition between two response regions is called the *decision bound*.

Decision-bound theory is a multivariate generalization of signal detection theory, with the decision bound playing the role of the signal detection response criterion. In signal detection theory, the most commonly made processing-time assumption is that RT decreases with the distance between the perceptual effect and the response criterion (Bindra, Donderi, & Nishisato, 1968; Bindra, Williams, & Wise, 1965; Emmerich, Gray, Watson, & Tanis, 1972; Smith, 1968). The RT-distance hypothesis states that RT

decreases with the distance between the perceptual effect and the decision bound.

Another major application of the RT-distance hypothesis is in the area of recognition memory, where it is called the *strength-latency* hypothesis. Strength theory assumes that each recognition memory decision is based on the strength of the memory trace of the probe item (Norman & Wickelgren, 1969; Wickelgren & Norman, 1966). Old items are assumed to have greater trace strength than do new items, and a response criterion partitions the strength dimension into *no* and *yes* response regions. The strength-latency hypothesis assumes that RT decreases symmetrically on either side of the *yes-no* criterion. In a comprehensive analysis, Murdock (1985) concluded that the strength-latency hypothesis is able to account for the major results of both the memory scanning task (Sternberg, 1966) and the study-test experiment (Murdock & Anderson, 1975).

The RT-distance hypothesis predicts certain testable relations between the RTs on correct and incorrect categorization trials. Under the conditions of Experiments 1 and 2, the RT-distance hypothesis predicts that median correct RT will be less than median incorrect RT (Thomas, 1971), because stimuli that are incorrectly categorized will tend to be closer to the decision bound than will stimuli that are correctly categorized.

RT Hypotheses Derived From Prototype Theory

Prototype theories of categorization assume that the exemplars of a category are distributed about a most typical member, called the *prototype* (Rosch, 1973a, 1973b; see also Homa, Sterling, & Trepel, 1981; Posner & Keele, 1968, 1970; Reed, 1972; Rosch, Simpson, & Miller, 1976). In its simplest form, prototype theory assumes that when asked to assign a stimulus to one of several categories, the subject responds with the category possessing the most similar prototype. Prototype models usually assume a multidimensional scaling (MDS) definition of similarity (see, e.g., Shin & Nosofsky, 1992). Specifically, they assume that stimuli can be represented as points in a multidimensional space and that the similarity of a pair of stimuli decreases with interpoint distance. Thus, given a certain stimulus, the most similar prototype is the *nearest* prototype.

Many studies have demonstrated that this simple prototype model provides an inadequate description of categorization accuracy (Ashby & Gott, 1988; Medin & Schaffer, 1978; Shin & Nosofsky, 1992). Even so, category prototypes are associated with a number of special properties, and it is possible that the prototypes could play some significant role in determining categorization RT. For example, it has been shown that the prototypes are the first exemplars learned by children (Mervis, 1980; Rosch, 1973a) and that they are likely to be named first when subjects are asked to recall all members of a category (Mervis, Catlin, & Rosch, 1976). In addition, if subjects are asked to verify whether a probe item is a member of a category, the fastest *yes* responses are to the category

prototypes (Rips et al., 1973; Rosch, 1973a). Given these results, a natural assumption is that RT decreases with the similarity of the stimulus to the prototype.

A popular measure of the similarity of stimulus i to j in models that use an MDS stimulus representation is

$$s_{ij} = exp(-cd_{ij}^{\alpha}), \qquad (1)$$

in which d_{ij} is the distance between the perceptual representations of the two stimuli (Nosofsky, 1986; Shepard, 1957). The parameter c is a measure of the overall discriminability between stimuli and is expected to increase with exposure duration and as the subject gains experience with the stimuli (Nosofsky, 1986). Finally, the parameter α defines the nature of the similarity function. The exponential similarity function occurs when $\alpha = 1$; the Gaussian function occurs when $\alpha = 2$. The most successful models have paired the exponential similarity function with either the city block or Euclidean distance metrics or have paired the Gaussian similarity function with the Euclidean metric (Ennis, 1988; Nosofsky, 1986; Shepard, 1987).

Consider a categorization task with only one category, say Category A, for which the subject is to decide whether each stimulus is a member of the category. We call this an $(A, not A)$ categorization. In such a task, the RT-prototype hypothesis assumes that on trials in which Stimulus X is presented, RT decreases with the similarity between X and the A prototype—that is, with s_{XA}. Now, if RT decreases with s_{XA}, it increases with any decreasing function of s_{XA}. In particular, Equation 1 implies that

$$d_{XA} = \left(-\frac{1}{c} \log s_{XA} \right)^{\frac{1}{\alpha}},$$

and thus that d_{XA} is a strictly decreasing function of s_{XA}. Therefore, an alternative statement of the RT-prototype hypothesis is that it assumes RT increases with the distance between the perceptual representations of the stimulus and the category prototype. Empirical testing of the RT-prototype hypothesis (either the distance or the similarity versions) is facilitated because its validity is unaffected by the numerical values of the α and c parameters of Equation 1. This is because d_{XA} is a decreasing function of s_{XA} for any positive values of α and c.

Next, consider a categorization task with two categories, A and B. The subject decides whether each stimulus is a member of Category A or B. We call this an (A, B) task. In this case, several RT hypotheses might be derived from prototype theory. One possibility is that RT increases with the distance between the representations of the stimulus and the *nearest* prototype. The *RT-nearest-prototype* hypothesis predicts that the two fastest responses should be to the two category prototypes. As one moves away from the prototypes, the stimuli become more atypical of the categories and so more difficult to categorize. Another possibility, however, is that the subject assigns a stimulus to the category containing the most similar prototype and, therefore, that RT decreases with the difference between the similarities of the stimulus to the Category A and B prototypes. This *RT-comparative-prototype* hypothesis

predicts that the fastest responses will be to the stimuli that are most discriminative with respect to the (A, B) categorization, even if these stimuli are atypical of both categories. In contrast to the RT–nearest-prototype hypothesis, the RT–comparative-prototype hypothesis predicts that, with respect to RT, the prototypes will have no favored status. In particular, they should have the same RT as do other stimuli that are associated with the same similarity difference.

The RT–comparative-prototype hypothesis can be formalized by assuming that, on trials in which the stimulus is X, RT decreases with the absolute value of the log similarity difference. Specifically, RT should decrease with

$$|\log s_{XA} - \log s_{XB}| = c|d_{XA}^{\alpha} - d_{XB}^{\alpha}|$$

or, equivalently, with the absolute difference between the prototype distances (DP) raised to the α power

$$DP = c^{-1}|\log s_{XA} - \log s_{XB}| = |d_{XA}^{\alpha} - d_{XB}^{\alpha}|. \quad (2)$$

The RT–comparative-prototype hypothesis predicts that the slowest responses will be to stimuli that are equally similar to the two prototypes—that is, to all stimuli X for which $s_{XA} = s_{XB}$, so that $DP = 0$. In addition, on one side of this isosimilarity contour, the subject always responds A (because $s_{XA} > s_{XB}$), whereas on the other side, the subject always responds B (because $s_{XA} < s_{XB}$). Thus, the isosimilarity contour plays the same role as a decision bound. In fact, Ashby and Maddox (1992b) showed that any model assuming the RT-comparative prototype hypothesis also satisfies the RT–distance hypothesis when α is greater than 1.0 and d is Euclidean distance.

In an $(A, \text{not } A)$ task, there is only one prototype, so $DP = |d_{XA}^{\alpha}| = d_{XA}^{\alpha}$. The latter equality holds because distances are never negative. Now, if RT increases with d_{XA}^{α}, it must also increase with d_{XA}. In other words, in an $(A, \text{not } A)$ task, the RT-comparative-prototype hypothesis predicts that RT increases with distance from the prototype. Thus, in this case, both RT-prototype hypotheses agree.

Familiarity Effects

Another factor that could possibly influence categorization RT is stimulus familiarity. One would expect RT to increase on trials when the subject is presented a stimulus that is very different from those he or she has already categorized. In fact, Nosofsky (1991) found that when subjects were trained to make a simple size categorization (large versus small), RT increased dramatically to the first presentation of a stimulus that was much larger than any of the training stimuli. Similarly, one might expect RT to be short on trials when the stimulus is familiar, perhaps because, in such cases, the correct response is well learned.

Although Nosofsky's results indicate some effect of familiarity, the overall contribution of familiarity to categorization RT is unknown. An extreme position assumes that there is some effect of familiarity on every trial. This position, which we call the *RT-familiarity* hypothesis, predicts that categorization RT decreases with stimulus

familiarity. Of course, to test this hypothesis one must explicitly define stimulus familiarity.

One convenient measure of stimulus familiarity can be derived from exemplar theory (see, e.g., Medin & Schaffer, 1978; Nosofsky, 1988). In exemplar theory, the familiarity of Stimulus X is defined as the sum of the similarities of X to every other stimulus in the ensemble—that is, to

$$\text{familiarity of } X = \sum_{i=1}^{n} s_{Xi}, \quad (3)$$

where s_{Xi} is the similarity between Stimuli X and i as given by Equation 1 and n is the number of stimuli in the ensemble. It is important to note that the sum is computed to all stimuli used in the experiment, without regard to their category membership.

Empirical Tests of the RT Hypotheses

A number of studies examined categorization RT in tasks in which the stimuli varied on only one relevant dimension (see, e.g., Bornstein & Monroe, 1980; Cartwright, 1941; Nosofsky, 1991). Bornstein and Monroe (1980) reported the results of an (A, B) categorization of Munsell color samples that equally supported the RT-distance hypothesis and both forms of the RT-prototype hypothesis. Cartwright (1941) reported the results of an $(A, \text{not } A)$ task in which the stimuli were two line segments radiating from a common point but varying in angle of separation. On each trial, subjects were shown a stimulus and were asked to determine whether its angle was within a specified range. Data from trials on which the subjects responded *yes* supported the RT-distance hypothesis and both forms of the RT-prototype hypothesis, but data from trials on which the subjects responded *no* supported only the RT-distance hypothesis.

In the case in which stimuli varied on more than one relevant dimension, the most important study of categorization RT was reported by Hyman and Frost (1975). They created two artificial categories by varying the height and width of an arbitrary configuration of nine dots. Their experiments were all of the (A, B) variety, and they specifically tested a number of RT hypotheses, including the RT-distance hypothesis and the RT-comparative-prototype hypothesis. All of the hypotheses that they tested received reasonable support, but they argued that, in one experiment, the RT-comparative-prototype hypothesis was favored, whereas the RT-distance hypothesis was favored in another experiment. There are several reasons, however, why the results of Hyman and Frost are of limited interest. First, rather than test ordinal versions of the RT hypotheses, they tested versions that assumed a specific functional form relating RT to the statistic of interest. For example, their version of the RT-distance hypothesis assumed that RT decreased exponentially with distance from the decision bound. Unfortunately, it is impossible to decide whether a failure of this model is due to a failure of the RT-distance hypothesis or whether the RT-distance hypothesis is valid but the function relating RT and

distance-to-bound is not exponential. For this article, the RT hypotheses were tested directly. No assumptions are made about the specific functional form relating RT to the statistics of interest (e.g., distance-to-bound). Second, Hyman and Frost arbitrarily selected one category bound for their test of the RT–distance hypothesis. Specifically, they chose one of an infinite number of bounds that perfectly separated the two categories. Third, because of the minor goodness-of-fit differences among the various RT hypotheses, they were unable to draw strong conclusions. Finally, they did not examine the role of stimulus familiarity on categorization RT.

EXPERIMENTS 1 AND 2

Experiments 1 and 2 used three different kinds of stimuli. Examples are shown in Figure 1. In Experiment 1, stimuli were vertical and horizontal line segments of varying length joined at an upper left corner. In Experiment 2, stimuli were either rectangles that varied in width and height or else semicircles of varying size that contained a radial line of varying orientation. The components of rectangles are thought to be perceptually integral (Burns & Hopkins, 1987; Dunn, 1983; Felfoldy, 1974), whereas the components of the circular stimuli are thought to be perceptually separable (Burns, Shepp, McDonough, & Weiner-Ehrlich, 1978; Garner & Felfoldy, 1970; Hyman
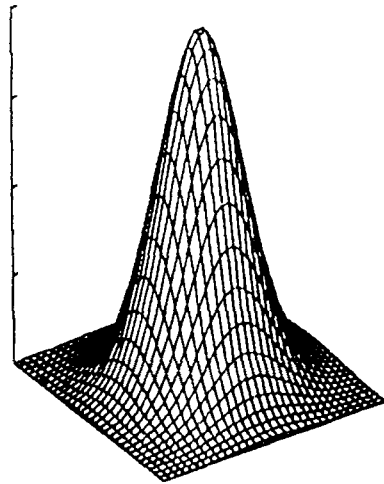
Figure 2. A bivariate normal distribution.

& Well, 1967; Shepard, 1964; however, see Ashby & Lee, 1991; Ashby & Maddox, 1990).

The most powerful tests of the RT hypotheses require an examination of RT throughout the perceptual space. Such analyses are possible only in experiments in which the stimuli vary continuously along each dimension. For this reason, all experiments reported in this article involved categories in which the exemplars were normally distributed around a prototypical value on each stimulus dimension (Ashby & Gott, 1988; Ashby & Maddox, 1990, 1992a). With two dimensions, each category is specified by a bivariate normal distribution, such as the one shown in Figure 2. Rather than draw a three-dimensional figure, it is conventional to depict a bivariate normal distribution by a contour of equal likelihood, which is created by taking a slice parallel to the stimulus plane at some arbitrary height and looking down at the result from above. The resulting figure is two-dimensional. With normal distributions, these contours are always circles or ellipses.

Bivariate normal distributions are specified by three kinds of parameters: (1) location (i.e., a mean on each dimension), (2) spread (i.e., a variance on each dimension), and (3) association (i.e., a covariance or correlation between dimensions). The contours of equal likelihood convey information about each of these parameters. Each contour is centered at the distribution mean. The ratio of the x to y standard deviations equals the ratio of the horizontal to vertical width of the contour, and the correlation coefficient determines the slope of the major axis.

In the experiments reported below, each category is defined by a specific bivariate normal distribution. Consider an (A, B) categorization task in which the exemplars of both categories are circular stimuli such as those shown in Figure 1. On trials when an exemplar from Category A is to be presented, a random sample (x, y) from the A
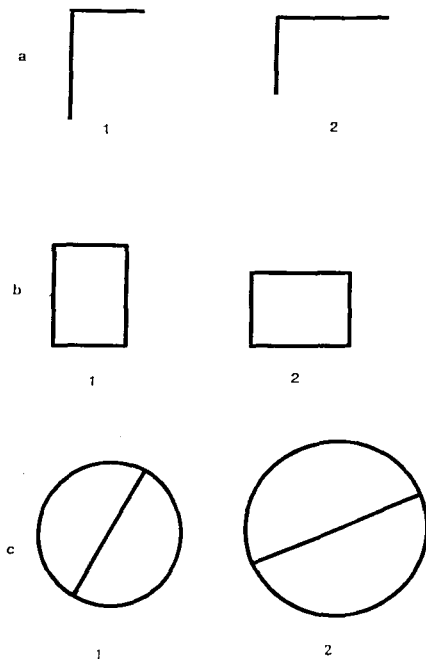

Figure 1. Examples of the stimuli used in the experiments reported in this article. The full circles shown in Figure 1C were used in Experiment 3, whereas the top halves were used as the semicircles in Experiment 2.
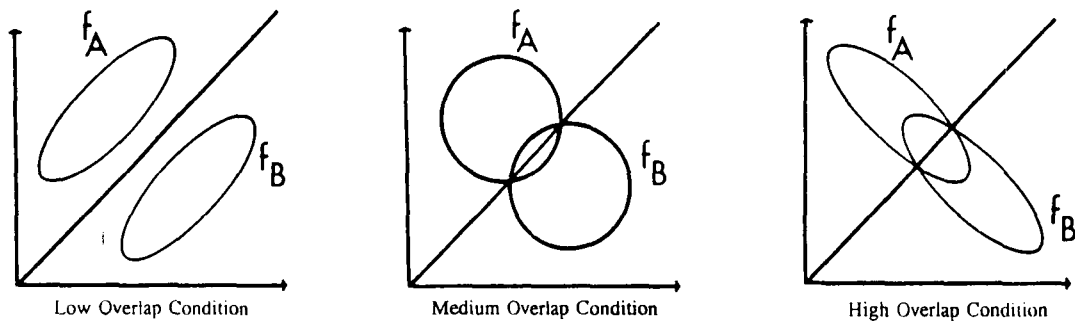
Figure 3. Contours of equal likelihood from the three conditions of Experiment 1.

distribution is used to construct a circle with size (i.e., diameter) x and orientation y. This stimulus is shown to the subject using high-contrast, noise-free, response-terminated displays. The subject is instructed to respond with the name of the category of which the exemplar is most likely a member. Feedback is given on each trial. Because the distributions overlap, perfect performance is impossible. Specifically, note that every stimulus has a nonzero likelihood of having been drawn from either category distribution. See Ashby (1992) for a more thorough discussion of categories as normal distributions.

In all experiments reported below, each prototype (i.e., the mean, median, and mode of each distribution) was displayed alternately with its category label five times, at the beginning of every session. This was done in an attempt to bias subjects toward prototype decision rules and category representations. A failure to find prototype effects on RT under these biased conditions would represent strong evidence against prototype theory.

The contours of equal likelihood used in the three conditions of Experiment 1 are displayed in Figure 3. Note that the category means are identical in all three conditions. In the high-overlap condition, the exemplars in each category have values on the two dimensions that are negatively correlated. In the medium-overlap condition, the dimensional values are uncorrelated; in the low-overlap condition, they are positively correlated. In all three conditions, the optimal classifier—that is, the device that maximizes overall categorization accuracy—uses the decision bound, $y = x$. Response A is given to any exemplar falling above this bound, and Response B is given to any exemplar falling below. The optimal classifier would correctly categorize 65% of the stimuli in the high-overlap condition, 80% of the stimuli in the medium-overlap condition, and 95% in the low-overlap condition. Note that this design varies overall accuracy but keeps total variation within each category approximately constant.

The stimuli used in Experiment 1 were the line figures of Figure 1A. With these stimuli, the decision rule,

Respond A if $y - x > 0$; otherwise respond B

is equivalent to the rule

Respond A if the stimulus is taller
than it is wide; otherwise respond B.

Ashby and Gott (1988) showed that subjects learn such a rule quickly and spontaneously and that they apply it consistently and in a nearly optimal fashion. Each subject in Experiment 1 completed one experimental session comprising 400 categorization trials. Feedback was given on each trial.

Experiment 2 used contours of equal likelihood similar to those of the low-overlap condition of Experiment 1 (see Table 1 for the exact parameter values). In both cases, the categories were characterized by a positive correlation, and the optimal decision bound was the line $y = x$. In Experiment 2, however, the optimal classifier would correctly classify 80% of the stimuli. There were two important differences between Experiment 2 and the low-overlap condition of Experiment 1. First, the two experiments used different stimuli. Half the subjects in Experiment 2 used the rectangular stimuli, and half used the semicircular stimuli. This allows a test of whether the results of Experiment 1 generalize to other stimulus components. Second, each subject in Experiment 2 completed three experimental sessions, whereas each subject in Experiment 1 completed only one session. This difference allows us to examine the effects of experience on categorization RT. In all other respects, the two experiments were identical.

General Method

Subjects

All subjects in the experiments were volunteers from the UCSB community who were paid at a base rate of $5 for each 1-h experimental session. All subjects had either 20/20 vision or vision corrected to 20/20. Twelve subjects participated in Experiment 1, and 4 subjects participated in Experiment 2. No subject participated in more than one experiment. Four subjects participated in each of the three conditions of Experiment 1 (low-, medium-, and high-overlap, all with the line stimuli). Each subject completed one session of 400 trials. Two subjects participated in each of the two conditions of Experiment 2 (medium-overlap rectangles and semi-circles), and each subject completed three sessions of 400 trials.

Stimuli

Figure 1 presents examples of the line and rectangle stimuli. The semicircles of Experiment 2 were identical to the top half of the Figure 2 circles. In both experiments, the stimuli were computer generated and displayed on a Mitsubishi Electric Color Display Monitor Model No. C-9918NB in a dimly lit room. Two categories,

Table 1
Parameter Values for the Categories of Experiments 1 and 2

| | Experiment 1 | | | |
Parameter	Low Overlap	Medium Overlap	High Overlap	Experiment 2
		Category A		
μ_x	200	200	200	92
μ_y	250	250	250	128
σ_x	67	42	67	42
σ_y	67	42	67	42
ρ_{xy}	.896	0	$-.896$	.492
		Category B		
μ_x	250	250	250	128
μ_y	200	200	200	92
σ_x	67	42	67	42
σ_y	67	42	67	42
ρ_{xy}	.896	0	$-.896$	.492

A and B, were devised by defining a pair of bivariate normal distributions.

The contours of equal likelihood for each of the three conditions of Experiment 1 are shown in Figure 3. Table 1 gives the exact parameter values describing the populations in each condition of both experiments. The size dimensions are given in pixel units based on a screen with a resolution of $1,024 \times 768$. There were approximately 27.8 pixels/cm, and the subjects sat approximately 140 cm from the screen. In Experiment 1, the average visual angle of all stimuli was 3.3°, with a standard deviation of 0.8°. In Experiment 2, the average visual angle was 1.6°, with a standard deviation of 0.6°. The orientation units of the semicircular stimuli were arbitrarily defined so that $1,000/\pi$ units was equal to 1.0 radian. Thus, a mean on the orientation dimension of 250 is equivalent to $\pi/4$ radians (i.e., 250 divided by $1,000/\pi$), so the bound $y = x$ equates a size of 250 pixels (about a quarter of the screen width) with a half of a semicircle ($\pi/4$ radians). To use the bound $y = x$, the subject must somehow compare size and orientation units. As Ashby and Maddox (1992a, p. 57) explain, no special problems with the data analysis arise if the subject equates 250 pixels with some orientation other than $\pi/4$ radians.

On each trial, stimulus generation proceeded as follows. First, a category (A or B) was randomly selected (each with probability 0.5). Next, a random sample (x, y) was drawn from the appropriate bivariate normal distribution. A stimulus was then constructed with either width (for the line figures and the rectangles) or radius (for the semicircles) equal to x and height or orientation equal to y.

Procedure

On every trial, the subject's task was to categorize the stimulus as an exemplar of category A or B by pressing an appropriate button. The subjects were told that RT was being recorded and that they should respond as quickly as possible without sacrificing accuracy. The stimulus display was terminated either by the subject's response or after 5 sec without a response. Feedback showing the correct response was displayed for 2 sec immediately following the subject's response. The next trial began 1 sec after the feedback disappeared.

At the beginning of each session, each prototype (i.e., the mean of each distribution) was displayed alternately with its category label five times. The first 100 trials of each session served as practice. A pause separated the practice trials from the 300 experimental trials. During the pause, the subject was allowed to ask questions about the procedure. The experimental session consisted of four 75-block trials. There was a 30-sec pause between blocks to allow subjects to rest.

Results and Discussion

Accuracy Analysis

Tables 2 and 3 describe characteristics of the responses given by the subjects in each condition. Column 3 gives each subject's percent correct and column 4 gives the percent correct to be expected from a subject who used the optimal bound perfectly. These percentages differ somewhat from the population values (i.e., 95%, 80%, and 65% for the low-, medium-, and high-overlap conditions, respectively) because of sampling variability. Table 2 indicates that all but one subject in Experiment 1 (Subject 4 in the high-overlap condition) performed at close to this optimal level. Table 3 indicates that, by Day 3, both subjects in the rectangle condition were responding nearly optimally but that both of the circle condition subjects responded suboptimally throughout the experiment.

For each subject, we also determined the linear bound that best separated the A and B responses, which we call the *discriminant* bound. An A response is correctly predicted if it is above the bound; a B response is correctly predicted if it is below the bound. The discriminant bound maximizes the number of correctly predicted responses. Tables 2 and 3 also describe the discriminant bounds, along with the percentage of responses accounted for by the discriminant and optimal bounds. For all subjects in Experiment 1, the number of responses correctly predicted by the discriminant bound is less than 4% greater than the number correctly predicted by the optimal bound, and for 10 of the 12 subjects, the difference is less than 2%. In Experiment 2, some of these differences are larger, suggesting that some subjects may have used a suboptimal bound. Even so, in both experiments, the discriminant bound slopes are all within .4 of the optimal slope. In addition, a number of the intercepts are large, suggesting the possibility of response biases.

In Experiment 1, and for the rectangular stimuli in Experiment 2, the discriminant bound correctly predicted at least 89% of the responses in 17 of the 18 cases. The only exception was Subject 4 in the high-overlap condition of

Table 2
Characteristics of the Responses Given in Experiment 1 (Line Stimuli)

Condition	Subject	% Correct	% Correct by Optimal Bound	% of Responses Accounted for by Optimal Bound	Discriminant Bound	% of Responses Accounted for by Discriminant Bound
Low	1	92.6	94.6	93.6	$y = x-23$	96.0
Overlap	2	94.3	94.3	94.0	$y = x+4.6$	95.0
	3	92.6	95.3	94.0	$y = x-23.8$	96.0
	4	93.7	96.3	96.0	$y = x-4.1$	97.0
Medium	1	82.7	83.3	91.3	$y = 1.3x-79.7$	94.7
Overlap	2	78.6	81.9	94.3	$y = x+2$	94.6
	3	78.3	82.3	90.7	$y = x-10$	92.0
	4	82.3	85.0	94.0	$y = x-11.8$	94.7
High	1	63.2	63.9	91.0	$y = 1.4x-78.2$	92.0
Overlap	2	59.2	62.2	88.0	$y = .7x+65$	90.0
	3	62.3	63.0	93.3	$y = x-10.3$	95.0
	4	61.3	67.0	77.7	$y = .8x+55$	78.3

Table 3
Characteristics of the Reponses Given in Experiment 2 (Medium Overlap)

Condition	Subject	Day	% Correct	% Correct by Optimal Bound	% of Responses Accounted for by Optimal Bound	Discriminant Bound	% of Responses Accounted for by Discriminant Bound
Rectangle	1	1	69.6	75.6	77.6	$y = 1.3x-3.3$	91.0
		2	74.7	79.3	85.3	$y = 1.2x+3.7$	89.3
		3	79.5	80.0	90.9	$y = 1.1x+6.6$	95.6
	2	1	76.8	79.1	85.2	$y = x-12.9$	90.9
		2	76.3	76.3	88.6	$y = .9x-6.9$	92.3
		3	81.0	82.7	92.3	$y = 1.1x-16.4$	94.3
Circle	1	1	73.6	78.6	76.8	$y = .8x+28.4$	80.6
		2	75.3	81.7	84.7	$y = 1.1x+3.4$	85.7
		3	73.7	82.6	87.0	$y = .9x+18.4$	89.3
	2	1	48.1	78.8	50.5	$y = .9x-3.9$	51.8
		2	68.0	81.3	73.8	$y = x+30.5$	81.6
		3	74.0	83.0	81.3	$y = .6x+50.4$	87.7

Experiment 1. For this subject, both the discriminant and optimal bounds incorrectly predicted many responses. This indicates that this subject either began guessing frequently or else gave up completely on the $y = x$ decision rule and adopted some other response strategy.

For the semicircular stimuli in Experiment 2, the discriminant bound less successfully predicted responding. Even so, in both conditions (rectangles, semicircles), the percentage of responses correctly predicted increased with the subject's level of experience. In fact, by the third experimental session, the performance of the discriminant bound was almost as good as that in the other stimulus conditions. See Ashby and Gott (1988) and Ashby and Maddox (1990, 1992a) for a more thorough discussion of accuracy analyses in experiments with normally distributed categories.

Response-Time Analysis

One important difficulty in testing the various RT hypotheses is that they all make predictions about the rela-tion of RT to the position of stimuli within the perceptual space—not about the position of stimuli within the stimulus space. Unfortunately, of course, the perceptual space is unobservable. Nevertheless, all is not lost. First, the Stevens exponent for length is very close to 1.0 (Stevens, 1961), so the perceived size dimensions should approximately equal the physical size dimensions (although see Krueger, 1989, on the dangers of interpreting magnitude estimates of line length as direct measures of subjective magnitude). Second, psychological scaling solutions that have been derived for the semicircular stimuli have shown that the perceptual space is very similar to the stimulus space (Nosofsky, 1986; Nosofsky, Clark, & Shin, 1989; Shepard, 1964). Third, the experiments use three different kinds of stimuli. If the results generalize across the different stimulus sets, concerns about some unexpected transformations between the stimulus and perceptual spaces are reduced. Fourth, virtually all of our RT analyses assume only an ordinal relation between RT and position of the stimulus in the stimulus space. As such, they

remain valid under a large class of monotonic transformations. Thus, these analyses do not require a close correspondence between the stimulus and perceptual spaces.

RT–distance hypothesis. The first RT analyses are shown in Tables 4 and 5. Median RTs for each condition are shown separately for both correct and incorrect responses. As described above, the RT–distance hypothesis predicts that in these experiments, incorrect RTs should be slower than correct RTs and that this difference should decrease as distributional overlap increases. In brief, this prediction is made because, when distributional overlap is small, most incorrectly categorized exemplars will be close to the decision bound and therefore will be associated with long RTs. In the high-overlap condition of Experiment 1, however, some incorrectly categorized exemplars will be far from the decision bound and therefore should be associated with short RTs. Table 4 indicates that these predictions are supported for every subject in Experiment 1. In Experiment 2, the correct median RTs are faster in 11 of the 12 cases (see Table 5). The only exception occurs during the first experimental session of Subject 2 in the rectangle condition.

An ideal test of the various RT hypotheses would be to fit the models to estimates of the mean RT to every stimulus and then to compare the resulting goodness-of-fit values. Unfortunately, this cannot be done without adding many extra assumptions. This is because each of the RT hypotheses makes predictions only at the ordinal level, whereas fitting the mean RTs requires predictions at the ratio scale level. For example, the RT–distance hypothesis predicts that RT decreases as one moves away from the decision bound, but it says nothing about the functional form that this decrease takes. One could fit a model to the mean RT surfaces that assumes RT decreases in an exponential fashion as one moves away from the decision bound, but a poor fit of this model is not necessarily evidence against the RT–distance hypothesis. For example, the RT–distance hypothesis may be correct, but RT may fall off as a power function of distance-to-bound rather than as an exponential function.

Table 4
Median RTs in Experiment 1 (Line Stimuli)

Condition	Subject	Correct Responses	Incorrect Responses
Low	1	872	1,528
Overlap	2	629	1,088
	3	871	1,537
	4	591	978
Medium	1	678	911
Overlap	2	671	996
	3	540	638
	4	449	486
High	1	805	818
Overlap	2	853	904
	3	499	519
	4	386	417

Table 5
Median RTs in Experiment 2 (Medium Overlap)

Condition	Subject	Day	Correct Responses	Incorrect Responses
Rectangle	1	1	795	824
		2	672	714
		3	642	725
	2	1	710	688
		2	457	503
		3	464	520
Circle	1	1	878	1,014
		2	1,025	1,358
		3	884	1,079
	2	1	1,109	1,412
		2	2,157	2,229
		3	1,151	1,725

Fortunately, it is possible to test the RT hypotheses directly by computing various rank-order correlations. For example, to test the RT–distance hypothesis, one rank orders the subject's responses by RT, then rank orders them by distance-to-bound, and finally computes the correlation between these two rankings. If the RT–distance hypothesis is correct, the resulting correlation should be negative, no matter what the function relating RT to distance-from-bound. The RTs from all responses (both correct and incorrect) should be included in these computations, since trials on which the subject responds with the incorrect category are not necessarily errors in the traditional sense. For example, because the category distributions overlap, it is possible that a subject processes the stimulus information optimally, makes the same response as the optimal classifier, but is incorrect. Some errors might be the result of mistakes in processing, but we expect most errors are due to the overlapping nature of the category distributions. Therefore, all RT correlations reported in this article are based on all responses made by the subject.

Tables 6 and 7 list some (Spearman) rank-order correlation coefficients relevant to the RT–distance and RT–prototype hypotheses. The first set of correlations are between the RT on each trial and the (Euclidean) distance between the stimulus that was presented and the optimal decision bound (i.e., the bound $y = x$). The second set of correlations is identical except that instead of computing distance to the optimal bound, distance was computed to the discriminant bound (given in Tables 2 and 3). The RT–distance hypothesis predicts that these correlations should be negative.

In Experiment 1, the RT–distance hypothesis is strongly supported for every subject,[1] except Subject 4 in the high-overlap condition. In general, it makes little difference whether distance is computed to the optimal or to the discriminant bound, probably because the discriminant bounds were not much different from the optimal bounds (i.e., see Table 2). The only exceptions are for Subjects

Table 6
Spearman Rank-Order Correlations for Experiment 1 (Line Stimuli)

Condition	Subject	RT Distance (Optimal Bound)	RT Distance (Discriminant Bound)	RT Prototype
Low Overlap	1	−0.432*	−0.538*	0.078
	2	−0.500*	−0.507*	0.134†
	3	−0.434*	−0.562*	0.046
	4	−0.454*	−0.436*	0.127†
Medium Overlap	1	−0.661*	−0.689*	−0.049
	2	−0.542*	−0.544*	−0.021
	3	−0.435*	−0.415*	−0.257
	4	−0.399*	−0.379*	−0.072
High Overlap	1	−0.421*	−0.399*	−0.357
	2	−0.339*	−0.329*	−0.231
	3	−0.241*	−0.256*	−0.124
	4	−0.047	−0.037	0.015

*$p < .001$.   †$.001 < p < .05$.

Table 7
Spearman Rank-Order Correlations for Experiment 2 (Medium Overlap)

Condition	Subject	Day	RT Distance (Optimal Bound)	RT Distance (Discriminant Bound)	RT Prototype
Rectangle	1	1	0.152	−0.006	−0.020
		2	−0.115†	−0.270*	0.162†
		3	−0.311*	−0.504*	0.055
	2	1	−0.101	−0.147†	0.011
		2	−0.308*	−0.206*	−0.080
		3	−0.316*	−0.352*	0.114
Circle	1	1	−0.099	−0.185†	0.046
		2	−0.369*	−0.409*	0.015
		3	−0.306*	−0.340*	−0.010
	2	1	0.029	0.087	−0.088
		2	−0.211*	−0.172†	−0.034
		3	−0.268*	−0.448*	0.030

*$p < .001$.   †$.001 < p < .05$.

1 and 3 in the low-overlap condition, in which the correlations computed from the discriminant bound are substantially more negative than are the correlations computed from the optimal bound. Because decision-bound theory predicts that the discriminant bound should be better than the optimal bound as an estimate of the subject's true decision bound, this result supports the RT–distance hypothesis.

In Experiment 2, there is a substantial effect of practice. In every case the RT–distance correlations are more extreme on Day 3 than on Day 1. In fact, by Day 3, for every subject, the RT–distance hypothesis is supported, and the correlations computed from the discriminant bound are more extreme than those computed from the optimal bound. On Day 1, however, the optimal-bound version of the RT–distance hypothesis is not supported by the data from any subjects. The discriminant-bound version receives marginal support from two subjects. With difficult categorization problems, a certain amount of instability in performance is expected while the subject is learning the appropriate decision bound. During this pe-

riod, the RT–distance hypothesis may be valid, but because of the changing nature of the decision bound, correlations of the type listed in Table 7 will be unimpressive. Recall, however, that, in Experiment 1, each subject only completed one experimental session, and yet the data of 11 of 12 subjects supported the RT–distance hypothesis. Thus, this hypothesis indicates that the Experiment 1 categorization task must have been easier to learn than was the Experiment 2 task. Recall that the optimal bound was the same in the two experiments but that the stimuli were different. Experiment 1 used the line stimuli, and Experiment 2 used rectangles and semicircles. The dimensions of the semicircles are in different units (i.e., orientation and size), therefore it makes sense that a task that requires integrating information from such dimensions would be more difficult than the same task with the line or rectangle stimuli (see Ashby & Maddox, 1990). It is not so obvious, perhaps, why the rectangle stimuli should be more difficult than the line stimuli. There is substantial evidence, though, that the dimensions of rectangles are integral (Burns & Hopkins, 1987; Dunn, 1983; Felfoldy,

1974), whereas the dimensions of the line stimuli are separable (Ashby & Gott, 1988; Ashby & Perrin, 1988; Townsend, Hu, & Ashby, 1981). In support of this possibility, note that all 4 subjects in Experiment 2 performed more poorly during Day 1 than did any of the medium-overlap subjects in Experiment 1 (see Tables 2 and 3), although in both cases optimal accuracy was 80%.

RT–comparative-prototype hypothesis. If the category prototypes are interpreted as the category means, medians, or modes, then in Experiments 1 and 2, the RT-comparative-prototype hypothesis is identical to the version of the RT-distance hypothesis that assumes the subject uses the optimal bound. A more general version of the prototype model allows the location of the category prototypes to be free parameters. In this case, the RT-comparative-prototype hypothesis is identical to the discriminant-bound version of the RT-distance hypothesis (Ashby, 1992; Ashby & Maddox, 1993). Thus, the design of Experiments 1 and 2 make it impossible to test between the RT-distance and RT-comparative-prototype hypotheses. Experiment 3 was designed specifically to make such a comparison possible.

RT–nearest-prototype hypothesis. The last column in Tables 6 and 7 lists correlations between each RT and the distance to the nearest prototype, with the prototypes defined as the means of the category distributions (or equivalently as the medians or modes). The RT-nearest-prototype hypothesis predicts that this correlation should be positive. Only 3 of the 24 correlations listed in Tables 6 and 7 marginally support this prediction. In fact, more than half of the correlations in Experiment 1 are negative. This result causes extreme difficulty for the RT-nearest-prototype hypothesis.

Although these results rule out the possibility that RT is determined by similarity to the nearest prototype, they do not rule out the possibility that the fastest responses were to the category prototypes—that is, they do not rule out all possible prototype effects. To test for this weaker type of prototype effect, we computed the RT rank of the stimulus most similar to each category prototype (the distribution means were not necessarily presented), which resulted in two numbers for each experimental session— the RT rank of the stimulus most similar to the Category A prototype and the RT rank of the stimulus most similar to the Category B prototype. During each session, 300 stimuli were presented, therefore a stimulus selected at random had an expected RT ranking of 150.5. In Experiment 1, the mean ranking of the stimulus most similar to the prototype was 142.7. In Experiment 2, the mean was 140.9. Since an RT ranking of 1 was assigned to the slowest response, these results indicate that RT to the prototypes was somewhat *greater* than would be expected from a stimulus drawn at random.

These results indicate that the prototypes have no special role within the entire set of stimuli. An even weaker type of effect would occur if the prototypes at least were

favored among all stimuli that were as equally discriminative as the prototype with respect to the A versus B categorization. To test this possibility, for each prototype in every data set, we computed the distance to the discriminant bound. We then found the 50 stimuli whose distance-to-bound was most similar to the prototype distance-to-bound. This procedure selects a set of stimuli that all fall in a swath that covers the prototype and is parallel to the decision bound. Within each of these sets, we rank ordered the stimuli by RT. A stimulus selected at random from one of these sets had an expected mean RT rank of 25.5. In Experiment 1, the prototypes had a mean RT ranking of 28; in Experiment 2, the mean rank of the prototypes was 23.8. Thus, even within the limited set of stimuli that are approximately equally discriminative with respect to the A versus B categorization, the prototypes have no favored status. In short, these results fail to find *any* special role for the prototypes in predicting categorization RT.

The RT–familiarity hypothesis. To investigate the role of stimulus familiarity on categorization RT, we computed the rank-order correlations between each RT and the familiarity score associated with each stimulus, with familiarity defined by Equation 3. The RT-familiarity hypothesis predicts that these correlations should be negative. Note that the validity of the RT-familiarity hypothesis may depend on the numerical values of the c and α parameters of Equation 1, and also on the nature of the distance metric. Because of this, we did separate analyses for the following three models: (1) Model EC assumed an exponential similarity function ($\alpha = 1$) and city-block distance, (2) Model EE assumed an exponential similarity function and Euclidean distance, and (3) Model GE assumed a Gaussian similarity function ($\alpha = 2$) and Euclidean distance. For each model, we estimated c directly from the RT data. Specifically, by using an iterative search procedure, we determined the value of c that minimized the rank-order correlation between RT and stimulus familiarity (i.e., made the correlation most negative). The performance of the three models was very similar. Nevertheless, Model GE consistently outperformed the other two models, so we restrict our attention to this version of the RT-familiarity hypothesis.

Tables 8 and 9 list the estimated values of c for Model GE, along with the resulting correlations between RT and familiarity. For comparison purposes, the correlations with distance to the discriminant bound are also listed. None of the correlations with familiarity were significantly less than zero. In fact, in both experiments, most correlations were positive. These data argue strongly against the validity of the RT-familiarity hypothesis.

In each condition of Experiment 1, the stimuli with the lowest familiarity scores came from the tails of the category distributions. In the medium- and high-overlap conditions, the category distributions are oriented so that the least familiar stimuli are far from the decision bound. In

Table 8
Results of Testing for Familiarity Effects on RT in Experiment 1 (Line Stimuli)

Condition	Subject	c	RT–Familiarity Correlation	RT–Distance Correlation (Discriminant Bound)	Distance–Familiarity Regression Model	Regression Weight on Familiarity
Low	1	0.043	−0.086	−0.538*	−0.544*	0.049
Overlap	2	0.268	−0.057	−0.507*	−0.512*	0.131
	3	0.116	−0.014	−0.562*	−0.562*	0.008
	4	0.047	−0.061	−0.436*	−0.444*	0.074
Medium	1	0.064	0.109	−0.689*	−0.690*	0.006
Overlap	2	35.76	0.078	−0.544*	−0.544*	0
	3	2.94	0.122	−0.415*	−0.415*	0
	4	0.005	0.090	−0.379*	−0.382*	0.060
High	1	5.48	0.160	−0.399*	−0.411*	0.114
Overlap	2	14.91	0.051	−0.329*	−0.330*	1.130
	3	5.48	0.046	−0.256*	−0.261*	0.417
	4	0.014	−0.062	−0.037	−0.054	7.375

$*p < .001.$

Table 9
Results of Testing for Familiarity Effects on RT in Experiment 2 (Medium Overlap)

Condition	Subject	Day	c	RT–Familiarity Correlation	RT–Distance Correlation (Discriminant Bound)	Distance–Familiarity Regression Model	Regression Weight on Familiarity
Rectangle	1	1	0.014	−0.056	−0.006	−0.056	∞
		2	0.002	−0.112	−0.270*	−0.276*	0.031
		3	14.905	0.007	−0.504*	−0.505*	0.006
	2	1	0.002	−0.038	−0.147†	−0.186†	0.374
		2	3.769	0.127	−0.206*	−0.206*	0
		3	0.005	−0.036	−0.352*	−0.354*	0.048
Circle	1	1	0.004	0.120	−0.185†	−0.185†	0
		2	0	0.134	−0.409*	−0.409*	0
		3	1.080	0.097	−0.340*	−0.340*	0
	2	1	0.005	−0.072	0.087	−0.076	2.20
		2	0.008	0.085	−0.172†	−0.172†	0
		3	45.91	0.097	−0.448*	−0.450*	0

$*p < .001.$ $†.001 < p < .05.$

this case, the RT–distance and RT–familiarity hypotheses are in opposition. The RT–distance hypothesis predicts that the least familiar stimuli will elicit the quickest responses, since they are furthest from the bound. Thus, the RT–distance hypothesis predicts that correlations between RT and familiarity should be positive in these two conditions. Table 8 indicates that seven of the eight RT–familiarity correlations in the medium- and high-overlap conditions are positive. In contrast, in the low-overlap condition, the tails of the category distributions are oriented so that the least familiar stimuli are close to the decision bound (i.e., see Figure 3). Thus, even in the absence of familiarity effects, if the RT–distance hypothesis holds, there will be a tendency for the correlations with familiarity to be negative. Table 8 indicates that all four RT–familiarity correlations are negative in the low-overlap condition. Thus, the results in Table 8 are con-

sistent with the hypothesis that the RT–distance hypothesis holds and that there is no effect of stimulus familiarity.

The RT–familiarity hypothesis predicts that stimulus familiarity is the primary determinant of categorization RT on every trial. Another, less ambitious view, which is suggested by Nosofsky's (1991) results, is that familiarity influences RT only for the least familiar stimuli. To test this hypothesis, we computed the familiarity rankings of the stimuli associated with the 10 slowest responses. As mentioned above, each experimental session consisted of 300 trials, so one stimulus selected at random had an expected ranking of 150.5. In Experiment 1, the stimulus associated with the slowest response had a mean familiarity ranking of 166 (averaged over subjects and conditions). The 10 slowest responses had a mean ranking of 120.4. Thus, the stimuli associated with the 10 slowest responses in Experiment 1 were somewhat more familiar

than were 10 stimuli selected randomly. In Experiment 2, the mean familiarity ranking of the stimuli associated with the 10 slowest responses was 155.4, about the same as one would expect if the stimuli had been chosen randomly. If the data of each day are examined separately, the picture looks only slightly better for this modified version of the RT-familiarity hypothesis. The mean familiarity rankings of the stimuli associated with the 10 slowest responses were 185.2 on Day 1, 137.9 on Day 2, and 143.0 on Day 3. Thus, on Day 1, the stimuli associated with the 10 slowest responses are somewhat *less* familiar than would be expected by chance. Even so, the effect is slight. The slowest response on Day 1 was to a stimulus with a mean familiarity ranking of 204.5, meaning that there were almost 100 stimuli that were less familiar but for which RT was nevertheless faster. Also note that the data from Experiment 1 did not show this marginal effect of familiarity, even though the subjects in Experiment 1 completed only one experimental session.

So far, the results indicate that familiarity is not the primary determinant of categorization RT in these experiments, even for the least familiar stimuli. Another possibility, however, is that familiarity has some small effect on every trial but that the primary determinant is some other principle, such as the RT-distance hypothesis. We tested this hypothesis with an ordinal version of multiple regression. Specifically, on each trial, we computed the weighted sum of distance-to-bound and familiarity (DF),

$$DF = \text{(distance to discriminant bound)}$$

$$+ \ (W \times \text{familiarity}),$$

where W is a nonnegative constant. We then computed the rank-order correlation of the statistic DF with RT. An iterative search procedure was used to find the value of W that minimized the rank-order correlation coefficient. To make the magnitude of the regression weight more meaningful, the distances and familiarities were first standardized to have the same mean and variance. Thus, a regression weight of $W = 1$ means that both distance and familiarity contributed equally to the RT prediction.

The rank-order correlations predicted by this regression model are given in Tables 8 and 9, along with the best-fitting value of the regression weight W (i.e., the value that made the correlation most negative). The regression model must always perform at least as well as the discriminant-bound version of the RT-distance hypothesis, because when $W = 0$, the two models are identical. In Experiment 1, the correlations predicted by the regression model are less than .02 more extreme than the correlations predicted by the RT-distance hypothesis in all 12 cases and are less than .01 more extreme in 10 of the 12 cases. Adding a familiarity component (and a free parameter) to the RT-distance hypothesis increases the average percent of variance accounted for by less than 1%. In Experiment 2, the regression model improves the rank-order correlation of the RT-distance hypothesis by

no more than .006 in 9 of 12 cases. The three exceptions are all from Day 1. In these three cases, however, the regression model accounts for only 1.45% of the variance in the data. Finally, note that in 8 of 24 cases in the two experiments, the value of the regression weight that maximized percent of variance accounted for was zero, indicating that the RT-distance hypothesis accounted for the data as well as did any form of the regression model.

The only evidence we obtained that indicated *any* effect of stimulus familiarity on categorization RT appeared in plots that estimated mean RT as a function of the stimulus coordinates. These plots were generated using the distance-weighted least squares procedure (McLain, 1974) provided by the statistical package, SYSTAT (Wilkinson, 1990). In general, the plots confirmed all of our other analyses. Specifically, they all showed a dramatic ridge hovering approximately over the discriminant bound, with RT falling off symmetrically on either side. In addition, there were no obvious depressions at the prototype coordinates. In some plots, however, there was a pronounced saddle in the ridge lying above the decision bound. A strict interpretation of the RT-distance hypothesis requires this ridge to be flat. There are several possible explanations. First, the saddles may be due to random variation in performance and have no real psychological meaning. Second, they may be due to a systematic bias in the distance-weighted least squares procedure. Third, the peaks may be due to familiarity effects. This alternative seemed especially likely since the low points of the saddles were between the category prototypes, above regions of the stimulus plane associated with high stimulus density (i.e., high familiarity).

In an effort to determine whether these saddles were statistical artifact or due to some real psychological phenomenon, we performed the following analysis. First, for each data set we selected the 50 stimuli that were nearest to the discriminant bound. This effectively isolates a swath of stimuli lying directly under the discriminant bound. Second, each of these stimulus sets was grouped into deciles with respect to distance from the origin (i.e., from the point (0,0)). This procedure is equivalent to making nine cuts perpendicular to the discriminant bound with the property that 5 stimuli lie between each pair of cuts. Third, mean RT was computed for the five stimuli comprising each decile. The resulting mean RTs are plotted[2] in Figure 4.

These plots show estimated mean RT along the subject's decision bound. If the saddles in the SYSTAT mean RT plots are due to a familiarity effect, the Figure 4 plots should be U-shaped, because the endpoints are associated with stimuli that are less familiar than the midpoints. As can be seen, however, there is no consistent U-shaped trend. Although some plots show a large deviation from the flat function predicted by the RT-distance hypothesis, the single function best fitting the entire set of plots would be very close to flat. In summary, Figure 4 indicates no support for the RT-familiarity hypothesis.[3]

y-axis = mean RT (in seconds)
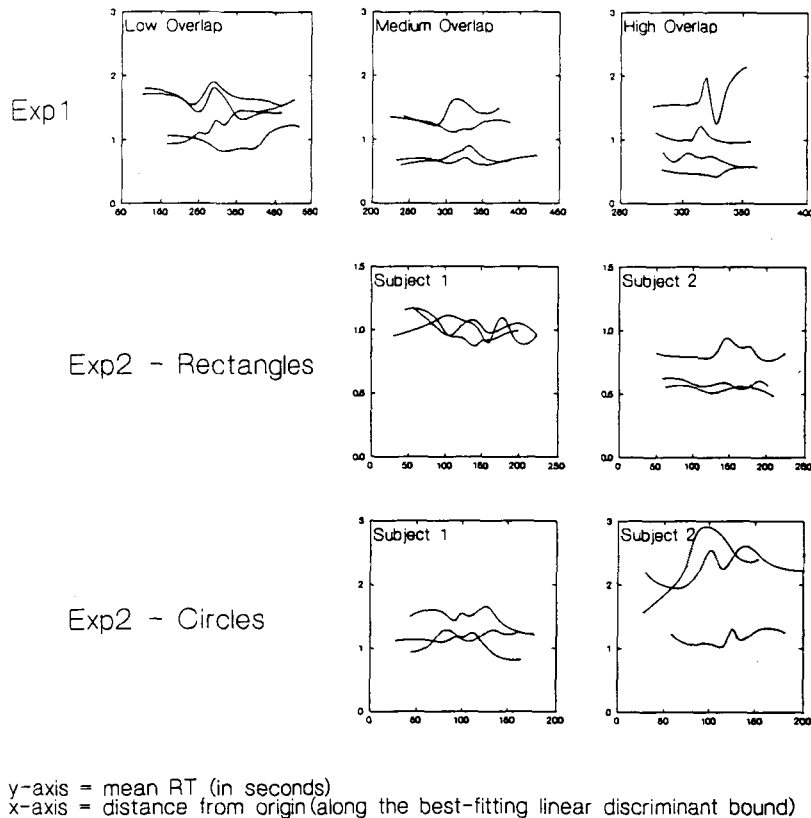x-axis = distance from origin (along the best-fitting linear discriminant bound)

Figure 4. Estimated mean response time profiles along the discriminant decision bound.

None of our analyses found any evidence that stimulus familiarity affects categorization RT. This does not mean that familiarity will never affect RT. For example, Nosofsky (1991) found long RTs to an unfamiliar stimulus far from the decision bound. However, in Nosofsky's experiment, the unfamiliar stimulus was many standard deviations away from its nearest neighbor. In our experiments, all stimuli were drawn randomly from (bivariate) normal distributions. Thus, the density of the stimuli in the stimulus space decreased continuously as one moved away from the category prototypes. Therefore, our results do not contradict the results of Nosofsky (1991). Taken together, the two studies suggest the following role of stimulus familiarity on categorization RT: Familiarity has negligible effects on RT except when the stimulus is so unfamiliar that its nearest neighbor is many standard deviations away.

EXPERIMENT 3

Experiments 1 and 2 falsified the RT-nearest-prototype and RT-familiarity hypotheses, but were unable to test between the RT-distance and RT-comparative-prototype hypotheses, because, in both experiments, these two hypotheses made identical predictions. As noted above, how-

ever, in an (A, not A) task, the RT-distance and RT-comparative-prototype hypotheses make strikingly different predictions. On trials in which the subject responds yes, both hypotheses predict that RT will increase with distance from the prototype. The RT-distance hypothesis makes this prediction because moving away from the prototype is equivalent to moving toward the decision bound. On trials in which the subject responds no, however, the two hypotheses make opposite predictions. The RT-comparative prototype hypothesis predicts that RT will continue to increase with distance from the prototype, but the RT-distance hypothesis predicts that RT will decrease with distance from the prototype, because, now, moving away from the prototype is equivalent to moving away from the decision bound. Experiment 3 was run to test these predictions.

Method

Subjects
There were 6 subjects. Three subjects participated in each of two conditions. Each subject completed one experimental session.

Stimuli
The contour of equal likelihood describing Category A was similar to the contours of the two categories in the medium–overlap condition of Experiment 1. There were two conditions. In one, stim-

uli were the rectangles of Figure 1B; in the other, stimuli were the circles of Figure 1C. In the circle condition, stimuli on *A* trials were drawn from a bivariate normal distribution with a mean of 250 and a standard deviation of 51.5 on each stimulus dimension. The same parameters were used in the rectangle condition except mean width was set to 350 (so the prototype was not square). In both conditions, stimuli on not *A* trials were drawn from a bivariate uniform distribution centered at the Category *A* mean. The uniform distribution extended from 0 to 500 on both stimulus dimensions (100 to 600 on the width dimension in the rectangle condition). The decision bound that maximizes categorization accuracy in this task is circular and centered at the Category *A* mean. The Category *A* variance was selected so that a subject using the optimal bound perfectly would have an overall accuracy rate of 90%. On average, the Category *A* stimuli were 5.1° wide and 3.7° high. The standard deviation on both dimensions was 0.76°.

Procedure

The procedures were the same as in Experiments 1 and 2 except that auditory rather than visual feedback was given. A correct response was indicated by a high-pitched tone, and an error was indicated by a low-pitched tone (both of 200-msec duration). The intertrial interval was 1.5 sec.

Results and Discussion

Accuracy Analysis

Table 10 describes characteristics of the subjects' responses in each condition. All subjects performed well below the 90% optimal accuracy level, which indicates that this was a difficult task. This conclusion is supported by a comparison of the discriminant and the optimal bounds. The discriminant bound was constrained to be a circle, but its location and radius were estimated from the data. Perhaps the most striking result of Table 10 is that the radius of the discriminant bound was significantly larger than the radius of the optimal bound for every subject, which indicates a large bias in favor of responding *yes*. There are also large discrepancies between the centers of the optimal and discriminant bounds, but these tend not to be systematic. The only exception is that in the rectangle condition, all subjects centered the discriminant bound to the right of the optimal center. On the other hand, given the large radii of the discriminant bounds, the discrepancies between the centers of the discriminant and optimal bounds should not be overemphasized. In almost all cases, the optimal bound was completely contained

within the discriminant bound. Thus, the subjects almost always agreed with the optimal classifier on *yes* responses.

Response Time Analysis

Four different sets of Spearman rank-order correlations were computed separately for *yes* and for *no* responses.[4] These included two versions of the RT–distance hypothesis and two versions of the RT–prototype hypothesis. Response times were correlated with two distance-to-bound measures and two distance-to-prototype measures: (1) distance to the optimal bound, (2) distance to the discriminant bound, (3) distance to the center of the optimal *yes* region (the optimal prototype), and (4) distance to the center of the best-fitting *yes* region (i.e., as defined by the discriminant bound; the resulting point is therefore called the discriminant prototype). Results are shown in Table 11.

For both *yes* and *no* responses, the RT–distance hypothesis predicts negative correlations for the two distance-to-bound measures and the RT–comparative-prototype hypothesis predicts positive correlations for the two distance-to-prototype measures. As Table 11 indicates, both of these predictions were supported on *yes* trials. Note also that for both hypotheses, correlations on *yes* trials were generally more extreme when computed with respect to the discriminant bound than when computed with respect to the optimal bound. A comparison of the discriminant bound correlations favors the RT–prototype hypothesis, but only slightly.

On *no* trials, the correlations with the two distance-to-bound measures were again negative, as predicted by the RT–distance hypothesis, but the correlations with the two distance-to-prototype measures were also negative, in contradiction of the RT–prototype hypothesis. The fastest *no* responses were to those stimuli furthest from the bound that separates the *yes* and *no* responses, not to those stimuli nearest to the category prototype.

These results clearly indicate that RT is not controlled by the category prototype. Specifically, knowledge of the prototype location is insufficient to allow one to predict which stimuli will be associated with the slowest RTs. On the other hand, if the decision bound is known, the stimuli associated with both the fastest and slowest RTs can be identified a priori. This does not mean that the pro-

Table 10
Characteristics of the Responses Given in Experiment 3

Condition	Subject	% Correct	Optimal Bound				Discriminant Bound			
			Radius	x Center	y Center	% of Responses Accounted For	Radius	x Center	y Center	% of Responses Accounted For
Rectangles	1	77	84.8	350	250	73	160.5	419.2	261.7	88
	2	74	84.8	350	250	75	135.4	411.6	204.8	80
	3	65	84.8	350	250	66	151.3	383.4	185.8	84
Circles	1	77	84.8	250	250	79	117.9	227.7	249.2	87
	2	79	84.8	250	250	74	139.4	227.9	268.9	88
	3	82	84.8	250	250	84	144.0	216.4	268.6	90

Table 11
Spearman Rank-Order Correlations for Experiment 3

Condition	Subject	RT–Distance (Optimal Bound)	RT–Distance (Discriminant Bound)	RT–Prototype (Optimal Prototype)	RT–Prototype (Discriminant Prototype)
		Yes Responses			
Rectangles	1	.054	−.170*	.036	.187*
	2	−.143†	−.124	.160†	.191†
	3	.060	−.148†	.090	.158†
Circles	1	−.174†	−.230†	.210†	.221†
	2	.015	−.108	.190†	.184†
	3	−.002	.074	.075	−.048
		No Responses			
Rectangles	1	−.462*	−.415*	−.465	−.477
	2	−.363*	−.282*	−.397	−.349
	3	.095	−.045	.080	−.029
Circles	1	−.467*	−.424*	−.432	−.382
	2	−.249*	−.242*	−.239	−.174
	3	−.345*	−.309*	−.340	−.306

*$p < .001$.   †$.001 < p < .05$.

totype played no role in the decision processes used by subjects in this experiment. In fact, both the optimal bound and the discriminant bound are compatible with the following rule: *Compute the similarity of the stimulus to the prototype. If this similarity exceeds a criterion value respond* yes; *if not, respond* no. However, with this rule, RT is controlled by the criterion value (i.e., the decision bound), not by the prototype. For example, consider another experiment in which task instructions and/or payoffs are used to induce subjects to adopt a higher criterion for responding *yes*. We expect such a manipulation to have no effect on the prototype location but a pronounced effect on the decision bound. Specifically, if the criterion for responding *yes* is increased, the diameter of the circular decision bound will decrease. The RT–distance hypothesis and the results of Experiment 3 suggest that decreasing the diameter of the decision bound would have a pronounced effect on RT. In contrast, the RT–comparative-prototype hypothesis predicts no effect on RT, since the prototype is unchanged.

GENERAL DISCUSSION

Although the RT–distance hypothesis received general support in all three experiments, other RT hypotheses may have been equally successful. For example, a general tack for deriving hypotheses relating RT to characteristics of the stimuli is to assume that RT increases with task difficulty. In a task with Categories A and B, a statistic that measures the ease of categorizing Stimulus X is $D = |P(A|X) - 0.5|$. If Stimulus X is easy to categorize, the subject will assign it consistently to the same category, and D will be large. If it is a difficult decision, the subject will sometimes assign it to one category and sometimes to the other, and D will be small. Therefore, RT hypotheses can be generated by assuming that RT is negatively correlated with the statistic D.

One example of a popular categorization theory for which no RT hypotheses have been derived is exemplar

theory (Estes, 1986; Medin & Schaffer, 1978; Nosofsky, 1986). Exemplar theory assumes that in the absence of response bias, $P(A|X)$ increases monotonically with the sum of the similarities between Stimulus X and all exemplars of Category A divided by the sum of the similarities between X and all exemplars of Category B. By incorporating this fact with the statistic D, a plausible RT–exemplar hypothesis could be derived from exemplar theory.

Unfortunately, the present data are not appropriate for testing between such an RT–exemplar hypothesis and the RT–distance hypothesis. The results of Ashby and Maddox (1993; see also Nosofsky, 1990) suggest that when the optimal decision bound is linear, decision-bound and exemplar models make predictions that are so similar that no ordinal tests could discriminate between them. Because the RT tests reported here are ordinal, the RT–exemplar hypothesis would fare as well as the RT–distance hypothesis with our data. A test between these two hypotheses requires complex nonlinear decision bounds and/or RT predictions at either the interval or ratio scale level.

In summary, no evidence was found that stimulus familiarity or the category prototypes play any special role in determining categorization RT. Even the slowest responses showed little or no familiarity effect, and the prototypes showed no RT advantage. In contrast, the RT–distance hypothesis was well supported by all data except those for Day 1 of Experiment 2. All of the RT hypotheses failed badly in this case.

REFERENCES

ASHBY, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449-483). Hillsdale, NJ: Erlbaum.

ASHBY, F. G., & GOTT, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 14, 33-53.

ASHBY, F. G., & LEE, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, 120, 150-172.

ASHBY, F. G., & LEE, W. W. (1992). On the relationship among iden-

tification, similarity, and categorization: Reply to Nosofsky and Smith (1992). *Journal of Experimental Psychology: General, 121*, 385-393.

ASHBY, F. G., & MADDOX, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception & Performance, 16*, 598-612.

ASHBY, F. G., & MADDOX, W. T. (1991). A response time theory of perceptual independence. In J. P. Doigon & J. C. Falmagne (Eds.), *Mathematical psychology: Current developments* (pp. 389-413). New York: Springer-Verlag.

ASHBY, F. G., & MADDOX, W. T. (1992a). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception & Performance, 18*, 50-71.

ASHBY, F. G., & MADDOX, W. T. (1992b). *A response time theory of separability and integrality in speeded classification.* Manuscript submitted for publication.

ASHBY, F. G., & MADDOX, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology, 37*, 372-400.

ASHBY, F. G., & PERRIN, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review, 95*, 124-150.

BINDRA, D., DONDERI, D. C., & NISHISATO, S. (1968). Decision latencies of "same" and "different" judgments. *Perception & Psychophysics, 3*, 121-130.

BINDRA, D., WILLIAMS, J. A., & WISE, J. S. (1965). Judgments of sameness and difference: Experiments on decision time. *Science, 150*, 1625-1627 (erratum, p. 1699).

BORNSTEIN, M. H., & MONROE, M. D. (1980). Chromatic information processing: Rate depends on stimulus location in the category and psychological complexity. *Psychological Research, 42*, 213-225.

BURNS, B., & HOPKINS, A. (1987). Integral and separable stimulus structure in iconic memory: A replication and extension of Burns (1987). *Perceptual & Motor Skills, 64*, 263-270.

BURNS, B., SHEPP, B. E., McDONOUGH, D., & WEINER-EHRLICH, W. K. (1978). The relation between stimulus analyzability and perceived dimensional structure. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 12, pp. 77-115). New York: Academic Press.

CARTWRIGHT, D. (1941). Relation of decision-time to the categories of response. *American Journal of Psychology, 54*, 174-196.

DUNN, J. C. (1983). Spatial metrics of integral and separable dimensions. *Journal of Experimental Psychology: Human Perception & Performance, 9*, 242-257.

EMMERICH, D. S., GRAY, J. L., WATSON, C. S., & TANIS, D. C. (1972). Response latency, confidence, and ROCs in auditory signal detection. *Perception & Psychophysics, 11*, 65-72.

ENNIS, D. M. (1988). Confusable and discriminable stimuli: Comment on Nosofsky (1986) and Shepard (1986). *Journal of Experimental Psychology: General, 117*, 408-411.

ESTES, W. K. (1986). Array models for category learning. *Cognitive Psychology, 18*, 500-549.

FELFOLDY, G. L. (1974). Repetition effects in choice reaction time to multidimensional stimuli. *Perception & Psychophysics, 15*, 453-459.

GARNER, W. R., & FELFOLDY, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1*, 225-241.

HOMA, D., STERLING, S., & TREPEL, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning & Memory, 7*, 418-439.

HYMAN, R., & FROST, N. H. (1975). Gradients and schema in pattern recognition. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and Performance V* (pp. 630-654). London: Academic Press.

HYMAN, R., & WELL, A. (1967). Judgments of similarity and spatial models. *Perception & Psychophysics, 2*, 233-248.

KRUEGER, L. E. (1989). Reconciling Fechner and Stevens: Toward a unified psychophysical law. *Behavioral & Brain Sciences, 12*, 251-320.

McLAIN, D. H. (1974). Drawing contours from arbitrary data points. *Computer Journal, 17*, 318-234.

MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.

MERVIS, C. B. (1980). Category structure and the development of categorization. In R. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 279-307). Hillsdale, NJ: Erlbaum.

MERVIS, C. B., CATLIN, J., & ROSCH, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society, 7*, 283-284.

MURDOCK, B. B., JR. (1985). An analysis of the strength-latency relationship. *Memory & Cognition, 13*, 511-521.

MURDOCK, B. B., JR., & ANDERSON, R. E. (1975). Encoding, storage, and retrieval of item information. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 145-194). Hillsdale, NJ: Erlbaum.

NORMAN, D. A., & WICKELGREN, W. (1969). Strength theory and decision rules and latency in retrieval from short-term memory. *Journal of Mathematical Psychology, 6*, 192-208.

NOSOFSKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.

NOSOFSKY, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14*, 700-708.

NOSOFSKY, R. M. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology, 34*, 393-418.

NOSOFSKY, R. M. (1991). Typicality in logically defined categories: Exemplar-similarity versus rule instantiation. *Memory & Cognition, 19*, 131-150.

NOSOFSKY, R. M., CLARK, S. E., & SHIN, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*, 282-304.

POSNER, M. I., & KEELE, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353-363.

POSNER, M. I., & KEELE, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83*, 304-308.

REED, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3*, 382-407.

RIPS, L. J., SHOBEN, E. J., & SMITH, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning & Verbal Behavior, 12*, 1-20.

ROSCH, E. (1973a). Natural categories. *Cognitive Psychology, 4*, 328-350.

ROSCH, E. (1973b). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111-144). New York: Academic Press.

ROSCH, E., SIMPSON, C., & MILLER, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception & Performance, 2*, 491-502.

SHEPARD, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika, 22*, 325-345.

SHEPARD, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology, 1*, 54-87.

SHEPARD, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*, 1317-1323.

SHIN, H. J., & NOSOFSKY, R. M. (1992). Similarity-scaling studies of "dot-pattern" classification and recognition. *Journal of Experimental Psychology: General, 121*, 278-304.

SMITH, E. E. (1968). Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin, 69*, 77-110.

STERNBERG, S. (1966). High-speed scanning in human memory. *Science, 153*, 652-654.

STEVENS, S. S. (1961). Psychophysics of sensory function. *American Scientist, 48*, 226-252.

THOMAS, E. A. C. (1971). Sufficient conditions for monotone hazard rate: An application to latency-probability curves. *Journal of Mathematical Psychology, 8*, 303-332.

TOWNSEND, J. T., HU, G. G., & ASHBY, F. G. (1981). Perceptual sampling of orthogonal straight line features. *Psychological Research, 43*, 259-275.

WICKELGREN, W. A., & NORMAN, D. A. (1966). Strength models and

serial position in short-term recognition memory. *Journal of Mathematical Psychology*, **3**, 316-347.

WILKINSON, L. (1990). *SYSTAT*. Evanston, IL: SYSTAT, Inc.

NOTES

1. At first, these correlations may seem unimpressive, but it should be recalled that, on every trial, the subject saw a different stimulus. During a session of 300 trials, there were 300 different RTs and 300 different distances to the decision bound. Therefore, even if the RT–distance hypothesis were exactly correct, variability in response execution time (i.e., motor time) would severely depress the RT–distance correlations.

2. Because of the small sample sizes in each decile (i.e., 5), the data were smoothed with a moving window. Let RT_k represent the mean RT in decile k. Then, this procedure replaces each RT_k with $(RT_{k-1} + RT_k + RT_{k+1})/3$, for $2 \leq k \leq 9$. RT_1 is replaced with $(RT_1 + RT_2)/2$, and RT_{10} is replaced with $(RT_9 + RT_{10})/2$.

3. On the basis of these results and some more extensive investigations, we believe that the distance-weighted least squares procedure (McLain, 1974) is biased, or at lease misleading, in the periphery. The problem is that the procedure extrapolates a function to a set of predefined endpoints, even in the absence of any data.

4. Median RTs are not reported because the RT–distance hypothesis makes no definite predictions about the relation of correct to incorrect median RT when the optimal decision bound is not linear.

Fourth International Conference on Visual Search
Koningshof, Eindhoven, The Netherlands
August 29–September 1, 1994

Call for Papers

The Fourth International Conference on Visual Search will be held in Eindhoven from August 29 to September 1, 1994.

The deadline for receipt of abstracts is February 5, 1994 and the deadline for full manuscripts is July 8, 1994.

For further information, contact the ICVS4 Conference Secretariat, Applied Vision Research Unit, University of Derby, Mickleover, Derby DE3 5GX, U.K. [Tel: 44(0)332-622286; Fax: 44(0)332-622287; Email: avru@derby.ac.uk]