Comparing decision bound and exemplar models of categorization

W. TODD MADDOX and F. GREGORY ASHBY
University of California, Santa Barbara, California

The performance of a decision bound model of categorization (Ashby, 1992a; Ashby & Maddox, in press) is compared with the performance of two exemplar models. The first is the generalized context model (e.g., Nosofsky, 1986, 1992) and the second is a recently proposed deterministic exemplar model (Ashby & Maddox, in press), which contains the generalized context model as a special case. When the exemplars from each category were normally distributed and the optimal decision bound was linear, the deterministic exemplar model and the decision bound model provided roughly equivalent accounts of the data. When the optimal decision bound was nonlinear, the decision bound model provided a more accurate account of the data than did either exemplar model. When applied to categorization data collected by Nosofsky (1986, 1989), in which the category exemplars are not normally distributed, the decision bound model provided excellent accounts of the data, in many cases significantly outperforming the exemplar models. The decision bound model was found to be especially successful when (1) single subject analyses were performed, (2) each subject was given relatively extensive training, and (3) the subject's performance was characterized by complex suboptimalities. These results support the hypothesis that the decision bound is of fundamental importance in predicting asymptotic categorization performance and that the decision bound models provide a viable alternative to the currently popular exemplar models of categorization.

Decision bound models of categorization (Ashby, 1992a; Ashby & Maddox, in press) assume that the subject learns to assign responses to different regions of perceptual space. When categorizing an object, the subject determines in which region the percept has fallen and then emits the associated response. The *decision bound* is the partition between competing response regions. In contrast, exemplar models assume that the subject computes the sum of the perceived similarities between the object to be categorized and *every* exemplar of each relevant category (Medin & Schaffer, 1978; Nosofsky, 1986). Categorization judgments are assumed to depend on the relative magnitude of these various sums.

This article compares the ability of decision bound and exemplar models to account for categorization response probabilities in seven different experiments. The aim is not to compare the performance of the decision bound model with *all* versions of exemplar theory; clearly this is beyond the scope of any single article. Rather, the goal is to compare the decision bound model with one of the most highly successful and widely tested versions of exemplar theory—namely, the generalized context model (GCM; Nosofsky, 1986, 1987, 1989; Nosofsky, Clark, & Shin, 1989; Shin & Nosofsky, 1992). In addition, a recently proposed deterministic exemplar model (DEM; Ashby & Maddox, in press), which contains the GCM as a special case, will also be applied to the data. The GCM, which has been applied to a wide variety of categorization conditions (see Nosofsky, 1992, for a review), provides good quantitative accounts of the data, in many instances accounting for over 98% of the variance. There have been cases, however, in which the model fits were less satisfactory, accounting for less than 85% of the variance in the data (see Nosofsky, 1986, Table 4).

Of the seven experiments, the first five involve categories in which the exemplars are normally distributed along each stimulus dimension and single subject analyses are performed. To date, the GCM only has been applied to data sets in which the category exemplars are not normally distributed, and in only one of these cases (Nosofsky, 1986) have single subject analyses been performed. This article provides the first test of the GCM's ability to account for data from normally distributed categories. The final two data sets were reported by Nosofsky (1986, 1989) and involve experiments in which the category exemplars are *not* normally distributed and only a small number of exemplars were utilized. Three of the

four categorization conditions are identical in the two studies, and the stimulus dimensions are the same. The main difference between the two studies is that single subject analyses were performed on the Nosofsky (1986) data, and the data were averaged across subjects in the Nosofsky (1989) data.

DECISION BOUND THEORY

Decision bound theory (also called general recognition theory) rests on one critical assumption—namely, that there is trial-by-trial variability in the perceptual information associated with every stimulus. In the case of threshold level stimuli, this assumption dates back to Fechner (1866) and was exploited fully in signal detection theory (Green & Swets, 1966). However, with the high-contrast stimuli used in most categorization experiments, the assumption might appear more controversial. There are at least two reasons, however, why even in this case, variability in the percept is expected. First, it is well known that the number of photons striking each rod or cone in the retina during presentation of a visual stimulus has trial-by-trial variability. In fact, the number has a Poisson probability distribution (Barlow & Mollon, 1982), and one characteristic of the Poisson distribution is that its mean equals its variance. Thus, more intense stimuli are associated with greater trial-by-trial variability at the receptor level. Specifically, the standard deviation of the number of photons striking each rod or cone is proportional to the square root of the mean stimulus intensity. Second, the visual system is characterized by high levels of spontaneous activity. For example, ganglion cells in the optic nerve often have spontaneous firing rates of as high as 100 spikes per second.

If this argument is accepted, then a second question to be asked is whether such variability is likely to affect the outcome of categorization judgments. For example, when one is classifying pieces of fruit such as apples or oranges, trial-by-trial variability in perceived color (i.e., in hue) is unlikely to lead to a categorization error. Even if perceptual variability does not affect the outcome of a categorization judgment, however, the existence of such variability has profound effects on the nature of the decision process. For example, in the presence of perceptual variability, the decision problem in a categorization task is identical to the decision problem in an identification task. In both cases, the subject must learn the many different percepts that are associated with each response. As a consequence, a theory of identification that postulates perceptual variability needs no extra structure to account for categorization data. In contrast, the most widely known versions of exemplar theory, including the context model (Medin & Schaffer, 1978) and the generalized context model (Nosofsky, 1986), postulate a set of decision processes that are unique to the categorization task.

To formalize this discussion, consider an identification task with two stimuli, A and B, which differ on two phys-

ical dimensions. In this case, stimulus i (i = A or B) can be described by the vector

$$\mathbf{v}_i = \begin{bmatrix} v_{1i} \\ v_{2i} \end{bmatrix}$$

where v_{1i} and v_{2i} are values of stimulus i on physical dimensions 1 and 2, respectively. Decision bound theory assumes that the subject misperceives the true stimulus coordinates $\mathbf{v}_i$ because of trial-by-trial variability in the percept (i.e., perceptual noise) and because the mapping from the physical space to the perceptual may be non-linear (e.g., because of response compression during sensory transduction).[1] Denote the subject's mean percept of stimulus i by $\mathbf{x}_i$. A natural assumption would be that $\mathbf{x}_i$ is related to $\mathbf{v}_i$ via a power or log transformation. In either case, the subject's percept of stimulus i is represented by

$$\mathbf{x}_{pi} = \mathbf{x}_i + \mathbf{e}_p, \tag{1}$$

where $\mathbf{e}_p$ is a random vector that represents the effects of perceptual noise.[2]

Given the Equation 1 model of the perceptual representation, the next step in building a theory of either identification or categorization is to postulate a set of appropriate decision processes. The ability to identify or categorize accurately is of fundamental importance to the survival of every biological organism. Plants must be categorized as edible or poisonous. Faces must be categorized as friend or foe. In fact, every adult has had a massive amount of experience identifying and categorizing objects and events. In addition, much anecdotal evidence testifies to the expertise of humans at categorization and identification. For example, despite years of effort by the artificial intelligence community, humans are far better at categorizing speech sounds or handwritten characters than are the most sophisticated machines.

These facts suggest that a promising method for developing a theory of human decision processes in identification or categorization is to study the decision processes of the optimal classifier—that is, of the device that maximizes identification or categorization accuracy. The optimal classifier has a number of advantages over any biological organism, so humans should not be expected to respond optimally. Nevertheless, one might expect humans to use the same strategy as the optimal classifier does, even if they do not apply that strategy with as much success.

The optimal classifier was first studied by R. A. Fisher more than 50 years ago (Fisher, 1936) and today its behavior is well understood (see, e.g., Fukunaga, 1990; Morrison, 1990). Consider an identification or categorization experiment with response alternatives A and B. Because of the Poisson nature of light, even the optimal classifier must deal with trial-by-trial variability in the stimulus information. Suppose the stimulus values recorded by the optimal classifier, on trials when stimulus i

is presented, are denoted by $\mathbf{w}_i$. Then the optimal classifier constructs a discriminant function $h_o(\mathbf{w}_i)$ and responds A or B according to the rule:

$$\text{if } h_o(\mathbf{w}_i) \begin{cases} < 0; & \text{then respond } A \\ = 0; & \text{then guess} \\ > 0; & \text{then respond } B \end{cases} \qquad (2)$$

The discriminant function partitions the perceptual space into two response regions. In one region [where $h_o(\mathbf{w}_i) < 0$], response A is given. In the other region [where $h_o(\mathbf{w}_i) > 0$], response B is given. The partition between the two response regions [where $h_o(\mathbf{w}_i) = 0$] is the *decision bound*. The position and shape of the decision bound depends on the stimulus ensemble and on the exact distribution of noise associated with each stimulus. If the noise is normally distributed and the task involves only two stimuli, the optimal bound is always either a line or a quadratic curve.[3]

Although identification and categorization are treated equivalently by the optimal classifier, in practice these two tasks provide insights about different components of processing. Identification experiments are good for studying the distribution of perceptual noise (Ashby & Lee, 1991). This is because, with confusable stimuli, small changes in the perceptual distributions can have large effects on identification accuracy (Ennis & Ashby, in press). In contrast, in most categorization experiments, overall accuracy is relatively unaffected by small changes in the perceptual distributions. Most categorization errors occur because the subject incorrectly learned the rule for assigning stimuli to the relevant categories. Thus, categorization experiments are good for testing the hypothesis that subjects use decision bounds.

In many categorization experiments, each category contains only a small number of exemplars (typically fewer than seven). Such a design is not the best for investigating the possibility that subjects use decision bounds, because with such few stimuli, many different bounds will typically yield identical accuracy scores. Ideally, each category would contain an unlimited number of exemplars and competing categories would overlap. In such a case, any change in the decision bound would lead to a change in accuracy.

Ashby and his colleagues (Ashby & Gott, 1988; Ashby & Maddox, 1990, 1992) have reported the results of a number of categorization experiments in which the exemplars in each category had values on each stimulus dimension that were normally distributed. All experiments involved two categories and stimuli that varied on two dimensions. Representative stimuli are shown in Figure 1. For example, in an experiment with the circular stimuli, on each trial a random sample is drawn from either the Category A or the Category B (bivariate) normal distribution. This specifies an order pair (v_{1i}, v_{2i}). A circle is then presented with diameter v_{1i} and with a radial line of orientation v_{2i}. The subject's task is to determine whether the stimulus is a member of Category A or Category B. Feedback is given on each trial. In each experi-
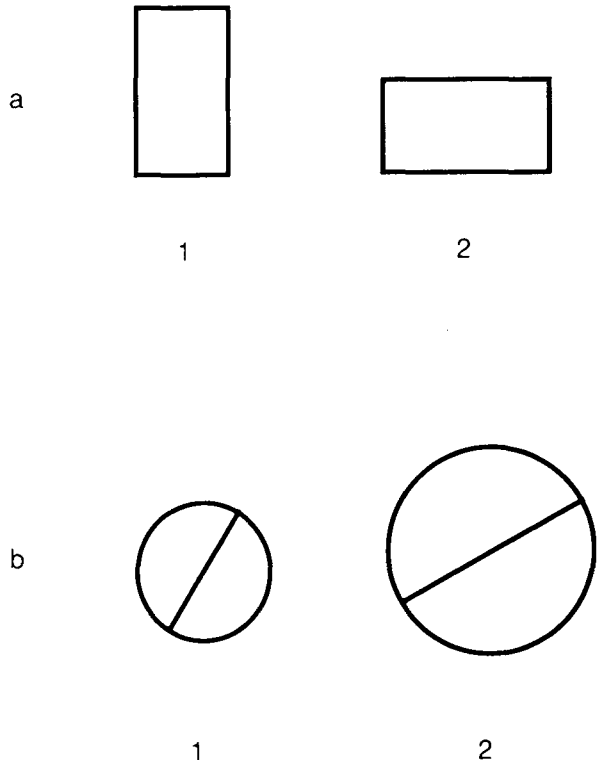

Figure 1. Representative stimuli from (a) Condition R (rectangular stimuli) and (b) Condition C (circular stimuli) of Application 1: Data Sets 1–5.

ment, Ashby and his colleagues tested whether each subject's A and B responses were best separated by the optimal bound[4] or by a bound predicted by one of several popular categorization models (e.g., prototype, independent decisions, and bilinear models). Although subjects did not respond optimally, the best predictor of the categorization performance of experienced subjects, across all the experiments, was the optimal classifier. When subjects responded differently than the optimal classifier, the bound that best described their performance was always of the same type as the optimal bound. Specifically, when the optimal bound was a quadratic curve, the best-fitting bound was a quadratic curve, and when the optimal bound was linear, the best-fitting bound appeared to be linear.

These facts suggest that the notion of a decision bound may have some fundamental importance in human categorization. That is, rather than compute similarity to the category prototypes, or add the similarity to all category exemplars, perhaps subjects behave as the optimal classifier and refer directly to some decision bound. According to this view, experienced categorization (i.e., after the decision bound is learned) is essentially automatic. When faced with a stimulus to categorize, the subject determines the region in which the stimulus representation has fallen and then emits the associated response. Exemplar information is not needed; only a response label is

retrieved. This does not mean that exemplar information is unavailable, however, because presumably exemplar information is used to construct the decision bound.

Although the data provide tentative support for the notion of a decision bound, the data also suggest that subjects do not respond optimally. A theory of human categorization must account for the fundamental suboptimalities of human perceptual and cognitive processes.

Decision bound theory assumes that subjects attempt to respond optimally but fail because, unlike the optimal classifier, humans (1) do not know the locations of every exemplar within each category, (2) do not know the true parameters of the perceptual noise distributions, (3) sometimes misremember or misconstruct the decision boundary, and (4) sometimes have an irrational bias against a response alternative.

The first two forms of suboptimality cause the subject to use a decision bound that is different from the optimal bound. Specifically, decision bound theory assumes that rather than using the optimal discriminant function $h_o(x_{pi})$ in the Equation 2 decision rule, the subject uses a suboptimal function $h(x_{pi})$ that is of the same functional form as the optimal function (i.e., a quadratic curve when the optimal bound is a quadratic curve, and a line when the optimal bound is linear). The third cause of suboptimal performance in biological organisms is imperfect memory. Trial-by-trial variability in the subject's memory of the decision bound is called *criterial noise*. In decision bound theory, criterial noise is represented by the random variable e_c (normally distributed with mean 0 and variance σ_c^2), which is assumed to have an additive effect on the discriminant value $h(x_{pi})$. A final cause of suboptimality is a bias against one (or more) response alternative. A response bias can be modeled by assuming that rather than compare the discriminant value to zero, as in the Equation 2 decision rule, the subject compares it to some value δ. A positive value of δ represents a bias against response alternative B.

In summary, decision bound theory assumes that rather than use the optimal decision rule of Equation 2, the subject uses the following rule:

$$\text{if } h(x_{pi}) + e_c \begin{cases} < \delta; \text{ then respond } A \\ = \delta; \text{ then guess} \\ > \delta; \text{ then respond } B \end{cases} \quad (3)$$

As a consequence, the probability of responding A on trials when stimulus i is presented is given by

$$P(A \mid \text{stimulus } i) = P[h(x_{pi}) + e_c < \delta \mid \text{stimulus } i]. \quad (4)$$

Several versions of this model can be constructed, depending on the form of the decision bound. In this article, we consider (1) the general quadratic classifier, (2) the general linear classifier, and (3) the optimal decision bound model (not to be confused with the optimal classifier discussed above). The three versions will be only briefly introduced here. More detailed derivations of each of these three models is given in Ashby and Maddox (in press).

In an experiment with normally distributed categories, the optimal classifier uses a quadratic decision bound if stimulus variability within Category A differs from the variability within Category B along any stimulus dimension or if the two categories are characterized by a different set of covariances. The general quadratic classifier assumes that the subject attempts to respond optimally but misestimates some of the category means, variances, or covariances and therefore uses a quadratic decision bound that differs from the optimal.

With the two perceptual dimensions x_1 and x_2, the decision bound of the general quadratic classifier satisfies

$$h(x_1, x_2) = a_1 x_1^2 + a_2 x_2^2 + a_3 x_1 x_2$$
$$+ b_1 x_1 + b_2 x_2 + c_0 \quad (5)$$

for some constants a_1, a_2, a_3, b_1, b_2, and c_0. An important property of this model is that the effects of perceptual and criterial noise can be estimated uniquely. Separate estimates of perceptual and criterial noise have been obtained in the past (see, e.g., Nosofsky, 1983), but these have required a comparison of performance across several different experiments.

If the variability within Category A is equal to the variability within Category B along both dimensions, and if the two categories are characterized by the same covariance, then the decision bound of the optimal classifier is linear. The general linear classifier assumes that the subject makes the inference (perhaps incorrectly) that these conditions are true, so he or she chooses a linear decision bound; but because the category means, variances, and covariance are unknown, a suboptimal linear bound is chosen. The general linear classifier is a special case of the general quadratic classifier in which the coefficients a_1, a_2, and a_3 in Equation 5 are zero.

At this point, decision bound theory makes no assumptions about categorization at the algorithmic level—that is, about the details of how the subject comes to assign responses to regions. There are a number of possibilities.[5] Some require little computation on the part of the subject. Specifically, it is not necessary for the subject to estimate category means, variances, covariances, or likelihoods, even when responding optimally. In this case, the only requirement is that the subject be able to learn whether a given percept is more likely to have been generated by an exemplar from Category A or B.

EXEMPLAR THEORY

Generalized Context Model

Exemplar theory (see, e.g., Estes, 1986a, 1986b; Smith & Medin, 1981) assumes that, on each trial of a categorization experiment, the subject performs some sort of global match between the representation of the presented stimulus and the memory representation of *every* exemplar of each category and chooses a response on the basis of these similarity computations. The assumption that the global matching operation includes *all* members of each

category seems viable in the kinds of categorization tasks used in many laboratory experiments, because it is a common experimental practice to construct categories with only four to six exemplars. With natural categories, however, the assumption seems less plausible. For example, when one is deciding that a chicken is a bird, it seems unlikely that one computes the similarity of the chicken in question to every bird one has ever seen. Of course, if performed in parallel, this massive amount of computation may occur, but it certainly disagrees with introspective experience.

Perhaps the most widely known of the exemplar models is the GCM, developed by Medin and Schaffer (1978) and elaborated by Estes (1986a) and Nosofsky (1984, 1986). According to the GCM, the probability that stimulus i is classified as a member of Category A, $P(A|i)$, is given by

$$P(A|i) = \frac{\beta \sum\limits_{j \in C_A} \eta_{ij}}{\beta \sum\limits_{j \in C_A} \eta_{ij} + (1-\beta) \sum\limits_{j \in C_B} \eta_{ij}}, \quad (6)$$

where $j \in C_J$ represents all exemplars of Category J, η_{ij} is the similarity of stimulus i to exemplar j, and β is a response bias. The similarity η_{ij} between a pair of stimuli is assumed to be a monotonically decreasing function of the psychological distance between point representations of the two stimuli. Thus, the GCM assumes no trial-by-trial variability in the perceptual representation. The psychological distance between stimuli i and j is given by

$$d_{ij} = c[w|x_{1i}-x_{1j}|^r + (1-w)|x_{2i}-x_{2j}|^r]^{1/r}, \quad (7)$$

where w is the proportion of attention allocated to Dimension 1 and the nonnegative parameter c scales the psychological space. The parameter c can be interpreted as a measure of overall stimulus discriminability and should increase with increased exposure duration or as subjects gain experience with the stimuli (Nosofsky, 1985, 1986). The exponent $r \geq 1$ defines the nature of the distance metric. The most popular cases occur when $r = 1$ (city-block distance) and when $r = 2$ (Euclidean distance).

Two specific functions relating psychological distance to similarity are popular. The *exponential decay function* assumes that the similarity between stimuli i and j is given by (e.g., Ennis, 1988; Nosofsky, 1988a; Shepard, 1957, 1964, 1986, 1987, in press)

$$\eta_{ij} = \exp(-d_{ij}). \quad (8)$$

In contrast, the *Gaussian function* assumes (e.g., Ennis, 1988; Ennis, Mullen, & Frijters, 1988; Nosofsky, 1988a; Shepard, 1986, 1987, 1988) that

$$\eta_{ij} = \exp(-d_{ij}^2). \quad (9)$$

In most applications of the GCM, the exponential similarity function is paired with either the city-block or the Euclidean distance metric or else the Gaussian similarity function is paired with the Euclidean metric (e.g.,

Nosofsky, 1986, 1987). The GCM has accounted successfully for the relationship between identification, categorization, and recognition performance with stimuli constructed from both separable and integral dimensions under a variety of different conditions (see Nosofsky, 1992, for a review).

A Deterministic Exemplar Model

The act of categorization can be subdivided into two components (see, e.g., Massaro & Friedman, 1990). The first involves accessing the category information that is assumed relevant to the decision-making process, and the second involves using this information to select a response. The decision bound model and the GCM differ on both of these components. First, the decision bound model assumes that the subject retrieves the response label associated with the region in which the stimulus representation falls, whereas the GCM assumes that the subject performs a global similarity match to all exemplars of each category. Second, the decision bound model assumes a deterministic decision process (i.e., Equation 3), whereas the GCM assumes a probabilistic decision process (i.e., Equation 6) that is described by the similarity-choice model of Luce (1963) and Shepard (1957). In a deterministic decision process, the subject *always* gives the same response, given the same perceptual and cognitive information. In a probabilistic process, the perceptual and cognitive information is used to compute the *probability* associated with each response alternative. Thus, given the same information, sometimes one response is given and sometimes another. A poor fit of one model relative to the other could be attributed to either component. Thus, it is desirable to investigate a model that differs from both the GCM and the decision bound model on only one of these two components.

Probabilistic versions of the decision bound model could be constructed and so could deterministic versions of exemplar theory. As we will see in the next section, however, the data seem to support a deterministic decision process, and for this reason, it is especially interesting to examine a deterministic version of exemplar theory.

Nosofsky (1991) proposed a deterministic exemplar model in which the summed similarity of the probe to all exemplars of each category are compared and the response associated with the highest summed similarity is given. This model, however, does not contain the GCM as a special case. Ashby and Maddox (in press) proposed a deterministic exemplar model in which the relevant category information is the log of the summed similarity of the probe to all exemplars of each category. Specifically, the model assumes that the subject uses the decision rule

Respond A if $\log(\sum \eta_{iA}) > \log(\sum \eta_{iB})$;

Otherwise respond B, (10)

where $\sum \eta_{iJ}$ represents the summed similarity of stimu-

lus i to all members of Category J. The log is important because the resulting model contains the GCM as a special case. With criterial noise and a response bias δ, the Equation 10 decision rule becomes

Respond A if $\log(\sum \eta_{iA}) - \log(\sum \eta_{iB}) > \delta + e_c$;

Otherwise respond B,   (11)

where the subject is biased against B if $\delta < 0$. Now, if e_c has a logistic distribution with a mean of 0 and variance σ_c^2, then the probability of responding A given stimulus i can be shown to equal (Ashby & Maddox, in press)

$$P(A \mid i) = \frac{\beta(\sum \eta_{iA})^\gamma}{\beta(\sum \eta_{iA})^\gamma + (1-\beta)(\sum \eta_{iB})^\gamma},   (12)$$

where

$$\gamma = \frac{\pi}{\sqrt{3}\sigma_c} \text{ and } \beta = \frac{e^{\delta\gamma}}{1+e^{\delta\gamma}}.$$

Thus, this model is equivalent to the GCM when $\gamma = 1$. In other words, the GCM can be interpreted as a deterministic exemplar model in which the criterial noise variance $\sigma_c^2 = \pi^2/3$.

The γ parameter indicates whether response selection is more or less variable than is predicted by the GCM. If $\gamma < 1$, the transition from a small value of $P(A \mid i)$ to a large value is more gradual than is predicted by the GCM; response selection is too variable. If $\gamma > 1$, response selection is less variable than the GCM predicts.

MODEL FITTING AND TESTING

When testing the validity of a model with respect to a particular data set, one must consider two problems. The first is to determine how unknown parameters will be estimated; the second is to determine how well the model describes ("fits") the data. The *method of maximum likelihood* is probably the most powerful method (see Ashby, 1992b; Wickens, 1982).[6] Consider an experiment with Categories A and B and a set of n stimuli, $S_1, S_2, \ldots, S_n$. For each stimulus, a particular model predicts the probabilities that the subject will respond A and B, which we denote by $P(A \mid S_i)$ and $P(B \mid S_i)$, respectively. The results of an experimental session are a set of n responses, $r_1, r_2, \ldots, r_n$, where we arbitrarily set $r_i = 1$ if response A was made to stimulus i and $r_i = 0$ if response B was made. According to the model, and assuming that the responses are independent, the likelihood of observing this set of n responses is

$$L(r_1, r_2, \ldots, r_n) = \prod_{i=1}^{n} P(A \mid S_i)^{r_i} P(B \mid S_i)^{1-r_i}.$$

The maximum likelihood estimators are those values of the unknown parameters that maximize $L(r_1, r_2, \ldots, r_n)$ [denoted $L(r)$ for short].

Maximum likelihood estimates are also convenient when one wishes to evaluate the empirical validity of a model. For example, under the null hypothesis that the

model is correct, the statistic $G^2 = -2\ln L(r)$ has an asymptotic chi-square distribution with degrees of freedom equal to the number of experimental trials (n) minus the number of free parameters in the model. Rather than assess the absolute ability of a model to account for a set of data, it is often more informative to test whether a more general model fits significantly better than a restricted model. Consider two models, M_1 and M_2. Suppose Model M_1 is a special case of Model M_2 (i.e., M_1 is *nested* within M_2) in the sense that M_1 can be derived from M_2 by setting some of the free parameters in M_2 to fixed values. Let G_1^2 and G_2^2 be the goodness-of-fit values associated with the two models. Because M_1 is a special case of M_2, note that G_2^2 can never be larger than G_1^2. If Model M_1 is correct, the statistic $G_2^2 - G_1^2$ has an asymptotic chi-square distribution with degrees of freedom equal to the difference in the number of free parameters between the two models. Using this procedure, one can therefore test whether the extra parameters of the more general model lead to a significant improvement in fit (see Wickens, 1982, for an excellent overview of parameter estimation and hypothesis testing using the method of maximum likelihood).

The G^2 tests work fine when one model is a special case of the other. For example, G^2 tests can be used to determine whether the extra parameters of the general quadratic classifier provide a "significant" improvement in fit over the general linear classifier. One goal of this article, however, is to test models that are not special cases of one another (i.e., models that are not nested), such as the proposed comparisons between the exemplar and decision bound models. Fortunately, a goodness-of-fit statistic called Akaike's (1974) information criterion (AIC) has been developed that allows comparisons between models that are not nested, such as the exemplar and decision bound models. The AIC statistic, which generalizes the method of maximum likelihood, is defined as

$$\begin{aligned} \text{AIC}(M_i) &= -2\ln L_i + 2N_i \\ &= G_i^2 + 2N_i,   (13) \end{aligned}$$

where N_i is the number of free parameters in Model M_i and $\ln L_i$ is the log likelihood of the data as predicted by Model i after its free parameters have been estimated via maximum likelihood. By including a term that penalizes a model for extra free parameters, one can make a comparison across models with different numbers of parameters. The model that provides the most accurate account of the data is the one with the smallest AIC. (See Sakamoto, Ishiguro, & Kitagawa, 1986, or Takane & Shibayama, 1992, for a more thorough discussion of the minimum AIC procedure.)

APPLICATION 1
NORMALLY DISTRIBUTED CATEGORIES

This section reports the results of fitting decision bound and exemplar models to the data from five experiments with normally distributed categories. In each experiment,

stimuli like those shown in Figure 1 were used. The dimensions of the rectangles (height and width) have been found to be integral (e.g., Garner, 1974; Wiener-Erlich, 1978), whereas the dimensions of the circles (size and orientation) have been found to be separable (Garner & Felfoldy, 1970; Shepard, 1964). In all experiments, each category was defined by a bivariate normal distribution. Such distributions can be described conveniently by their *contours of equal likelihood*, which are always circles or ellipses. Although the size of each contour is arbitrary, the shape and location conveys important category information. The center of each contour is always the category prototype (i.e., the mean, median, or mode), and the ratio formed by dividing the contour width along Dimension 1 by the width along Dimension 2 is equal to the ratio of the standard deviations along the two dimensions. Finally, the orientation of the principle axis of the elliptical contour provides information about the correlation between the dimensional values.

Data Sets 1 and 2:
Linear Optimal Decision Bound

The contours of equal likelihood that describe the categories of the first two data sets are shown in Figure 2. Note that, in both experiments, variability within each category is equal on the two dimensions, and the values on the two dimensions are uncorrelated. In both experiments, the optimal stimulus bound (represented by the broken line in Figure 2) is linear ($v_2 = v_1$ and $v_2 = 450 - v_1$, for the first and second experiments, respectively, where v_1 corresponds to the width or size dimension, and v_2 corresponds to the height or orientation dimension). A subject perfectly using the optimal bound would correctly classify 80% of the stimuli in each experiment.[7]

Six subjects participated in each experiment, 3 with the rectangular stimuli (Condition R; see Figure 1a) and 3 with the circular stimuli (Condition C; see Figure 1b). At the beginning of every experimental session, each subject was shown the stimulus corresponding to the Category *A* and Category *B* distribution means, along with their category labels. Each of these stimuli were presented 5 times each in an alternating fashion, for a total of 10 stimulus presentations. This was followed by 100 trials of practice and then 300 experimental trials. Feedback was given after every trial. Only the experimental trials were included in the subsequent data analyses. The exact experimental methods were identical to those described by Ashby and Maddox (1992, Experiment 3). All subjects using the circular stimuli completed three experimental sessions. For the rectangular stimuli, 2 subjects from Experiment 1 completed two sessions and 1 subject completed one session, whereas the 3 subjects from Experiment 2 completed four, two, and three sessions, respectively. All subjects achieved at least 75% correct during their last experimental session.

Three decision bound models were fit to the data from each subject's last experimental session: the general linear classifier, the optimal decision bound model, and the gen-
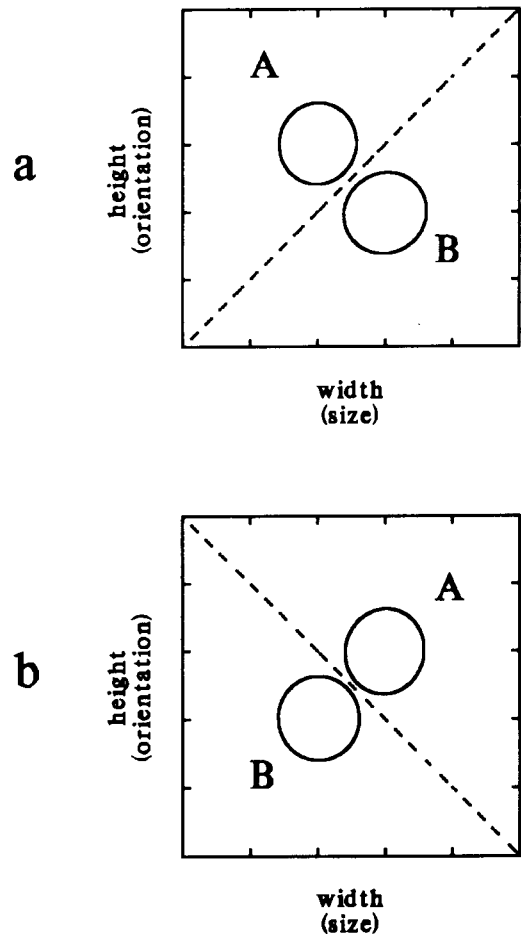
Figure 2. Contours of equal likelihood and optimal decision bounds (broken lines) for Application 1: (a) Data Set 1 and (b) Data Set 2.

eral quadratic classifier. Each model assumed that the amount of perceptual noise did not differ across stimuli or stimulus dimensions. In both experiments, exemplars from each category were presented equally often, so there was no a priori reason to expect a response bias toward either category. Thus, the response bias was set to zero in all three models. The general linear classifier has three free parameters: a slope and intercept parameter that describe the decision bound, and one parameter that represents the combined effect of perceptual and criterial noise.[8] Because the optimal decision bound model uses the decision bound of the optimal classifier, the slope and intercept are constrained by the shape and orientation of the category contours of equal likelihood; thus, the model has only one free parameter, which represents the sum of the perceptual and criterial noise (see Note 8). The general quadratic classifier has seven parameters: five of the six parameters in Equation 5 (one can be set to 1.0 without loss of generality), a perceptual noise parameter, and a criterial noise parameter.

In both experiments, the optimal classifier uses a linear decision bound, and thus decision bound theory predicts

that subjects will use some linear decision bound in both experiments. If the theory is correct, the extra parameters of the general quadratic classifier should lead to no significant improvement in fit. Goodness-of-fit tests (G^2) strongly supported this prediction. For 10 of the 12 subjects, the three-parameter general linear classifier and the seven-parameter general quadratic classifier yielded identical G^2 values. For the other 2 subjects (both from Experiment 2), the general quadratic classifier provided a slightly better absolute fit to the data, but in both cases, the improvement in fit was not statistically significant ($p > .25$). In order to determine the *absolute* goodness-of-fit of the general linear classifier, G^2 tests were performed between the general linear classifier and the "null" model (i.e., the model that perfectly accounts for the data). For all 12 subjects, the G^2 tests were not statistically significant ($p > .25$). (In other words, the null hypothesis that the general linear classifier is correct could not be rejected.) In addition, for 9 of the 12 subjects, the fits of the general linear classifier were significantly better ($p < .05$) than those for the optimal decision bound model. Thus, these analyses strongly support the hypothesis that subjects used a suboptimal decision bound that was of the same form as the optimal bound (in this case, linear).

The DEM has four parameters in this application: a scaling parameter c; a bias parameter β and attention parameter w, and the response selection parameter γ (see Equations 7 and 12). The GCM has only the first three of these free parameters ($\gamma = 1$ in the GCM). Three versions of each model were tested. One version assumed city-block distance and an exponential decay similarity function; one version assumed Euclidean distance and a Gaussian similarity function; and the third assumed Euclidean distance and an exponential similarity function.

Krantz and Tversky (1975) suggested that the perceptual dimensions of rectangles may be shape and area rather than height and width. A transformation from the dimensions of height and width to shape and area is accomplished by rotating the height and width dimensions by 45°. In order to test the shape-area hypothesis, versions of the GCM and DEM that included an additional free parameter corresponding to the degree of rotation were also applied to the data.[9] The resulting models, which we call the GCM(θ) and DEM(θ), were fitted to the data by using the three distance-similarity function combinations described above. The details of the fitting procedure are described in the Appendix.

Of the three distance-similarity function combinations tested, the Euclidean-Gaussian version of the GCM and GCM(θ) provided the best account of the data for both the rectangular (8 of 12 cases) and circular (10 of 12 cases) stimuli. The Euclidean-Gaussian version of the DEM and DEM(θ) provided the best account of the data for the circular stimuli (11 of 12 cases), and the Euclidean-exponential version fitted best for the rectangular stimuli (9 of 12 cases).

Table 1 presents the goodness-of-fit values for the general linear classifier and for the best-fitting versions of

Table 1
Goodness-of-Fit Values (AIC) for Application 1: Data Sets 1 and 2

Condition/Subject	GLC	GCM	DEM	GCM(θ)	DEM(θ)
			Data Set 1		
R/1	120.8	153.8	119.9	143.5	118.9
R/2	119.5	177.3	135.3	159.5	127.9
R/3	69.1	138.0	70.7	131.2	72.8
Mean	103.1	156.4	108.6	144.7	106.5
C/1	167.5	195.3	169.3	187.0	171.7
C/2	251.1	250.0	245.1	249.0	246.0
C/3	179.3	199.3	181.9	196.5	183.1
Mean	199.3	214.9	198.8	210.8	200.3
			Data Set 2		
R/1	129.0	155.0	133.6	154.7	131.5
R/2	191.8	208.5	195.0	194.7	197.1
R/3	221.2	226.9	223.2	224.2	220.7
Mean	180.7	196.8	183.9	191.2	183.1
C/1	172.1	195.8	176.3	182.6	178.8
C/2	218.4	229.2	219.7	225.8	221.0
C/3	248.2	248.0	250.2	250.7	252.7
Mean	212.9	224.3	215.4	219.7	217.5
Mean*	174.0	198.1	176.7	191.6	176.9

Note—Rows correspond to subjects and columns to models. GLC, general linear classifier; GCM, generalized context model; DEM, deterministic exemplar model; GCM(θ), GCM with additional θ parameter; DEM(θ), DEM with additional θ parameter. *Across 12 subjects.

the GCM, DEM, GCM(θ), and DEM(θ). The DEM [or DEM(θ)] provides the best fit for 3 subjects, and the GCM provides the best fit for 1 subject. For the remaining 8 subjects, the general linear classifier provides the best fit. Note, however, that the general linear classifier performs only slightly better than the DEM or DEM(θ).

In general, the fits of the GCM [and GCM(θ)] are worse than those for the DEM [and DEM(θ)]. In fact, the DEM fits better than the GCM(θ) for 11 of the 12 subjects, suggesting that the GCM is more improved by the addition of the γ parameter (a parameter associated with the decision process) than by the addition of the θ parameter (a parameter associated with the perceptual process). The poor performance of the GCM [and GCM(θ)] apparently occurs because response selection was less variable than predicted by these models; a fact that is reflected in the estimates of the DEM's γ parameter. Table 2 presents the median γ estimates for the best-fitting DEM and DEM(θ) from Table 1. As predicted, in every case the median $\gamma \geq 1$. The γ values are quite large for Data Set 1, espe-

Table 2
Median γ Estimates for Best-Fitting DEM and DEM(θ)
Reported in Table 1

Stimuli	Data Set 1		Data Set 2	
	DEM	DEM(θ)	DEM	DEM(θ)
Rectangles	5.12	2.59	2.13	1.00
Circles	2.27	2.46	1.40	1.49

Note—Rows correspond to stimuli and columns to models. DEM, deterministic exemplar model; DEM(θ), DEM with additional θ parameter.

cially for the rectangular stimuli. As one might expect, given this result, the difference in goodness of fit between the DEM and GCM(θ) is also quite large in Data Set 1.

Two important conclusions can be drawn from these data. First, the superior fits of the general linear classifier and the DEM over the GCM suggest that response selection is less variable than that predicted by the GCM, especially in Data Set 1. Second, when the optimal decision bound is linear, it is difficult to distinguish between the performance of a deterministic exemplar model and a decision bound model that assumes subjects use linear decision bounds in the presence of perceptual and criterial noise.

Data Sets 3–5:
Nonlinear Optimal Decision Bound

If the amount of variability along any dimension *or* if the amount of covariation between any pair of dimensions differs for the two categories, then the optimal bound will be nonlinear (Ashby & Gott, 1988; Morrison, 1990). To examine the ability of subjects to learn nonlinear decision bounds, Ashby and Maddox (1992) conducted three experiments using normally distributed categories. Both experiments involved two conditions: one with the rectangular stimuli (Condition R; see Figure 1a) and one with the circular stimuli (Condition C; see Figure 1b). Four subjects participated in each condition. Each subject completed between three and five experimental sessions.

The contours of equal likelihood used in Data Sets 3–5 (Ashby & Maddox, 1992, Experiments 1–3) are shown in Figures 3a–3c, respectively. Because the shape and orientation of the Category A and B contours of equal likelihood differ, the optimal bound is highly nonlinear (represented by the broken line in Figures 3a–3c). A subject perfectly using this bound would correctly classify 90%, 78%, and 90% of the stimuli in Data Sets 3–5, respectively. In contrast, a subject perfectly using the most accurate linear bound would correctly classify 65%, 60%, and 75% of the stimuli in the three experiments, respectively. There were large individual differences in accuracy during the final session, but, for 23 of 24 subjects, accuracy exceeded that predicted by the most accurate linear bound. Accuracy ranged from 68% to 82%, 59% to 73%, and 81% to 91%, for Data Sets 3–5, respectively. Details of the experimental procedure and summary statistics are presented in Ashby and Maddox (1992) and will not be repeated here.

The three decision bound models described in the last section were fitted separately to the data from each subject's first and last experimental sessions. Decision bound theory predicts that the best-fitting model should be the general quadratic classifier because the optimal bounds are all quadratic.

With normally distributed categories, the optimal decision bound model and the general linear classifier are a special case of the general quadratic classifier, so G^2 tests were performed to determine whether the extra parameters of the general quadratic classifier led to a significant improvement in fit. These results, which are re-
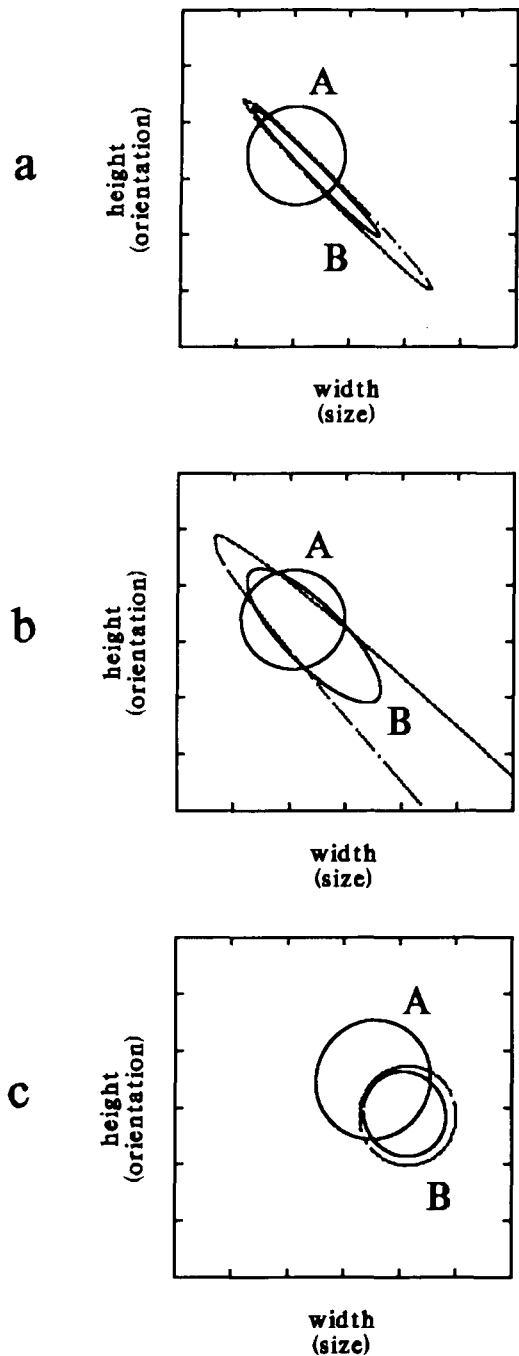
Figure 3. Contours of equal likelihood and optimal decision bounds (broken lines) for Application 1: (a) Data Set 3, (b) Data Set 4, and (c) Data Set 5.

ported in Table 3, confirm the superiority of the general quadratic classifier with these data.

There is evidence that 3 subjects in Data Set 3 used linear bounds during their first session, but, for the data from the last session, the general linear classifier is rejected in every case. The superiority of the general quadratic classifier over the optimal decision bound model

Table 3
Proportion of Times That the General Quadratic Classifier Fits Significantly Better ($p < .05$) Than the General Linear Classifier or Optimal Decision Bound Model

Data Set	General Linear Classifier		Optimal Decision Bound Model	
	First Session	Last Session	First Session	Last Session
3	5/8	8/8	7/8	8/8
4	8/8	8/8	8/8	8/8
5	8/8	8/8	7/8	6/8

suggests that, except for 2 subjects in Data Set 5, the subjects did not use optimal bounds.

Over the three data sets, the best fit clearly is provided by the general quadratic classifier. In fact, the null hypothesis that this is the correct model could not be rejected for any of the subjects in their last session. For the first session, the model was rejected only for 3 subjects in

Data Set 3 ($p < .05$). Thus, in addition to being the best of the three decision bound models, the general quadratic classifier also provides an adequate absolute account of the data. These results support the hypothesis that subjects use a suboptimal decision bound of the same form as the optimal bound (in this case, quadratic).

For the GCM [and GCM(θ)], the Euclidean–exponential version provided the best fit for 38% and 39% of the subjects who classified the rectangular and circular stimuli, respectively. The Euclidean–exponential version of the DEM [and DEM(θ)] also provided the best fit for 38% and 44% of the subjects who classified the rectangular and circular stimuli, respectively.

Table 4 compares the goodness-of-fit values for the general quadratic classifier, and the Euclidean–exponential version of the GCM, DEM, GCM(θ), and DEM(θ). For the data from the first session, the general quadratic classifier fits best in 16 of 24 cases, the DEM [or DEM(θ)]

Table 4
Goodness-of-Fit Values (AIC) for Application 1: Data Sets 3–5

C/S	GQC		GCM		DEM		GCM(θ)		DEM(θ)	
	First	Last	First	Last	First	Last	First	Last	First	Last
					Data Set 3					
R/1	291.6	326.3	344.5	327.6	337.3	329.4	344.1	323.3	325.2	325.7
R/2	271.6	306.3	296.7	312.4	295.5	314.0	288.5	311.3	291.1	307.3
R/3	365.3	199.9	367.2	206.5	368.4	208.4	361.8	197.2	358.6	198.9
R/4	285.2	255.9	352.2	268.4	312.9	269.1	353.9	261.1	302.6	260.7
Mean	303.4	272.1	340.2	278.7	328.5	280.2	337.1	273.2	319.4	273.2
C/1	410.1	273.3	402.4	295.9	404.4	284.1	404.4	296.3	405.6	285.2
C/2	312.3	216.2	370.0	212.1	371.8	213.2	360.8	213.2	363.6	215.1
C/3	326.2	217.9	322.4	246.3	319.8	236.8	324.4	238.1	321.6	231.2
C/4	386.2	273.8	405.8	280.9	402.4	279.1	407.4	281.8	403.3	280.8
Mean	358.7	245.3	375.2	258.8	374.6	253.3	374.3	257.4	373.5	253.1
					Data Set 4					
R/1	326.1	290.7	384.1	304.3	370.6	289.9	384.0	306.5	365.6	287.7
R/2	294.6	297.9	340.6	325.9	320.1	325.6	331.2	327.2	302.1	319.1
R/3	246.8	259.6	358.0	275.7	327.6	275.5	345.8	276.5	331.2	275.4
R/4	312.0	336.8	349.0	362.9	345.4	340.4	346.0	365.2	337.5	342.3
Mean	294.9	296.3	357.9	317.2	340.9	307.9	351.8	318.9	334.1	306.1
C/1	257.8	212.2	361.1	338.6	299.1	229.7	359.0	341.3	291.0	217.6
C/2	328.4	207.7	379.0	286.2	371.8	246.3	381.0	288.3	358.6	239.9
C/3	242.2	215.2	316.3	282.2	293.0	227.0	318.8	284.3	266.4	232.1
C/4	185.1	224.5	317.9	341.0	178.1	331.2	321.2	342.8	180.3	309.5
Mean	253.4	214.9	343.6	312.0	285.5	258.6	345.0	314.2	274.1	249.8
					Data Set 5					
R/1	203.2	112.5	209.8	119.4	207.9	110.4	182.4	116.3	184.2	94.1
R/2	257.1	114.2	241.2	161.9	243.2	146.1	236.4	148.4	226.7	146.7
R/3	210.0	133.9	204.2	165.5	194.9	160.6	184.6	144.1	187.1	132.2
R/4	189.7	219.1	205.4	207.3	206.1	206.5	205.4	208.4	206.9	207.7
Mean	215.0	144.9	215.2	163.5	213.0	155.9	202.2	154.3	201.2	145.2
C/1	190.9	128.8	213.7	144.5	205.9	143.9	216.2	147.6	207.5	145.8
C/2	113.5	57.5	93.7	51.3	95.5	42.5	91.4	52.6	92.8	41.3
C/3	63.0	122.3	75.3	152.1	75.9	127.5	73.8	136.9	77.0	124.1
C/4	91.6	79.4	100.9	88.9	102.3	84.4	97.6	91.5	97.1	87.2
Mean	114.8	97.0	120.9	109.2	119.9	99.6	119.8	107.2	118.6	99.6
Mean*	256.7	211.8	292.2	239.9	277.1	225.9	288.4	237.5	270.2	221.2

Note—Rows correspond to subjects and columns to models. GQC, general quadratic classifier; GMC, generalized context model; DEM, deterministic exemplar model; GCM (θ), GCM with additional θ parameter; DEM(θ), DEM with additional θ parameter. *24 subjects

fits best in 4 cases, and the GCM [or GCM(θ)] fits best in 4 cases. For the data from the last session, the general quadratic classifier fits best in 16 of 24 cases, the DEM [or DEM(θ)] fits best in 5 cases, and the GCM [or GCM(θ)] fits best in 3 cases.

In Data Sets 1 and 2, where the optimal decision bound was linear, the general linear classifier fits only slightly better than the DEM [or DEM(θ)]. In fact, the average goodness-of-fit value of the general linear classifier was less than 3 AIC points better than the average goodness-of-fit value of the DEM. However, in Data Sets 3–5, where the optimal decision bound is nonlinear, the decision bound model enjoys a clear advantage over the DEM and DEM(θ). For the data from the first experimental session, the average goodness-of-fit value of the general quadratic classifier is more than 20 AIC points better than the average value of the DEM [and about 14 AIC points better than the DEM(θ)]. For the final experimental session, the general quadratic classifier betters the DEM by about 14 AIC points and the DEM(θ) by about 10 AIC points.[10]

As with Data Sets 1 and 2, the GCM and GCM(θ) fits are worse than the DEM [and DEM(θ)] fits. However, a closer examination reveals that the largest discrepancy occurs in Data Set 4. One possible explanation for this result is that response selection is less variable in Data Set 4 than in Data Sets 3 and 5. Since larger values of γ are associated with less variability in response selection, this hypothesis predicts that the γ values for Data Set 4 should be larger than those for Data Sets 3 and 5. Table 5 presents the median γ estimates for the best-fitting DEM and DEM(θ) from Table 4. As predicted, the largest median γ values occurred in Data Set 4. Note also that in 9 of the 12 cases reported in Table 5, the median γ value increased from the first to the last experimental session.

Several important conclusions can be drawn from the quantitative analysis of Application 1. First, the GCM was consistently outperformed by the DEM and by the decision bound model. Much of this disparity can be attributed to the fact that when the optimal decision bound was linear, response selection was less variable than predicted by the GCM. When the optimal decision bound was nonlinear, both the GCM and the DEM were outperformed by the decision bound model.

Table 5
Median γ Estimates for Best-Fitting DEM and DEM(θ)
Reported in Table 4

Stimuli	Data Set	DEM		DEM (θ)	
		First	Last	First	Last
Rectangles	3	1.59	1.07	1.25	1.51
	4	2.71	2.19	2.62	2.80
	5	1.14	1.46	1.13	1.90
Circles	3	1.85	1.61	1.60	2.07
	4	3.45	4.48	4.29	5.67
	5	1.03	2.12	1.30	1.65

Note—Rows correspond to stimuli and columns to models. DEM, deterministic exemplar model; DEM(θ), DEM with additional θ parameter.

One might ask why the DEM performed more poorly than the decision bound model when the optimal decision bound was nonlinear. One way to answer this question is to compare their respective $P(A \mid x) = .5$ contours [i.e., the set of all x for which $P(A) = .5$], under the assumption that no response bias exists. The points that make up this contour favor the two response alternatives equally, in other words, they are equivocal with respect to category membership. As a result, we call such a contour the *equivocality contour* (Ashby & Maddox, in press). In the decision bound model, the equivocality contour is the unbiased decision bound. In the exemplar model, the equivocality contour is the set of coordinates for which summed similarity to the two categories is equal (i.e., the set of all x satisfying $\sum \eta_{xA} = \sum \eta_{xB}$). If the equivocality contours for the decision bound model and the exemplar model agree, then the performance of the models should be similar.

As throughout this article, assume that the amount of perceptual noise is constant across all perceptual dimensions and is uncorrelated. Ashby and Maddox (in press; see also Nosofsky, 1990) showed that, under these conditions, there exist parameter values that allow the Euclidean–Gaussian exemplar models to mimic exactly the equivocality contour of the optimal decision bound model (in the perceptual space). Thus, if subjects respond optimally, it should be very difficult to discriminate between decision bound and exemplar models.

Although the equivocality contours for the optimal decision bound model and the Euclidean–Gaussian exemplar models are identical in the case of independent perceptual noise, the decision bound and exemplar models treat suboptimality differently. The exemplar models stress the importance of selective attention (i.e., the stretching and shrinking of the perceptual dimensions), which indirectly affects the equivocality contour. The decision bound model assumes that the subject operates on the decision bound directly. When the optimal bound is linear, as in Data Sets 1 and 2, manipulating attention is essentially equivalent to changing the decision bound slope and intercept. However, if the optimal bound is nonlinear, the effects of manipulating attention will be limited. In this case, the general quadratic classifier is more powerful than the exemplar models. Of course, if the direct action of the subject is one of selective attention, the extra power of the general quadratic classifier (in the form of extra free parameters) is wasted. The success of the general quadratic classifier in Data Sets 3–5 supports the hypothesis that the decision bound is a fundamental construct of human categorization.

A second conclusion to be drawn from Application 1 is that the decision bound model consistently outperformed the GCM and DEM. Of course, these results do not falsify exemplar theory. Although we tested an important class of exemplar models, other versions may have been more successful. We can conclude, however, that the decision bound models provide a viable alternative to the exemplar-similarity-based models of categorization.

One weakness of the present analysis is that some rather strong assumptions were needed about the mapping from the stimulus space to the perceptual space. When fitting the GCM, it is common practice to first collect similarity judgments or identification responses on the stimulus ensemble and then to subject the data to some sort of multidimensional scaling (MDS) analysis. The coordinates of the stimuli from the MDS solution are then assumed to estimate the coordinates of the stimuli in the perceptual space of the categorization task (e.g., Nosofsky, 1986).[11] This approach is untenable with normally distributed categories, owing to the unlimited number of exemplars. To test the decision bound model more completely, it is of interest to apply the model to data where this sort of MDS-based analysis was performed. Data of this sort are examined in Application 2, to which we turn now.

APPLICATION 2
NONNORMALLY DISTRIBUTED CATEGORIES

This section compares the performance of the decision bound model with that of the GCM and DEM at predicting categorization data from experiments in which the category exemplars are not normally distributed and contain only a small number of exemplars. Two data sets collected by Nosofsky (1986, 1989) were chosen. In both cases, the complete stimulus ensemble consisted of 16 circles of the type shown in Figure 1 (constructed by combining factorially 4 levels of each dimension). All categorization conditions involved two categories with four exemplars each. Thus, the training ensemble contained 8 stimuli. Feedback was given on each trial. After categorization accuracy reached a criterion level, the ensemble was enlarged to include all 16 stimuli (the additional 8 stimuli were termed "transfer" stimuli by Nosofsky, 1986).

In the Nosofsky (1986) experiment, 2 highly practiced subjects participated in a large number of identification sessions, followed by several sessions of categorization. Because decision bound theory is a theory of the performance of *individual* subjects, these data are highly appropriate as a test of the theory. The second data set (Nosofsky, 1989), which consists of data averaged across a large number of inexperienced subjects, is less appropriate for testing decision bound theory, but will serve to test the theory's generalizability.

In the Nosofsky experiments (1986, 1989), decision bound theory predicts that the optimal decision bound is neither a linear nor a quadratic function of the dimensional values. Even so, it seems reasonable to assume that subjects might use linear or quadratic bounds in these experiments. This is because it has been hypothesized that the multivariate normal distribution provides a good model of many natural categories (Ashby, 1992a; see also Fried & Holyoak, 1984), and with normally distributed categories, the optimal decision bound is always linear or quadratic. Thus, if humans frequently categorize at near-optimal levels, they will have much experience with linear and quadratic bounds. If so, it makes sense that they would use these bounds when confronted with the artificial categories constructed by Nosofsky (1986, 1989). We begin by describing the general method used in fitting the various models and then turn to the model comparisons.

General Method

Application of the exemplar models is straightforward. The MDS solution derived from the identification conditions of Nosofsky (1986, 1989) will be used in conjunction with Equations 6, 7, 9, and 12 to generate predicted response probabilities. Following Nosofsky (1986, 1989), the Euclidean distance metric and Gaussian similarity function are assumed. Two augmented versions of the GCM and DEM were applied to the data as well. The first, proposed by Ashby and Maddox (in press), allowed for oblique perceptual dimensions. The second allowed the scaling constant c (from Equation 7) to differ for training and transfer stimuli. One hypothesis is that experience with category exemplars increases their perceptual dissimilarity (Nosofsky, 1986). If so, c should be larger for training than for transfer stimuli. In many cases, the goodness-of-fit values for these two models were worse than those for the standard GCM and DEM. When the goodness of fit was improved, however, inclusion of these models never affected the qualitative results (see Tables 6 and 9), so they will not be discussed further.

When applying the decision bound model, the MDS coordinates Nosofsky (1986, 1989) obtained from the identification confusions were used as estimates of the perceptual means (i.e., x_i from Equation 1). As in Application 1, perceptual variability was assumed to be constant across dimensions, and to be uncorrelated. Thus, only one perceptual variance parameter was estimated. This is the simplest perceptual representation allowed in decision bound theory, and in light of the results of Ashby and Lee (1991; see Figure 6, p. 161), it is surely incorrect. We chose this perceptual representation for two reasons. First, Ashby and Perrin (1988) showed that these distributional assumptions produce a dissimilarity metric that is equivalent to the measure used by the GCM and DEM when equal amounts of attention are allocated to each stimulus dimension (i.e., when $w = .5$; see Equation 7). Nosofsky (1986, 1989) argued that it is necessary to incorporate selective attention components (at least within the framework of the exemplar–similarity model) in order to predict data from several of the "dimensional" (i.e., size and angle) categorization conditions (Nosofsky, 1986, 1989; see Figures 4–6 in the present paper). This selective attention manifests itself as a stretching of distance relations along the attended dimension, and a shrinking of distance relations along the unattended dimension. Ashby and Lee (1991, 1992) argue that data in these "dimensional" categorization conditions can be accounted for equally well (at least within a framework of the decision bound model) without postulating any stretching or shrinking of distance relations (i.e., selective attention), but rather by acknowledging the different decision bounds required for identification and categorization. The fact that

the decision bound model we propose predicts identical similarity relations as the GCM with no selective attention allows a test of this hypothesis. Second, such a simple perceptual representation forces the decision bound to account for most of the variance in the data. Therefore, this is a good method for testing the hypothesis that the decision bound is of fundamental importance in predicting asymptotic categorization performance.

Data Set 6: Nosofsky (1986)

The stimulus dimensions used by Nosofsky (1986) were the same as those shown in Figure 1b, except that only the upper half of the circle and radial line were presented (see Nosofsky, 1989, Figure 2, for an example). Sixteen stimuli were constructed from the factorial combination of 4 levels of circle size and 4 levels of orientation of the radial line. Two subjects participated in a large number of identification sessions, followed by several sessions of categorization.

In the categorization conditions, 4 stimuli were assigned to Category 1 and a different 4 were assigned to Cate-

gory 2. During the training phase of the experiment, any of these 8 stimuli could appear on a given trial and corrective feedback was provided following the subject's response. During the transfer phase of the experiment, all 16 stimuli were included, and subjects were given corrective feedback only when a training exemplar was presented. The data of interest are those collected during the transfer phase only.

The following four categorization conditions, illustrated in Figure 4 (ignore the line or curve, which will be discussed later), were utilized. (1) *Size*: Category 1 exemplars were small in size, whereas Category 2 exemplars were large in size. (2) *Criss-cross*: Category 1 contained exemplars with large size/small angle or small size/large angle, whereas Category 2 contained exemplars with large size/large angle or small size/small angle. (3) *Interior-exterior*: Category 1 contained exemplars with intermediate size and angles, whereas Category 2 contained exemplars with extreme values on each dimension. (4) *Diagonal*: Category 1 contained exemplars that fell below a line with a slope of approximately −1, whereas Cate-

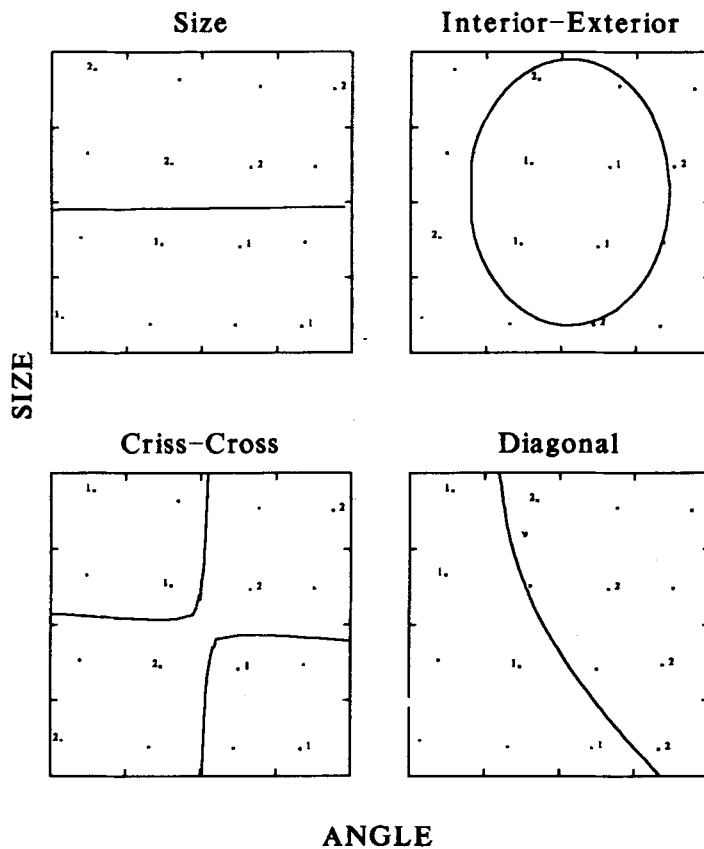## Nosofsky(1986) Categorization Conditions, Subject 1

Figure 4. MDS coordinates (Euclidean distance) for Nosofsky (1986), Subject 1. Labeled exemplars were presented during training *and* transfer for each of the four categorization conditions. The line or curve denotes the decision bound predicted by the best-fitting decision bound model.

Table 6
Goodness-of-Fit Values (AIC) for Application 2, Data Set 6 (Nosofsky, 1986)

	Subject 1				Subject 2			
	Size	Criss-Cross	Interior–Exterior	Diagonal	Size	Criss-Cross	Interior–Exterior	Diagonal
Exemplar Models								
GCM	70.4	208.3	247.7	131.9	102.6	275.8	231.3	130.3
DEM	71.9	183.6	185.7	129.4	104.8	277.3	224.2	118.4
Decision Bound Models								
GQC	76.7	112.3	135.6	110.5	102.6	119.2	111.6	105.8
GLC	73.4	1,925.0	971.0	146.2	97.6	1,237.0	614.0	103.4

Note—Rows correspond to models and columns to subjects and conditions. GCM, generalized context model; DEM, deterministic exemplar model; GQC, general quadratic classifier; GLC, general linear classifier.

gory 2 contained exemplars that fell above the line. In Figure 4, each training exemplar is labeled with a 1 or 2, depending on its category membership.

Model Fits and Comparisons

Table 6 presents the goodness-of-fit values for the GCM, DEM, general quadratic classifier, and general linear classifier for each subject and categorization condition. The GCM provides the best fit for Subject 1 in the dimensional condition, but in the seven other applications, a decision bound model provides the best account of the data. In the criss-cross, interior–exterior, and diagonal conditions, the decision bound model performed substantially better than the GCM or DEM. In fact, across both subjects, the average fit value in these three conditions was 115.4, 204.2, and 186.4 for the decision bound model, the GCM, and the DEM, respectively.[12]

Table 7 presents the observed probability of responding "1" for each of the 16 stimuli by subject and condition. In addition, the predicted probability of responding "1" is presented below the observed response probabilities for the best-fitting decision bound model, GCM, and DEM, respectively. Figures 4 and 5 present the MDS coordinates for each stimulus and the decision bound predicted by the best-fitting decision bound model for each subject in each categorization condition.

The predictions of the three models are most dissimilar for the criss-cross and interior–exterior conditions, so we will examine these conditions in greater detail. For purposes of elaboration, the stimuli can be numbered from 1 to 16. The numbering scheme for the stimuli is presented in Table 8.

First, consider the interior–exterior condition. Stimuli 3, 5, 12, and 14 are all exemplars of Category 2, and they all have approximately the same similarity relations to the exemplars of Category 1. Therefore, exemplar models predict that accuracy should be nearly equal for these stimuli, a prediction that is not supported by the data. The average accuracy for Stimuli 3, 5, and 12 was 72%, but for Stimulus 14 it was only 50%. As a consequence, the exemplar models failed badly for Stimulus 14. The GCM predicted an average accuracy to Stimulus 14 of 72%, and

the DEM predicted 67%. In contrast, the decision bound model successfully predicted the low accuracy to Stimulus 14. (The decision bound model predicted an average accuracy of 51%.) It did this by assuming that the decision bound passed close to the mean of the Stimulus 14 perceptual distribution.

Next, consider the criss-cross condition. For each of the transfer stimuli, the two nearest training exemplars are always from the same category, and the average observed probability with which the subjects assigned these stimuli to the same category as these nearest neighbors was .755. The exemplar models can account for these high observed probabilities, but if they do, they must predict that accuracy is near chance for training Stimuli 6, 7, 10, and 11. This is because these stimuli have the property that the two nearest training exemplars are from the contrasting category. Exemplar models predict that if the response probabilities of transfer stimuli are dominated by the category membership of their nearest neighbors, it must make the same predictions for training stimuli. In fact, average accuracy for responses to Stimuli 6, 7, 10, and 11 was 67.5% but the GCM predicted an average accuracy of only 55.5%. In contrast, the decision bound model predicted an average accuracy for responses to Stimuli 6, 7, 10, and 11 of 65.5%.

These results agree with those from Application 1. When the best-fitting decision bound was linear (or nearly linear), the goodness-of-fit difference between the exemplar and decision bound models was small. However, when the best-fitting decision bound was highly nonlinear (as in the criss-cross and interior–exterior conditions), the decision bound model fitted substantially better than the exemplar models.

There is good evidence that both subjects responded suboptimally in this experiment.[13] Suppose that they used a suboptimal decision bound. The only way the exemplar models can account for this fact is by uniformly expanding or contracting the space (by manipulating the parameter c), by stretching or shrinking one of the perceptual dimensions (by manipulating the attention weight w), or by changing the intercept of the $P(A) = .5$ contour (by manipulating the response bias β). If the subject's bound

Table 7
Observed and Predicted Probability of Responding "1" by Subject and
Categorization Condition for Nosofsky (1986)

	Subject 1				Subject 2			
Stimulus	Size	Criss-Cross	Interior–Exterior	Diagonal	Size	Criss-Cross	Interior–Exterior	Diagonal
1	.982	.033	.074	1.000	.966	.136	.096	.986
	.989	.032	.066	.994	.982	.105	.112	.994
	.994	.016	.111	.993	.985	.082	.150	.992
	.993	.018	.076	.994	.986	.081	.120	.993
2	.996	.126	.325	.920	.979	.371	.372	.930
	.992	.131	.288	.944	.969	.399	.288	.934
	.996	.242	.485	.918	.970	.359	.401	.936
	.995	.235	.459	.933	.971	.358	.390	.925
3	.990	.750	.296	.705	.960	.863	.318	.706
	.992	.766	.375	.711	.972	.798	.282	.644
	.996	.830	.309	.610	.972	.689	.294	.649
	.995	.816	.410	.624	.973	.686	.331	.613
4	.995	.958	.229	.354	.961	.932	.169	.174
	.993	.952	.178	.385	.964	.906	.189	.203
	.996	.971	.266	.296	.961	.853	.338	.220
	.996	.965	.227	.294	.962	.856	.344	.213
5	.765	.264	.143	.973	.852	.421	.264	.976
	.764	.293	.206	.970	.831	.416	.303	.984
	.759	.215	.199	.974	.824	.325	.210	.978
	.762	.218	.209	.968	.824	.324	.207	.984
6	.804	.326	.713	.755	.811	.407	.704	.843
	.815	.291	.663	.766	.797	.435	.737	.820
	.805	.401	.713	.833	.783	.504	.703	.814
	.808	.370	.728	.835	.783	.502	.709	.816
7	.824	.738	.712	.419	.734	.689	.618	.229
	.834	.710	.722	.360	.786	.680	.661	.304
	.821	.627	.658	.425	.765	.624	.702	.365
	.823	.645	.712	.440	.766	.621	.719	.322
8	.818	.799	.384	.172	.764	.732	.293	.061
	.812	.827	.394	.124	.710	.745	.310	.052
	.794	.733	.425	.143	.686	.640	.506	.073
	.797	.749	.410	.158	.688	.636	.509	.064
9	.110	.902	.241	.912	.330	.775	.424	.948
	.089	.885	.246	.903	.349	.797	.400	.964
	.100	.822	.321	.894	.392	.719	.330	.959
	.098	.829	.273	.882	.386	.717	.305	.966
10	.143	.697	.769	.444	.426	.526	.891	.693
	.134	.705	.746	.462	.367	.508	.887	.649
	.147	.614	.688	.528	.397	.601	.704	.577
	.145	.649	.723	.514	.394	.599	.715	.622
11	.140	.229	.695	.105	.297	.290	.731	.157
	.156	.303	.742	.134	.334	.332	.663	.126
	.167	.393	.652	.138	.354	.482	.635	.151
	.164	.367	.679	.150	.353	.477	.653	.127
12	.181	.300	.402	.047	.254	.338	.252	.025
	.155	.257	.362	.042	.248	.329	.261	.020
	.162	.221	.279	.033	.262	.330	.309	.027
	.159	.204	.301	.041	.262	.324	.316	.022
13	.000	.982	.119	.836	.045	.896	.274	.887
	.003	.968	.135	.814	.048	.899	.261	.884
	.001	.989	.218	.785	.047	.932	.227	.910
	.001	.988	.143	.781	.045	.935	.198	.917
14	.000	.772	.504	.271	.060	.547	.497	.344
	.005	.754	.470	.293	.056	.535	.514	.358
	.002	.785	.282	.331	.053	.669	.277	.280
	.002	.800	.354	.295	.051	.669	.304	.353

Table 7 (continued)

	Subject 1				Subject 2			
Stimulus	Size	Criss-Cross	Interior-Exterior	Diagonal	Size	Criss-Cross	Interior-Exterior	Diagonal
15	.000	.141	.425	.058	.035	.100	.252	.025
	.007	.139	.449	.067	.054	.112	.314	.045
	.003	.191	.421	.054	.047	.265	.359	.045
	.004	.197	.379	.045	.046	.258	.334	.039
16	.000	.057	.149	.012	.020	.076	.108	.012
	.007	.035	.144	.017	.024	.046	.093	.005
	.004	.021	.158	.007	.015	.049	.157	.005
	.004	.025	.113	.007	.014	.044	.124	.004

Note—Rows correspond to stimuli and columns to categorization conditions. Top row: observed probability of responding "1." Second row: predicted probability of responding "1" for the best-fitting decision bound model. Third row: predicted probability of responding "1" for the GCM. Bottom row: predicted probability of responding "1" for the DEM.

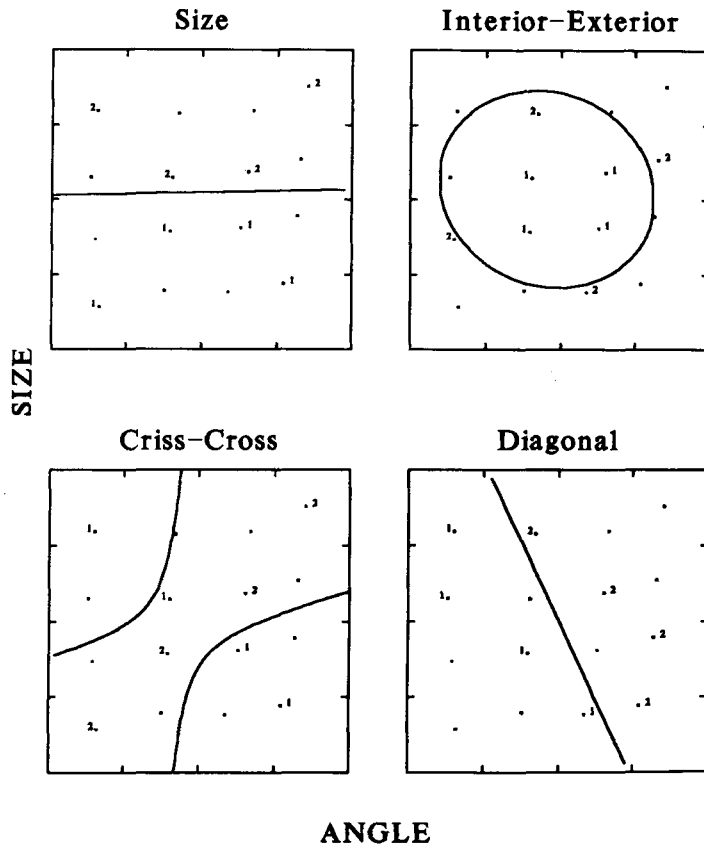## Nosofsky(1986) Categorization Conditions, Subject 2

Figure 5. MDS coordinates (Euclidean distance) for Nosofsky (1986), Subject 2. Labeled exemplars were presented during training and transfer for each of the four categorization conditions. The line or curve denotes the decision bound predicted by the best-fitting decision bound model.

Table 8
Numbering Scheme for Stimuli From Application 2,
Data Sets 6 and 7 (Nosofsky, 1986, 1989)

	Angle			
	13	14	15	16
Size	9	10	11	12
	5	6	7	8
	1	2	3	4

Note—Nosofsky (1986, 1989) for the specific size and angle values.

is linear, these transformations will be effective, but if the bound is highly nonlinear, these transformations will often be too crude.

Data Set 7: Nosofsky (1989)

Our final empirical application is to the categorization data collected by Nosofsky (1989). The stimulus dimensions used by Nosofsky (1989) were identical to those from the 1986 study, although the actual stimuli were somewhat more discriminable. The experimental procedure was the same, but with three important exceptions. First, a large number of subjects were run, and different

subjects participated in the identification and categorization conditions. Second, each subject received little training (only one experimental session). Finally, the data were collapsed across subjects. In addition, the interior–exterior condition was replaced with another dimensional condition in which angle was relevant (see Figure 6).

The fact that the categorization data were averaged across subjects causes problems for the decision bound models, because decision bound theory is a theory of *individual* categorization performance. The theory assumes that the experienced subject utilizes a fixed decision bound and that trial-by-trial fluctuations in performance are due to the effects of perceptual and criterial noise. If 2 subjects each use a linear bound with a different slope, the averaged data will be inconsistent with any linear (or quadratic) bound. In fact, the averaged data will be consistent with decision bound theory only in the special case in which each subject uses a bound that is a simple intercept shift of the others. In this case, the intercept shifts will be absorbed into the criterial and perceptual noise parameters. In the simpler categorization conditions, such as the dimensional (size and angle) conditions, it seems plausible that subjects will use bounds of the same shape

Nosofsky(1989) Categorization Conditions
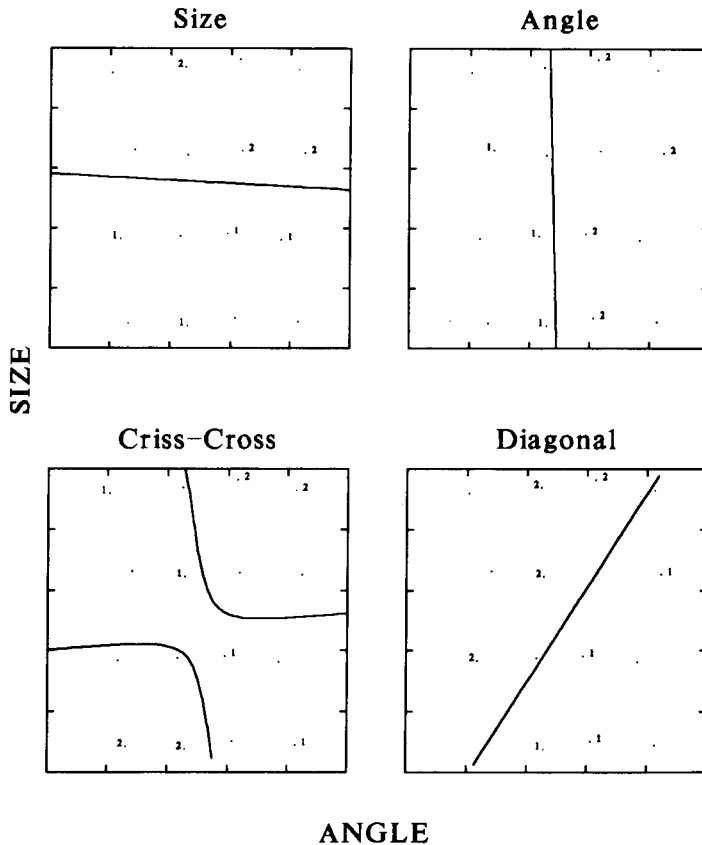
Figure 6. MDS coordinates (Euclidean distance) for Nosofsky (1989). Labeled exemplars were presented during training *and* transfer for each of the four categorization conditions. The line or curve denotes the decision bound predicted by the best-fitting decision bound model.

but with simple intercept shifts, and so the decision bound models should do better in these conditions than in the criss-cross and diagonal conditions. Certainly, though, we expect performance of the decision bound models to be poorer when the data are averaged across subjects.

Model Fits and Comparisons

Table 9 presents the goodness-of-fit values for the GCM, DEM, general quadratic classifier, and general linear classifier for each categorization condition. The GCM provides the best fit for the criss-cross and diagonal conditions, and the general linear classifier provides the best fit for the two dimensional (size and angle) conditions.

Table 10 presents the observed probability of responding "1" for each of the 16 stimuli by subject and condition. In addition, the predicted probability of responding "1" is presented (below the observed probabilities) for the best-fitting decision bound model (general quadratic classifier or general linear classifier), GCM, and DEM, respectively. Figure 6 presents the MDS coordinates for each stimulus and the best-fitting decision bound for each categorization condition.

The modeling results for Nosofsky (1989) differ in several important ways from the results for Nosofsky (1986). First, the DEM never fit the Nosofsky (1989) data better than the GCM. This result suggests that response selection was neither more nor less variable than predicted by the GCM. The γ values, which ranged from .98 for the angle condition to 1.28 for the criss-cross condition, support this hypothesis. In contrast, the DEM provided a better fit than did the GCM in several of the Nosofsky (1986) conditions (see Table 6). In each of these cases, the γ values differed considerably from $\gamma = 1$ (Subject 1, criss-cross, $\gamma = .21$, interior-exterior, $\gamma = 3.07$; Subject 2, interior-exterior, $\gamma = 1.57$, diagonal, $\gamma = 1.51$). In light of these results, it is likely that the Nosofsky (1989) results are due to the small amount of training given each subject, to the fact that the data were averaged across subjects, or to some combination of both.

Second, in each condition, the goodness-of-fit values for the best-fitting exemplar and decision bound models

Table 9
Goodness-of-Fit Values (AIC) for Application 2, Data Set 7 (Nosofsky, 1989)

	Size	Angle	Criss-Cross	Diagonal
Exemplar Models				
GCM	87.5	95.0	101.4	102.4
DEM	87.7	96.9	102.2	104.6
Decision Bound Models				
GQC	80.8	86.5	101.9	105.3
GLC	77.4	84.3	483.5	104.0

Note—Rows correspond to models and columns to conditions. GCM, generalized context model; DEM, deterministic exemplar model; GQC, general quadratic classifier; GLC, general linear classifier.

Table 10
Observed and Predicted Probability of Responding "1" by Subject and Categorization Condition for Nosofsky (1989)

Stimulus	Size	Angle	Criss-Cross	Diagonal
1	.973	.963	.222	.465
	.989	.943	.240	.497
	.992	.933	.219	.435
	.991	.937	.220	.434
2	.985	.572	.405	.775
	.989	.596	.397	.767
	.991	.554	.367	.782
	.990	.561	.371	.782
3	.973	.157	.608	.906
	.985	.174	.584	.884
	.986	.191	.551	.892
	.986	.193	.552	.892
4	.987	.024	.768	.965
	.985	.015	.783	.966
	.986	.014	.762	.961
	.985	.014	.755	.961
5	.870	.988	.460	.222
	.868	.955	.443	.219
	.865	.956	.400	.219
	.873	.959	.419	.219
6	.892	.662	.460	.570
	.848	.644	.502	.474
	.839	.608	.489	.514
	.848	.617	.496	.514
7	.842	.229	.575	.700
	.828	.210	.556	.688
	.814	.224	.555	.689
	.823	.227	.554	.689
8	.853	.024	.635	.814
	.842	.028	.626	.876
	.825	.031	.601	.852
	.831	.031	.606	.852
9	.311	.913	.716	.093
	.291	.915	.670	.107
	.313	.916	.632	.131
	.325	.921	.648	.131
10	.243	.573	.538	.213
	.307	.556	.555	.267
	.327	.532	.572	.286
	.339	.541	.572	.286
11	.244	.134	.446	.488
	.263	.119	.431	.503
	.282	.134	.474	.465
	.292	.138	.476	.466
12	.245	.015	.324	.748
	.257	.010	.344	.788
	.277	.007	.358	.777
	.285	.007	.367	.777
13	.027	.927	.776	.035
	.037	.963	.794	.072
	.026	.974	.808	.035
	.032	.976	.811	.035
14	.034	.573	.595	.100
	.028	.560	.524	.100
	.020	.542	.564	.098
	.024	.553	.565	.098
15	.041	.096	.283	.253
	.020	.118	.295	.230
	.014	.130	.329	.205
	.017	.135	.326	.206

Table 10 (Continued)

Stimulus	Size	Angle	Criss-Cross	Diagonal
16	.041	.024	.174	.640
	.025	.012	.172	.527
	.022	.009	.186	.559
	.026	.009	.181	.560

Note—Rows correspond to stimuli and columns to categorization conditions. Top row: observed probability of responding "1." Second row: predicted probability of responding "1" for the best-fitting decision bound model. Third row: predicted probability of responding "1" for the GCM. Bottom row: predicted probability of responding "1" for the DEM.

are very similar. The biggest difference occurs in the angle condition, where the goodness-of-fit value of the general linear classifier is 10.1 AIC points better than the goodness-of-fit value of the GCM. This difference is substantially smaller than the differences observed in the Nosofsky (1986) experiment, where the best and worst AIC values differed by as much as 100 or more points.

Third, the exemplar models fit the averaged data from the criss-cross and diagonal conditions (i.e., the Nosofsky, 1989 data) better than the single subject data from these same two conditions (i.e., the Nosofsky, 1986 data). For the decision bound models, however, the fits were similar for the averaged and single subject data sets. For the dimensional (size and angle) conditions, each of the models fit the averaged data about as well as the single subject data. This suggests that in complicated categorization tasks (such as the criss-cross and diagonal conditions), exemplar models may fit data averaged across subjects better than the data of any single subject.

Finally, the fact that the decision bound model accounted for the two dimensional (size and angle) categorization conditions better than the GCM or DEM supports the claim that, at least within the framework of a decision bound approach to categorization, shifts in the decision bound are more important than shifts in selective attention. However, it is possible, as suggested by Nosofsky (1989; see also Ashby & Lee, 1992; Nosofsky & Smith, 1992), that other decision bound models, which assume selective attention shifts (see Nosofsky, 1989, Figure 7, p. 288) might also provide good fits to these data.

SUMMARY AND CONCLUSIONS

The goal of this article was to develop and test a decision bound theory of categorization by: (1) applying the model to data from normal, as well as nonnormally distributed categories, and (2) comparing the performance of the model with a currently popular exemplar model of categorization. In Application 1, the models were fit to data from five categorization experiments (36 subjects total). In every experiment, the category exemplars were normally distributed and each subject was given extensive training. The performance of the decision bound model was compared with the currently popular and widely tested GCM (see Nosofsky, 1992, for an extensive review) and with a deterministic exemplar model (DEM; Ashby & Maddox, in press) that contains the GCM as a special case, but includes a parameter that allows it to predict data in which response selection is either more or less variable than what is predicted by the GCM. When the optimal decision bound was linear, the DEM and a decision bound model (the general linear classifier) provided nearly equivalent accounts of the data, and they both significantly outperformed the GCM. When the optimal decision bound was highly nonlinear, both the GCM and the DEM were outperformed by a decision bound model that postulated quadratic decision bounds. Taken together, these results suggest (1) that the poor performance of the GCM is due partly to the fact that it postulates a probabilistic rather than a deterministic decision rule, and (2) that the exemplar models have only a limited ability to account for suboptimal performance. In particular, they are inferior to the decision bound model at accounting for complex suboptimalities.

In Application 2, we began a preliminary investigation of the ability of the decision bound models to account for categorization data in which the exemplars from each category were *not* normally distributed. When applied to data in which single subjects were given extensive training (Nosofsky, 1986), the decision bound model provided excellent accounts of the data, and in many cases it significantly outperformed the GCM and DEM. When applied to data that were averaged across subjects, each of whom received little training, the performance of the decision bound model was still quite impressive, especially when applied to data from conditions in which only one stimulus dimension was relevant.

Future research could extend the present work in a number of directions. First, it would be interesting to expand the type of analysis presented in Application 2 to other categorization conditions and stimulus dimensions. Second, it would be of interest to examine the tradeoff between parameters associated with perceptual and decisional processes. Throughout this article, we assumed a very restricted perceptual representation. Although this was appropriate for the goals of the article and yielded good accounts of the data, it is possible that models with more general perceptual representations might provide better fits to the present data sets and might be necessary to predict data from other experiments, especially those involving highly confusable stimuli.

REFERENCES

AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716-723.

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

ASHBY, F. G. (1992b). Multivariate probability distributions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 1-34). Hillsdale, NJ: Erlbaum.

ASHBY, F. G., & GOTT, R. (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 between identification, 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. (1992). 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. (in press). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*.

ASHBY, F. G., & PERRIN, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, **95**, 124-15.

BARLOW, H. B., & MOLLON, J. D. (1982). Psychophysical measurements of visual performance. In H. B. Barlow & J. D. Mollon (Eds.), *The senses* (pp. 114-132). Cambridge: Cambridge University Press.

BUSEMEYER, J. R., & MYUNG, I. J. (1992). An adaptive approach to human decision making: Learning theory, decision theory, and human performance. *Journal of Experimental Psychology: General*, **121**, 177-194.

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

ENNIS, D. M., & ASHBY, F. G. (in press). The relative sensitivities of same-different and identification judgment models to perceptual dependence. *Psychometrika*.

ENNIS, D. M., MULLEN, K., & FRIJTERS, J. E. R. (1988). Variants of the method of triads: Unidimensional Thurstonian models. *British Journal of Mathematical & Statistical Psychology*, **41**, 25-36.

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

ESTES, W. K. (1986b). Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology: General*, **115**, 155-174.

FECHNER, G. T. (1866). *Elements of psychophysics*. New York: Holt, Rinehart & Winston.

FISHER, R. A. (1936). The use of multiple measurement in taxonomic problems. *Annals of Eugenics*, **7**, 179-188.

FRIED, L. S., & HOLYOAK, F. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 234-257.

FUKUNAGA, K. (1990). *Statistical pattern recognition* (2nd ed.). San Diego, CA: Academic Press.

GARNER, W. R. (1974). *The processing of information and structure*. New York: Wiley.

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

GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

HINTZMAN, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, **93**, 411-428.

KRANTZ, D. H., & TVERSKY, A. (1975). Similarity of rectangles: An analysis of subjective similarity. *Journal of Mathematical Psychology*, **12**, 4-34.

LUCE, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103-189). New York: Wiley.

MASSARO, D. W., & FRIEDMAN, D. (1990). Models of integration given multiple sources of information. *Psychological Review*, **97**, 225-252.

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

MORRISON, D. F. (1990). *Multivariate statistical methods* (3rd ed.). New York: McGraw-Hill.

NOSOFSKY, R. M. (1983). Information integration and the identification of stimulus noise and criterial noise in absolute judgement. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 299-309.

NOSOFSKY, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 299-309.

NOSOFSKY, R. [M]. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception & Psychophysics*, **38**, 415-432.

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

NOSOFSKY, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 87-108.

NOSOFSKY, R. M. (1988a). On exemplar-based representations: Reply to Ennis (1988). *Journal of Experimental Psychology: General*, **117**, 412-414.

NOSOFSKY, R. M. (1988b). Similarity, frequency, and the category representation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 54-65.

NOSOFSKY, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. *Perception & Psychophysics*, **45**, 279-290.

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

NOSOFSKY, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 3-27.

NOSOFSKY, R. M. (1992). Exemplar-based approach to relating categorization, identification and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363-393). Hillsdale, NJ: Erlbaum.

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.

NOSOFSKY, R. M., & SMITH, J. E. K. (1992). Similarity, identification, and categorization: Comment on Ashby and Lee (1991). *Journal of Experimental Psychology: General*, **121**, 237-245.

SAKAMOTO, Y., ISHIGURO, M., & KITAGAWA, G. (1986). *Akaike information criterion statistics*. Dordrecht: Reidel.

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. (1986). Discrimination and generalization in identification and classification: Comment on Nosofsky. *Journal of Experimental Psychology: General*, **115**, 58-61.

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

SHEPARD, R. N. (1988). Time and distance in generalization and discrimination: Reply to Ennis (1988). *Journal of Experimental Psychology: General*, **117**, 415-416.

SHEPARD, R. N. (in press). Integrality versus separability of stimulus dimensions: From an early convergence of evidence to a proposed theoretical basis. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure*.

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., & MEDIN, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

STEVENS, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.

TAKANE, Y., & SHIBAYAMA, T. (1992). Structure in stimulus identification data. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 335-362). Hillsdale, NJ: Erlbaum.

TVERSKY, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-352.

WICKENS, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: W. H. Freeman.

WIENER-ERLICH, W. K. (1978). Dimensional and metric structures in multidimensional stimuli. *Perception & Psychophysics*, **24**, 399-414.

NOTES

1. It may be inaccurate to define a nonlinear mapping from the physical space to the perceptual space as a *misperception* of the true dimensional values. Although it is true that a psychophysical mapping of this sort may yield a dimensional representation that differs from the physical dimensions of the stimulus, it is generally assumed (e.g., Stevens, 1975) that this mapping is one to one. In other words, each set of physical coordinates corresponds to a unique set of perceptual coordinates. Perceptual noise, on the other hand, is assumed to have a one-to-many mapping (i.e., each presentation of the same stimulus yields a unique perceptual effect).

2. Throughout this article we assume that perceptual noise, represented by the random variable e_p, is multivariate normally distributed with mean 0 and variance σ_p^2.

3. With normally distributed categories, the optimal decision bound is always quadratic. In 2 dimensions, the general quadratic equation is written $h_o(w_i) = a_1 w_{1i}^2 + a_2 w_{2i}^2 + a_3 w_{1i} w_{2i} + b_1 w_{1i} + b_2 w_{2i} + c_o$. The equation for an ellipse, a circle, and a line are a special case of the general quadratic equation. For example, when the coefficients a_1, a_2, and a_3 are zero, $h_o(w_i)$ is linear.

4. In the Ashby and Gott (1988) and Ashby and Maddox (1990) experiments, the optimal bound was linear. In the Ashby and Maddox (1992) experiments, the optimal bound was a quadratic curve.

5. An example is Busemeyer and Myung's (1992) *rule competition model*, which is composed of two parts. The first is an adaptive network model that describes how individuals choose different types of decision schemes (e.g., linear or quadratic). The second part uses a "hill climbing" algorithm to predict how subjects learn to fine tune their decision rule by adjusting the decision bound parameters.

6. Maximum likelihood estimators have many desirable properties. First, they are always consistent. Second, if an efficient unbiased estimator exists, maximum likelihood estimation will generally find it.

7. When calculating the optimal bound for the circular stimuli, we arbitrarily assumed that one quarter of a semicircle (i.e., $\pi/4$ radian) is psychologically equal to one quarter of the screen width (i.e., 250 pixels). If the subject assumes some other relation between the two dimensions, the optimal bound in the perceptual space will have some nonunit slope (see Ashby & Maddox, 1992, for a fuller discussion of this point).

8. In the general linear classifier, the separate effects of perceptual and criterial noise cannot be estimated due to the nonidentifiability of these two parameters.

9. This transformation was accomplished by premultiplying the dimensional values by the matrix

$$\mathbf{w} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \cos\theta & \sin\theta \end{bmatrix},$$

which rotates the perceptual dimensions by $\theta°$.

10. As in Data Sets 1 and 2, the general quadratic classifier was fit with the response bias parameter $\delta = 0$, whereas the fits of the exemplar models include a bias parameter. It is possible that the fits of the exemplar models could be reduced by up to 2 AIC points if one assumes no response bias (i.e., $\beta = .5$). There were two cases in which the fit of the DEM(θ) was within 2 AIC points of that for the general quadratic classifier. Refitting the data for these 2 subjects, under the assumption that $\beta = .5$, did not improve the fit.

11. Tversky (1977) and many others have argued that this method of data analysis also involves strong (and, in general, untenable) assumptions about the mapping from the stimulus space to the perceptual space.

12. Because category base rates are equal in these data sets, it is possible, as in Application 1, that assuming no response bias could improve the fit of the GCM and DEM. However, even if the bias *and* attention weight parameters were set equal to .5, the maximum improvement in fit would be 4 AIC points, which would not change any of the qualitative results in Table 6.

13. Ashby and Lee (1991) fitted the optimal decision bound model to these data. Although it performed somewhat better than the GCM, the optimal model performed more poorly than the suboptimal decision bound models of Table 6.

APPENDIX
Fitting the Exemplar Models

This appendix describes the specific techniques used in fitting the GCM, GCM(θ), DEM, and DEM(θ) to Application 1: Data Sets 1-5. Two considerations are important. First, one must determine whether repeated presentations of a given exemplar should lead to independent memory traces (e.g., Hintzman, 1986) or to a single memory trace. Exemplar models that assume independent memory traces are called *token* exemplar models, whereas those that assume a single memory trace are called *type* exemplar models. Because token models are frequency sensitive (Nosofsky, 1988b) and are thus more powerful, token exemplar models were fitted to the data.

Second, one must decide which exemplars to include in the summed similarity computations. The answer to this question differed for the data from the first and last experimental sessions, so each is dealt with in turn.

First Experimental Session

Each experimental session consisted of 5 presentations of the stimulus corresponding to the Category *A* and *B* means (10 presentations total), 100 practice trials, and 300 experimental trials. Although the probability of responding Category *A* for stimulus *i*, $P(A|i)$ was estimated for the 300 experimental trials only, it was assumed that a unique memory trace was formed for the 110 exemplars presented during the preexperimental learning sessions, and thus, these were included in the summed similarity computations. When estimating $P(A|i)$ on the *i*th experimental trial, all exemplars presented on Trials 1 through $(i-1)$, as well as those presented during the preexperimental learning sessions, were included in the Equation 6 and 12 sums.

Final Experimental Session

Ideally, when fitting the exemplar models to the data from the last experimental session, one might like to include the memory trace of all exemplars encountered by the subject up to that point in the experiment. By the last trial of a fifth experimental session, however, a subject will have seen 2,049 exemplars. It seems plausible that at some sample size, predictions of the exemplar models will asymptote. The most efficient fitting algorithm uses the smallest sample size after asymptote occurs.

To determine the point at which additional exemplar information provided no improvement in fit, the data from 8 subjects (those with the highest and lowest accuracies from each experiment and each condition of Data Sets 3 and 4) were fit using a procedure where the similarity of stimulus *i* was taken with respect to all 410 exemplars from the last session. The AIC

value for this model was compared with the AIC value for a model where the similarity was taken with respect to the 820 exemplars from the final and penultimate sessions. This procedure was repeated for the three versions of the GCM model (city-block-exponential, Euclidean-Gaussian, and Euclidean-exponential). The difference between the AIC for both methods was computed for each of the 24 cases (8 subjects × 3 versions). The mean AIC difference was −.044, the standard deviation was 3.013, and the range was −4.4 to 6. These results suggest exemplar information asymptotes at or before 410 samples, and so the similarity computations were taken with respect to the 410 exemplars during each subject's last experimental session.

The Local Minima Problem

To reduce the possibility of local minima in the exemplar model fits, a grid search of the parameter space was conducted.

For the GCM, 10 values of each parameter were chosen. The fit of the model was then recorded at each of these 1,000 (10 × 10 × 10) locations, and the parameter values corresponding to the minimum of these 1,000 values were then used as starting values for the minimization routine. Because the GCM is a special case of the GCM(θ) and DEM, the best-fitting values from the GCM were used as starting values for the GCM(θ) and DEM in a two-step procedure. First, the best-fitting parameters (c, w, and β) were held constant, and the γ or θ parameter (depending on the model) was left free to vary. Once the best fitting γ or θ parameter was obtained, all four parameters were left free to vary. A similar procedure was used for the DEM(θ), but in this case, the best-fitting parameters from the DEM were used as starting values for the DEM(θ) in a two-step procedure.