Journal of Experimental Psychology: General
1991, Vol. 120, No. 2, 150–172

0096-3445/91/$3.00

Predicting Similarity and Categorization From Identification

F. Gregory Ashby and W. William Lee
University of California, Santa Barbara

In this article, the relation between the identification, similarity judgment, and categorization of multidimensional perceptual stimuli is studied. The theoretical analysis focused on general recognition theory (GRT), which is a multidimensional generalization of signal detection theory. In one application, 2 Ss first identified a set of confusable stimuli and then made judgments of their pairwise similarity. The second application was to Nosofsky's (1985b, 1986) identification-categorization experiment. In both applications, a GRT model accounted for the identification data better than Luce's (1963) biased-choice model. The identification results were then used to predict performance in the similarity judgment and categorization conditions. The GRT identification model accurately predicted the similarity judgments under the assumption that Ss allocated attention to the 2 stimulus dimensions differently in the 2 tasks. The categorization data were predicted successfully without appealing to the notion of selective attention. Instead, a simpler GRT model that emphasized the different decision rules used in identification and categorization was adequate.

The perceptual processes involved when subjects identify, categorize, or judge the pairwise similarity of multidimensional perceptual stimuli are closely related (e.g., Ashby & Perrin, 1988; Getty, Swets, Swets, & Green, 1979; Nosofsky, 1986; Shepard & Chang, 1963; Shepard, Hovland, & Jenkins, 1961). Roughly speaking, as the similarity between a pair of stimuli increases, so too does the probability that one will be misidentified as the other and the probability that they will be assigned to the same category. This observation suggests a possible close relationship between these three tasks.

During the past several years, a number of theories have been developed that attempt to simultaneously account for data from all three of these tasks. Such theories are important because they represent attempts to integrate a broad spectrum of psychological data within one theoretical framework. In this article, we (a) explore the empirical relation between identification, categorization, and similarity judgment and (b) examine the ability of the more powerful of these theories to predict categorization performance and judgments of perceived similarity from the confusions that subjects make in an identification task.

The models that we focus most heavily on are derived from general recognition theory (GRT; Ashby & Gott, 1988; Ashby & Perrin, 1988; Ashby & Townsend, 1986). They assume that the perceptual effect associated with each presentation of a stimulus can be represented as a point in a multidimensional space but that perceptual noise causes the percept to vary over trials. Thus, GRT assumes that a distribution of percepts is the appropriate perceptual representation of a stimulus. During identification or categorization, the subject is assumed to divide the perceptual space into response regions. On each trial, the subject determines in which region the perceptual effect falls and then emits the associated response.

If the same stimuli are used in different experimental tasks and if the subject's perceptual system is unchanged, then the resulting perceptual representations should be related. Nosofsky (1986) effectively argued that to account for this relation, one must allow for systematic shifts in selective attention across the various tasks. However, we argue below that the key to understanding the relationship between identification, categorization, and similarity judgment is to understand the manner in which response regions change with the experimenter's instruction. As shown below, to a large extent, these changes are accurately predicted by assuming that subjects are trying to maximize response accuracy.

The models are tested against two separate data sets. The first are identification-similarity data from an experiment reported below, and the second are the identification-categorization data reported by Nosofsky (1985b; 1986). In both experiments, subjects began by participating in an identification task that lasted for a number of experimental sessions. After the last identification session, the same subjects began a series of sessions in which they performed a new experimental task with the same stimuli. In our data set, the new task required subjects to judge the similarity of stimulus pairs. In the Nosofsky data set, the new task required subjects to learn four different rules for categorizing the stimuli. Nosofsky (1986) demonstrated that categorization performance can be predicted from the errors made in an identification task. To our knowledge, however, no one has attempted to predict similarity judgments from identification performance (but see Getty et al., 1979). In addition, GRT models have never been simultaneously fit to data from two separate experiments.

Both experiments involved stimuli of the type shown in Figure 1. On the basis of a number of independent tests, the size and orientation components of the Figure 1 stimuli have been assumed to be perceptually separable (e.g., Burns, Shepp, McDonough, & Wiener-Ehrlich, 1978; Garner & Felfoldy, 1970; Hyman & Well, 1967; Shepard, 1964; see, however,

This research was supported in part by National Science Foundation Grant BNS88-19403 to F. Gregory Ashby. We thank Daniel Ennis, Todd Maddox, Robert Nosofsky, and Keith Smith for their helpful comments. We also thank Cindy Castillo for her help preparing the article.

Correspondence concerning this article should be addressed to F. Gregory Ashby, Department of Psychology, University of California, Santa Barbara, California 93106.

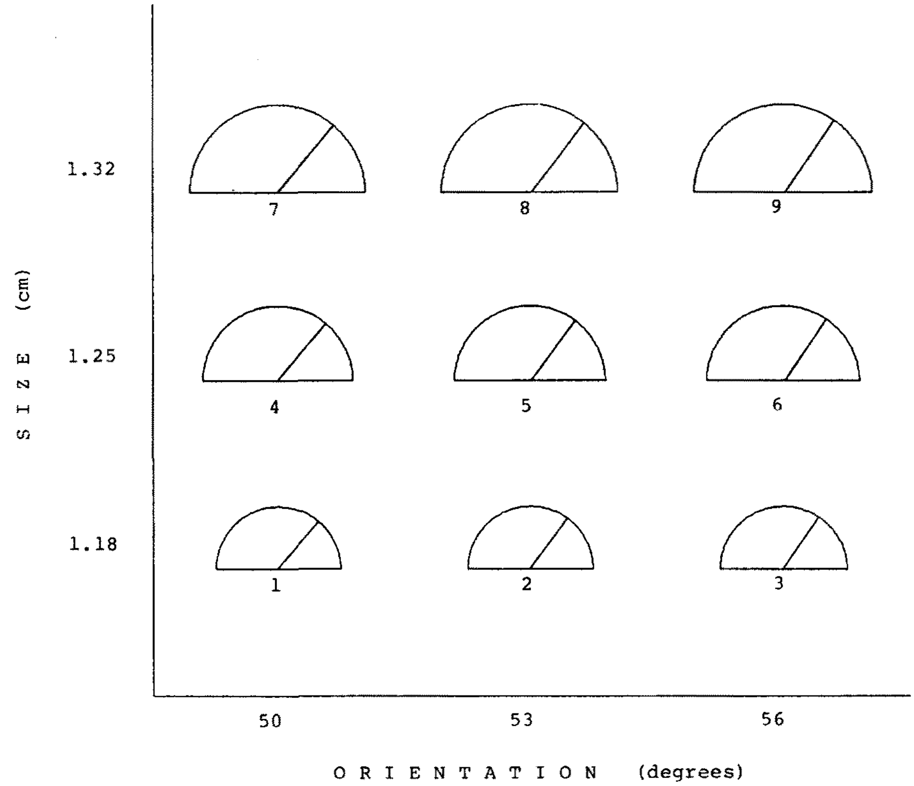

Figure 1. Stimulus set for Experiment 1.

Ashby & Maddox, 1990). Later in this article, however, we question this assumption.

On each trial of an identification experiment, subjects are shown one of n possible stimuli and are asked to identify it uniquely. Call the n stimuli $S_1, S_2, \cdots, S_n$ and the n associated responses $R_1, R_2, \cdots, R_n$. The data from an identification experiment are summarized in an $\mathbf{n} \times \mathbf{n}$ confusion matrix in which the entry in Row i and Column j gives the frequency with which subjects emit Response R_j on trials when Stimulus S_i is presented. In a categorization experiment, the n stimuli are partitioned by the experimenter into m categories $C_1, C_2, \cdots, C_m$ (where $m < n$). On each trial, subjects are shown one of the n stimuli and are asked to name the category to which it belongs. Data from a categorization task are summarized in an $\mathbf{n} \times \mathbf{m}$ confusion matrix in which the entry in Row i and Column j gives the frequency with which subjects categorize Stimulus S_i into Category C_j. On each trial of a similarity judgment experiment, subjects are shown two of the n stimuli. The subjects' task is to rate the similarity of the pair on a scale from 1 to R, with 1 meaning very dissimilar and R meaning very similar. The relevant data from a similarity judgment experiment is a rank ordering of all stimulus pairs according to their judged similarity. The identification, categorization, and similarity judgment models studied in this article were designed to account for the observed identification and categorization confusion matrices and for the similarity rank orderings.

Identification Models

General Recognition Theory (GRT)

General recognition theory is a multivariate extension of signal detection theory (Green & Swets, 1966; Tanner, 1956). It assumes that, on any given trial, the perceptual effect of a stimulus can be represented as a point in a multidimensional space. However, like signal detection theory, GRT assumes that repeated presentations of the same stimulus do not always lead to the same perceptual effect. Thus, over trials, the perceptual effects of a stimulus are represented by a multivariate probability distribution.

With the Figure 1 stimuli, GRT models naturally postulate two perceptual dimensions: perceived size and perceived orientation. In this case, each stimulus is represented as a bivariate probability distribution. An example in which the perceptual distributions are normal is shown in Figure 2. Note that the distribution has a three dimensional bell-like structure. The height of the bell at any particular size and orientation represents the likelihood that presentation of the stimulus will elicit a percept having that particular size and orientation. Rather than draw a three-dimensional figure, it is conventional to depict a bivariate normal distribution by its contours of equal likelihood, each of which is created by taking a slice parallel to the perceptual plane and looking down at the result from above.

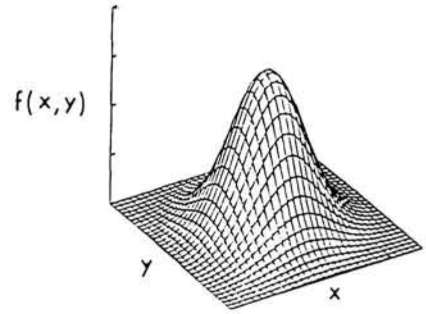

Figure 2. Bivariate normal distribution.

Multivariate normal distributions are specified by three kinds of parameters: (a) location parameters (i.e., a mean on each dimension), (b) spread parameters (i.e., a variance on each dimension), and (c) association parameters (i.e., a covariance or correlation for each pair of dimensions). With bivariate normal distributions, the contours of equal likelihood are always circles or ellipses. If the covariance term is zero, the major and minor axes of the ellipse are always parallel to the coordinate axes. With a positive covariance, the major axis has a positive slope, and with a negative covariance, it has a negative slope. The relative magnitude of the variances determines whether the contours are wider in the direction of the *x*- or *y*-axis.

In an identification task, GRT assumes the subject divides the perceptual space into regions and associates a response label with each region. On each trial, the subject determines in which region the stimulus representation falls and then emits the associated response. The *decision bounds* are the lines or curves that separate one response region from another. With nine stimuli arranged in a 3 × 3 configuration (as in Figure 1), nine response regions are required. Figure 3 shows the nine response regions in a hypothetical perceptual space (for other examples, see Figure 6).

The probability of responding R_j on trials in which Stimulus S_i was presented is equal to the probability that a random sample from the perceptual distribution associated with Stimulus S_i falls in the response region associated with response R_j. Formally,

$$P(R_j \mid S_i) = \int_{\mathscr{R}_j}\int f_i(x, y)\, dx\, dy \qquad (1)$$

where $f_i(x, y)$ is the bivariate normal probability density function for Stimulus S_i, and $\mathscr{R}_j$ signifies the R_j response region.

Consider the Figure 1 case in which nine stimuli are arranged in a 3 × 3 configuration and in which there are two relevant perceptual dimensions. In this case, each perceptual distribution has five parameters (i.e., μ_x, μ_y, σ_x^2, σ_y^2, and ρ), and in addition, some parameters must be used to specify the decision bounds. Although a 9 × 9 confusion matrix has enough degrees of freedom to test such a model, for a number of reasons it is desirable to also test more restrictive versions of GRT. For example, by selectively fixing certain parameters, various assumptions about the perceptual processing of the

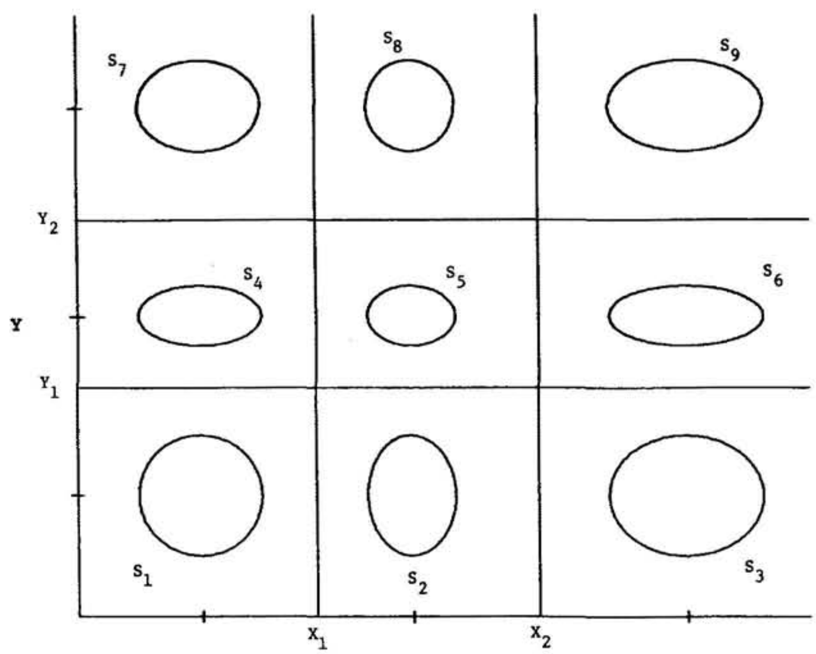

Figure 3. Hypothetical decision bounds and contours of equal likelihood that satisfy perceptual independence, perceptual separability, and decisional separability.

stimuli can be tested, such as whether the stimulus components are separable and whether they are perceived independently.

In many models, the process of adding simplifying assumptions is guided by mathematical simplicity rather than by empirical validity. One important attribute of GRT is that many simplifying assumptions can be tested empirically before being considered for incorporation into the model. If the data do not support a particular assumption, it can be replaced by one with greater empirical validity. A number of techniques for testing many plausible simplifying assumptions have been developed (Ashby, 1988; Ashby & Townsend, 1986; Kadlec & Townsend, in press; Wickens & Olzak, in press). This article focuses on three such assumptions: perceptual independence, perceptual separability, and decisional separability.

Perceptual independence (PI) holds for a pair of components in a particular stimulus if the perceptual effects of the two components are statistically independent (Ashby & Townsend, 1986). When the perceptual distributions are bivariate normal, perceptual independence holds for a given stimulus if—and only if—the associated covariance parameter is zero. This in turn means that the contours of equal likelihood must be either circles or ellipses in which the major and minor axes are parallel to the coordinate axes. All nine contours in Figure 3 satisfy PI. Note that PI is a property of a single stimulus.

Perceptual separability (PS) holds if the perceptual effects of a given level of one component do not depend on the level of the other component. For example, consider a stimulus A_iB_j that is constructed from Component A at Level i and from Component B at Level j. Component A is perceptually separable from B at Level i if the perceptual effects of A_i do not depend on the level of B. Therefore, PS is a property of an entire set of stimuli (i.e., of $A_iB_1, A_iB_2, \cdots, A_iB_r$, where r is the number of levels of Component B). If Dimension x corresponds to Component A and Dimension y to B and if the perceptual distributions are normal, then the PS of A_i from Component B implies that Stimuli $A_iB_1, A_iB_2, \cdots, A_iB_r$ all have the same mean and variance on Dimension x. Components A and B are mutually perceptually separable if A_i is perceptually separable from B for all values of i and B_j is separable from A for all values of j. Note that this condition implies a rectangular gridlike configuration of the perceptual means. The contours illustrated in Figure 3 satisfy mutual PS.

If decisional separability (DS) holds, then the subject's decision about the level of one component does not depend on the value of the perceptual effect associated with any other component. This property, which often leads to suboptimal performance, implies that the decision bounds are parallel to the coordinate axes. Note that DS also holds in Figure 3.

A convenient analogy that provides psychological intuition for these concepts is as follows (Ashby, 1989). Suppose a processing channel exists for each stimulus component. Let the critical set of a channel be the set of all possible stimuli that directly stimulate that channel. If there is no mutual interaction between the two channels, PI holds. If the critical sets associated with the two channels do not overlap, PS holds. If the subject makes a decision about the level of a component by setting a criterion on the output of the channel associated with that component, DS occurs.

Luce-Shepard Choice Models

To evaluate the validity of the GRT models, they are compared with a number of models derived from the classic biased-choice model (Luce, 1963; Shepard, 1957). In the biased-choice model, $P(R_j|S_i)$ is a function of the similarity of Stimulus S_i to Stimulus S_j, denoted η_{ij}, and of the bias toward response R_j, denoted β_j. Specifically,

$$P(R_j|S_i) = \frac{\beta_j\eta_{ij}}{\sum_{k=1}^{n} \beta_k\eta_{ik}} \quad (2)$$

where n is the number of stimuli in the ensemble. When applying this model, it is typically assumed that similarity is symmetric, so that $\eta_{ij} = \eta_{ji}$. In addition, without loss of generality, it is assumed that all self-similarities equal 1.0 (i.e., $\eta_{ii} = \eta_{jj} = 1.0$, for all i and j). Despite the fact that a number of researchers have questioned the empirical validity of the symmetry assumption (e.g., Krumhansl, 1978; Tversky, 1977), the biased-choice model has been the most successful identification model of the past 25 years (when the criterion is goodness-of-fit to an identification confusion matrix; e.g., Smith, 1980; Townsend, 1971; Townsend & Ashby, 1982).

In fact, the biased-choice model has been so successful that one might wonder why other identification models should be investigated. For example, Nosofsky (1986) found that the biased-choice model accounted for more than 99% of the variance in the identification confusion matrices obtained from both subjects in an experiment that used stimuli like those in Figure 1. Despite results such as these, we believe there are at least four reasons to investigate other models. First, the biased-choice model, as given in Equation 2, is cognitively sterile in the sense that a good fit to a particular data set provides little insight into the perceptual and cognitive processes involved in the identification task. Second, although the biased-choice model usually provides excellent fits to identification data, it occasionally encounters difficulty. For example, Ashby and Perrin (1988) found that it provided poor fits to 3 of the 8 confusion matrices reported by Townsend, Hu, and Ashby (1981). Third, even a model that accounts for 99% of the variance in a particular data set might be improved. An alternative model that accounts for 99.5% of the variance in the same data accounts for 50% of the unexplained variance. Newtonian physics accounts for most of the variance in astronomical data, but it was rejected in favor of a more powerful model many years ago. Finally, when accounting for data in an identification confusion matrix, percentage of variance accounted for is a misleading statistic. In most cases, the main diagonal contains the largest entries in an identification confusion matrix. Thus, any model that predicts that the probability of a correct response is greater than the probability of an error will account for a large percentage of the observed variance.

The MDS-choice model (i.e., multidimensional scaling-choice model) is a special case of the biased-choice model that in identification. In the MDS-choice model, η_{ij} is replaced with a more structured similarity measure (Getty, Swets, & Swets 1980; Getty et al., 1979; Nosofsky, 1985b; Shepard, 1957). Aside from symmetry, the biased-choice model makes no assumptions about the relation of various stimulus similarities. In contrast, the MDS-choice model assumes that stimuli can be represented as points in a multidimensional perceptual space. Pairwise similarities can be inferred from the interpoint distances. If the MDS space is of low dimensionality, the MDS-choice model may require fewer parameters to derive all the pairwise similarities than the biased-choice model. In addition, the MDS representation may yield insights about the perceptual processing of the stimuli that cannot be obtained from the biased-choice model.

With the stimuli used in our experiments, the model naturally postulates a perceptual space with two dimensions: orientation and size. Perceived similarity, as represented by η_{ij} is assumed to be inversely related to the distance between the point-representations of S_i and S_j. The exact relation between similarity and distance is defined by the similarity function. Two alternatives are prominent: the exponential decay and the Gaussian.

Let d_{ij} be the distance between the point-representations of Stimuli S_i and S_j. The exponential decay similarity function assumes

$$\eta_{ij} = e^{-d_{ij}} \qquad (3)$$

and the Gaussian similarity function assumes

$$\eta_{ij} = e^{-d^2_{ij}}. \qquad (4)$$

Shepard (1957, 1958a, 1958b, 1987, 1988) proposed the exponential decay function as a universal principle. Nosofsky (1985b) found evidence that with highly confusable stimuli, better fits are obtained with the Gaussian similarity function, but Ennis (1988) showed that with probabilistic stimulus representations, the two models are difficult to discriminate. Finally, Shepard (1986, 1988) and Takane and Shibayama (in press) argued that the Gaussian function is most appropriate in cases in which subjects are highly practiced and the models are fit to the confusion matrices of single subjects. If the data are aggregated over subjects, or if the subjects are not highly practiced, the exponential function is likely to fit better. The argument runs as follows. As d increases, $\exp(-d^2)$ approaches zero more quickly than $\exp(-d)$, and thus, with the Gaussian similarity function, the contribution to the response probabilities of very dissimilar stimuli is less than with the exponential function. Thus, any factor that increases confusions between dissimilar stimuli will tend to favor the exponential similarity function.

Different versions of the MDS-choice model can also be formulated depending on how distance is defined. In this article, two specific distance metrics are considered. Both are special cases of the so-called Minkowski metric. Let (u_i, y_i) be the coordinates of the perceptual representation of Stimulus S_i. Then the Minkowski metric defines the distance d_{ij} by

$$d_{ij} = [|u_i - u_j|^r + |v_i - v_j|^r]^{1/r}. \qquad (5)$$

The exponent r defines the nature of the distance metric. Euclidean distance results when $r = 2$ and city-block distance when $r = 1$. Previous applications of the MDS-choice model have paired the exponential similarity function with the city-block distance metric (the exponential–city-block MDS-choice model; Nosofsky, 1985a; Shepard, 1957, 1987; Takane & Shibayama, 1986), the Gaussian similarity function with the Euclidean metric (the Gaussian–Euclidean MDS-choice model; Nosofsky 1985b, 1986), or the exponential similarity function with the Euclidean metric (the exponential–Euclidean MDS-choice model; Nosofsky, 1987). These are the three versions of the MDS-choice model tested in this article.

Traditionally, the exponential–city-block model was thought to fit best when the dimensions are separable and the exponential–Euclidean model to fit best when the stimulus dimensions are integral (Attneave, 1950; Garner, 1974; Handel & Imai, 1972; Hyman & Well, 1968; Shepard, 1964; Torgerson, 1958). However, more recently a number of researchers have found that with separable dimensions the best fitting MDS-choice model is the Gaussian–Euclidean (Ashby & Perrin, 1988; Nosofsky, 1985b).

The MDS-choice model is a special case of the biased-choice model in the following sense: An identification confusion matrix that is generated from any MDS-choice model can be fit perfectly by some biased-choice model, but identification confusion matrices can be generated from the biased-choice model that cannot be fit perfectly by any MDS-choice model with the same number of parameters. To see this, note from Equations 2, 3, and 4 that the MDS-choice model satisfies each of the restrictions that the biased-choice model places on the similarity parameters. Namely, $\eta_{ij} = \eta_{ji}$ and $\eta_{ii} = 1$ for all i. Thus, in the biased-choice model, there exists a unique similarity parameter, η_{ij}, for each unique distance d_{ij} in the MDS-choice model. A consequence of this equivalence is that the MDS-choice model can never provide a better absolute fit to a set of data than the biased-choice model. It can, however, provide a better relative fit. Specifically, it could provide a fit that is almost as good, but at the savings of many free parameters.

Although the MDS-choice model and GRT both assume that stimuli can be represented in a multidimensional perceptual space, the similarity between the two approaches ends with this assumption. The GRT models (a) emphasize variability in the percept, (b) assume that confusability is a fundamental construct, (c) stress the importance of the decision bound, and (d) assume that response selection is a deterministic process (i.e., if a percept falls in Response Region A, then Response R_A is given with probability 1). On the other hand, the MDS-choice models (a) assume a static perceptual representation, (b) assume that similarity is a fundamental construct, (c) stress the importance of interpoint distance, and (d) assume that response selection is a probabilistic process (i.e., a given percept always leads to an R_A response probability between 0 and 1).

Similarity Models

General Recognition Theory

Confusability is closely related to similarity. In general, the greater the similarity between a pair of stimuli, the more often they are confused. In fact, GRT assumes that, in the absence of response bias, confusability and perceived similarity are proportional (Ashby & Perrin, 1988). Thus, when the stimuli are two dimensional, the similarity of Stimulus S_i to Stimulus S_j is given by

$$s(S_i, S_j) = k \int_{\mathscr{R}_j}\int f_i(x, y)\, dx\, dy \qquad (6)$$

where k is a positive constant, $f_i(x, y)$ is the bivariate normal probability density function representing the perceptual effects of Stimulus S_i, and $\mathscr{R}_j$ signifies the region for responding R_j in an unbiased identification experiment. Without loss of generality, k can be set to 1. Thus, in GRT, perceived similarity is closely related to identification accuracy. Ashby and Perrin (1988) showed that this model contains the Euclidean MDS models as a special case, even those that assume that subjects differentially weight potentially oblique perceptual dimensions. Even so, the Equation 6 similarity measure is not constrained by any of the so-called distance axioms. In particular, the predicted similarities need not be symmetric and not all self-similarities need be equal.

The relationship between GRT and MDS is summarized in Figure 4. The left side of the figure shows the hypothetical stimulus representation of four stimuli—A, B, C, and D—which vary on two physical dimensions, X and Y. Four possible GRT perceptual representations of these stimuli are depicted in the middle of Figure 4, and the equivalent MDS representation is shown on the right. In the top GRT representation, perceptual independence holds for each stimulus, and all stimuli have equal amounts of perceptual variability on both dimensions. In this case, the equivalent MDS representation comes from the simple Euclidean MDS model (i.e., Equation 5 with $r = 2$). In the second GRT representation, the Perceptual Dimensions x and y are always perceived independently, and all stimuli are associated with the same amount of perceptual variability along Dimension x and also along Dimension y. Note, however, that there is greater variability along Dimension y than along Dimension x. The equivalent MDS representation comes from the weighted Euclidean scaling model (Carroll & Chang, 1970; Horan, 1969). This model allows subjects to differentially weight the dimensions. In Figure 4, greater weight is placed on Dimension u. In the third GRT representation, x and y exhibit a positive dependence, but the magnitude of this dependence (i.e., ρ) does not depend on which stimulus was presented. The equivalent MDS representation comes from the so-called general Euclidean scaling (GEM) model (Carroll & Chang, 1972; Tucker, 1972). In the GEM representation, the angle between Perceptual Dimensions u and v satisfies $\rho = -\cos\theta$. Finally, in the bottom GRT representation, Perceptual Dimensions x and y are perceived independently when Stimulus A is presented, but when B is presented x and y exhibit a negative dependence. In this case, there is no equivalent MDS representation.

Note that, although the Stimulus Vectors **B** and **C** are orthogonal in the stimulus representation, they are not orthogonal in any GRT or MDS representation. This illustrates that the dimensions of the GRT or MDS representation are not stimulus dimensions but instead are perceptual dimensions. The MDS-choice models and the GRT models are alternative theories of the perceptual representation of multidimensional stimuli. In each theory, a perceptual representation mediates a choice process.

MDS-Choice Model

In the MDS-choice model, similarity is specifically defined by η_{ij}, which is a function of the psychological distance between Stimuli S_i and S_j (see Equations 2, 3, and 4). The estimates of the coordinate values of each point-representation (the u_i and v_i) are substituted into Equation 3 or 4 to generate the similarity predictions.

Categorization Models

General Recognition Theory

The GRT identification model can be generalized to predict categorization performance. Consider the situation in which there are two categories, C_A and C_B, and a total of n stimuli. In this case, GRT postulates that the subject divides the perceptual space into two regions: one associated with Response R_A and one associated with Response R_B. Thus, the probability of responding R_J when Stimulus S_i was presented is

$$P(R_J|S_i) = \int_{\mathscr{R}_J}\int f_i(x, y)\, dx\, dy \qquad (7)$$

where $\mathscr{R}_J$ refers to the response region of Category C_J, and $f_i(x, y)$ is again the bivariate normal probability density function representing Stimulus S_i. (Note that lowercase subscripts refer to individual stimuli and uppercase subscripts refer to categories.)

A number of different versions of this model can be formulated depending on how the boundary that separates the two response regions is constructed. As a concrete example, suppose the experimenter assigns Stimuli $S_1 - S_4$ of Figure 1 to Category C_A and Stimuli $S_5 - S_9$ to Category C_B. Suppose that during the categorization phase of the experiment, there are an equal number of Category C_A and C_B trials, and within each category, each stimulus is presented with equal probability. Then the distribution of percepts on Category C_A trials is

$$f_A(x, y) = \tfrac{1}{4}f_1(x, y) + \tfrac{1}{4}f_2(x, y) + \tfrac{1}{4}f_3(x, y) + \tfrac{1}{4}f_4(x, y).$$

Similarly, the distribution of percepts on Category C_B trials is

$$f_B(x, y) = \tfrac{1}{5}f_5(x, y) + \tfrac{1}{5}f_6(x, y) + \tfrac{1}{5}f_7(x, y) + \tfrac{1}{5}f_8(x, y) + \tfrac{1}{5}f_9(x, y).$$

The optimal classifier uses the decision rule

$$\text{Respond } R_A \text{ if } \frac{f_A(x, y)}{f_B(x, y)} > 1; \text{ otherwise respond } R_B$$

and so the response region $\mathscr{R}_A$ includes all points (x, y) such that $f_A(x, y)/f_B(x, y) > 1$. The decision bound is the set of all points (x, y) such that $f_A(x, y)/f_B(x, y) = 1$. In all but a very few special cases, this bound will be highly nonlinear.

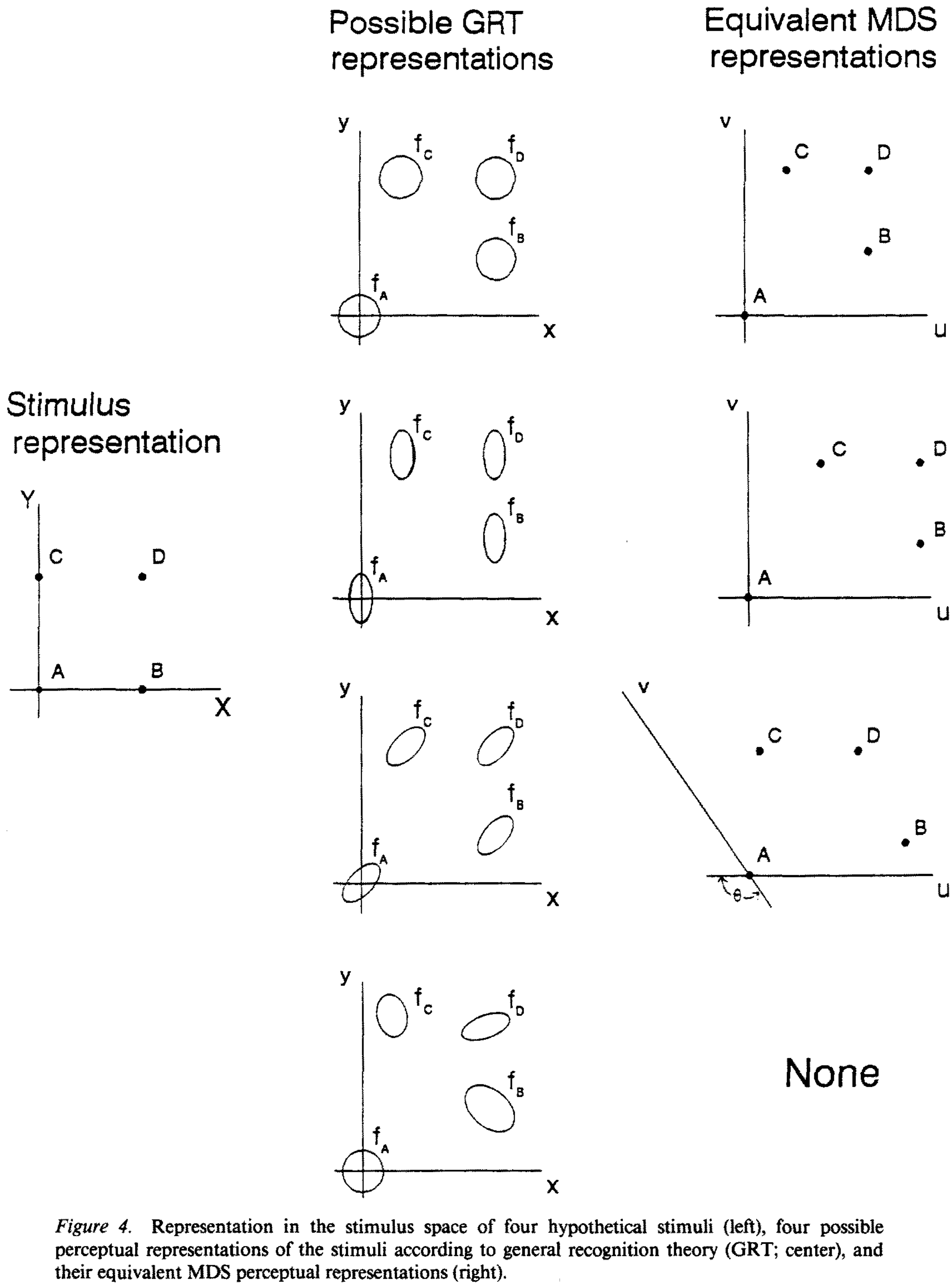

Figure 4. Representation in the stimulus space of four hypothetical stimuli (left), four possible perceptual representations of the stimuli according to general recognition theory (GRT; center), and their equivalent MDS perceptual representations (right).

To generate categorization predictions from identification performance, the GRT identification model is first fit to the identification confusion matrices. Next, estimates of the parameters describing the perceptual distribution associated with each stimulus are obtained from the best fitting model. These are then used in conjunction with Equation 7 to obtain the predicted accuracies in the categorization condition.

Generalized Context Model

Medin and Schaffer (1978) developed the context model as an application and extension of the biased-choice model to the categorization paradigm. Consider a categorization task with two categories, C_A and C_B. According to the context model, the probability that Stimulus S_i is classified as a member of Category C_A, $P(R_A|S_i)$, is given by

$$P(R_A|S_i) = \frac{\beta_A \sum_{j \in C_A} \eta_{ij}}{\beta_A \sum_{j \in C_A} \eta_{ij} + (1 - \beta_A) \sum_{k \in C_B} \eta_{ik}} \quad (8)$$

where as before, β_A is the bias for responding R_A, and η_{ij} is the similarity between Stimuli S_i and S_j.

Nosofsky (1984, 1986) generalized the context model to continuous valued perceptual dimensions. The resulting model, known as the *generalized context model* (GCM) assumes

$$\eta_{ij} = e^{-(d_{ij})^{\alpha}} \quad (9)$$

where, as before, d_{ij} is the distance between the perceptual representations of Stimuli S_i and S_j and α is generally set to $\alpha = 1$ (the exponential similarity function) or $\alpha = 2$ (the Gaussian similarity function).

The GCM postulates a more powerful distance measure than the MDS-choice model. Nosofsky (1986) assumed that whereas the coordinates of the perceptual representation of each stimulus are the same in identification and categorization, there is a possibility that the proportion of attention allocated to the various perceptual dimensions might be different in the two tasks. For example, accurate identification of the Figure 1 stimuli requires that each perceptual dimension be allocated approximately equal amounts of attention. Consider, however, a categorization task in which three categories are created, one for each circle size. In this case, there is no reason to allocate any attention to the orientation dimension.

Let the proportion of attention allocated to Dimensions u and v be denoted by w_u and w_v, respectively (where $w_u + w_v = 1$). Then the GCM assumes that the distance between the perceptual representations of Stimuli S_i and S_j is given by

$$d_{ij} = c[w_u|u_i - u_j|^r + w_v|v_i - v_j|^r]^{1/r}. \quad (10)$$

As in the MDS-choice model, the Parameter r is generally set to $r = 1$ (city-block distance) or $r = 2$ (Euclidean distance). The nonnegative Parameter c scales the perceptual space. It can be interpreted as a measure of overall stimulus discriminability and so is expected to increase with increased exposure duration or as subjects gain experience with the stimuli.

Although Nosofsky (1990) showed that under certain restrictive conditions, the GCM and the GRT models make similar predictions, the two models nevertheless differ fundamentally in their assumptions about the categorization process. The GCM is an exemplar model. It assumes that on each trial, the subject performs a global match between the representation of the presented stimulus and the memory representation of every exemplar in each category. The outcome of this matching process determines the strength of each response alternative. As long as the parameters have finite values, response strength is always positive valued, and thus, response selection is probabilistic. The GRT model assumes that exemplar information is available, but unnecessary for experienced subjects.[1] As soon as the subject determines in which region the percept has fallen (which is sometimes difficult), response selection occurs automatically (and deterministically). For example, when categorizing faces by ethnicity, GRT assumes that through experience, subjects have learned to associate certain ethnicities with certain regions of the "face" space. So long as a particular face is not near a bound that separates regions associated with different ethnicities, categorization is automatic. In particular, there is no need to access the representations of the many faces in each candidate ethnicity (as GCM assumes).

Experiment 1

The basic idea behind the two experiments described in this article is that, if the processes and representations underlying the various psychophysical tasks (such as identification, categorization, and similarity judgment) are related, then it should be possible to predict performance in one task from performance in another task. Specifically, in Experiment 1, we wanted to predict performance in a similarity judgment task from identification confusions. Thus, the goals of Experiment 1 were to (a) collect identification confusion data, (b) fit the models to the data, (c) generate similarity predictions using the results of the identification model-fits, (d) collect similarity-judgment data, and (e) compute the correlation between the predicted and observed similarity judgments. The analysis of these data allows us to compare the models on how well each accounts for the identification data and how well each predicts the similarity-judgments.

The stimulus set consisted of the nine semicircles shown in Figure 1. Note that the stimuli vary on two dimensions: the size of the circle and the orientation of the radial line. These two dimensions have been found to be perceptually separable in a number of independent tests (e.g., Burns et al., 1978; Garner & Felfoldy, 1970; Hyman & Well, 1967; Shepard, 1964; however, see Ashby & Maddox, 1990; Nosofsky, 1985b).

Two subjects each participated in five sessions of identification and five sessions in which they judged the similarity of pairs of the stimuli (on a scale from 1 to 10). On the average, each of the 9 stimuli were presented about 165 times during the course of the identification sessions, and each of the 81 stimulus pairs were presented about 22 times during the course of the similarity sessions.

Virtually all identification models are able to account for data in which accuracy is perfect. The challenge is to account for the specific confusions that subjects make when accuracy is degraded. Accurate estimation of specific error probabilities requires a high overall error rate. A number of specific experimental manipulations effectively increase error rate. These include choosing highly similar stimuli, limiting exposure duration, and using a backward mask. Following Nosofsky (1986), we chose to use all three of these techniques. General recognition theory predicts that each technique will increase perceptual variability (or decrease the distance between perceptual means), but that they should not fundamentally disrupt normal identification processes.[2]

[1] This is the case at least in a categorization task. In other tasks, for example, when subjects are asked to make typicality judgments, exemplar information is necessary.

[2] The presence of a mask is most likely to disrupt the natural identification process, particularly if the mask shares features with the stimulus. Therefore, our mask consisted only of a rectangular grid of white dots.

Typically, when collecting judgments of similarity, one does not limit exposure duration or terminate the display with a mask. However, prediction of similarity judgments from identification data is facilitated if the perceptual representation is invariant across the two tasks. The best way to minimize differences in the perceptual representations is to keep the stimulus display conditions the same in the two conditions. Unfortunately, however, because an identification trial involves only one stimulus and a similarity judgment trial involves two stimuli, identical stimulus conditions are impossible. The best one can do is to select a longer exposure duration in the similarity condition.

Method

Subjects

Two subjects participated in this experiment. Subject 1 was a female undergraduate who had never been in a psychology experiment. Subject 2 was a male graduate student, experienced in experiments similar to this one. Both were students at the University of California at Santa Barbara and were paid for their participation. Both subjects had normal vision.

Stimuli

The stimulus set consisted of nine semicircular figures (see Figure 1), constructed by factorially combining three sizes (i.e., radii) and three orientations. The radii were 1.18, 1.25, and 1.32 cm; the orientations were 50°, 53°, and 56°. Average visual angle was about 1°, and thus these stimuli were similar to those used by Nosofsky (1985b, 1986). The stimuli were computer generated and displayed on a Mitsubishi Electric Color Display Monitor Model No. C-9918NB in a dimly lit room.

Procedure

Identification trials. Before the actual trials began, subjects were shown a drawing of how the stimuli were arranged in a 3 × 3 configuration (similar to Figure 1) and how each stimulus could be identified by two numbers: the level on the orientation dimension and the level on the size dimension. They were told that to respond correctly, they must correctly identify the levels along both dimensions.

Each trial began with a fixation dot that appeared in the center of the screen for 500 ms. Next, one of the nine stimuli was randomly selected and then displayed for 150 ms. This was followed immediately by a mask (i.e., a grid of dots). Subjects had up to 10 s to make a response. As soon as the subjects responded, the correct response appeared for 1.5 s. This was followed by a 2-s pause, and then the next trial was initiated. Subjects made their responses by pressing one of nine buttons on a 3 × 3 keypad. There were a total of five sessions (days) of identification. Each session consisted of 300 trials in four blocks of 75 trials; there was a 30-s break between blocks. On the average, each stimulus was shown about 165 times during the five sessions. Because we were interested in asymptotic identification performance and not in the learning process, the first session was considered practice, and data from this session were not included in the subsequent analysis.

Similarity-judgment trials. After the end of the identification sessions, each subject completed five sessions (days) of judging similarity. Each trial began with a fixation dot that appeared for 500 ms.

Table 1
Observed and Predicted Identification Confusions

S_i	1	2	3	4	5	6	7	8	9
				Subject 1					
1(1, 1)[a]	61	18	0	41	6	1	6	1	0
	61	19	1	40	6	1	6	1	0
	60	17	0	44	7	0	6	0	0
2(2, 1)	29	41	4	32	23	3	4	3	0
	27	41	7	31	23	3	4	3	0
	31	43	6	31	22	0	5	1	0
3(3, 1)	2	10	68	3	19	36	1	1	1
	1	3	69	4	19	38	2	2	2
	0	11	69	0	18	36	0	3	3
4(1, 2)	4	4	2	65	13	4	38	8	1
	8	4	1	64	13	4	36	7	1
	10	3	1	63	13	4	38	6	0
5(2, 2)	4	7	6	18	37	21	12	15	5
	2	4	9	18	37	21	12	16	6
	2	5	6	17	35	20	14	19	7
6(3, 2)	0	1	14	3	11	56	2	16	26
	0	0	10	3	12	57	2	17	28
	0	1	13	0	15	60	0	16	25
7(1, 3)	0	0	0	13	5	1	101	27	0
	0	0	0	16	4	1	100	24	1
	0	0	0	13	5	1	101	27	0
8(2, 3)	0	1	0	2	6	7	15	72	33
	0	0	0	2	4	7	18	72	31
	0	0	0	1	9	8	15	71	33
9(3, 3)	0	0	1	0	2	13	1	19	74
	0	0	0	0	1	9	0	23	76
	0	0	0	0	2	12	0	20	76
				Subject 2					
1(1, 1)	49	14	2	33	8	1	11	6	1
	49	19	0	30	9	0	10	6	2
	50	15	0	33	11	0	12	4	0
2(2, 1)	35	40	4	15	27	0	2	10	0
	32	40	5	15	29	1	2	10	0
	35	40	3	15	28	0	4	9	0
3(3, 1)	3	28	66	0	16	20	2	6	9
	3	28	66	0	16	20	2	6	9
	4	28	66	2	16	20	1	5	9
4(1, 2)	12	7	1	31	36	1	33	29	5
	17	5	0	31	37	0	30	28	6
	10	9	2	33	35	0	33	29	5
5(2, 2)	3	21	3	7	43	6	4	20	10
	1	3	0	10	49	0	7	28	19
	5	20	5	6	43	5	2	20	10
6(3, 2)	2	8	33	0	18	31	1	11	22
	2	8	33	0	18	31	1	11	22
	0	8	34	0	19	31	0	11	23
7(1, 3)	2	1	1	13	16	0	51	39	6
	3	0	0	17	16	0	50	34	7
	0	1	1	13	16	2	51	39	6
8(2, 3)	0	4	1	3	27	10	5	39	44
	1	1	0	7	26	0	15	40	43
	1	4	1	3	27	10	4	39	44
9(3, 3)	0	0	5	0	5	25	1	3	92
	0	0	0	1	8	0	1	18	103
	0	1	5	0	5	25	0	3	92

Note. Rows correspond to stimuli and columns to responses. *Top row*: Observed frequency; *Middle row*: Predicted frequency of the biased choice model; *Bottom row*: Predicted frequency of the most efficient general recognition theory model.
[a] (angle level, size level).

Table 2
Summary of the Assumptions Underlying Each of the General Recognition Theory Models

Model	Perceptual independence?	Perceptual separability?	Decisional separability?
GRT(PI, PS, DS)	Yes	Yes	Yes
GRT(PI, PS_s, DS_s)	Yes	Only on size dimension	Only on size dimension
GRT(PI, DS)	Yes	No	Yes
GRT(PS, DS)	No	Yes	Yes
GRT(PI)	Yes	No	No
GRT(PS_s, DS_s)	No	Only on size dimension	Only on size dimension
GRT(DS)	No	No	Yes
GRT(●)	No	No	No

Note. GRT = general recognition theory; PI = perceptual independence; PS = perceptual separability; DS = decisional separability.

Next, a pair of stimuli was randomly selected and then displayed side by side for 400 ms. This was followed immediately by the mask, which was displayed until the subjects made a response. The two stimuli were about 1 cm apart. Horizontally, the visual angle of the entire display was about 2.5°. The subjects' task was to rate the similarity of the two stimuli on a scale from 1 to 10, with 1 meaning very dissimilar and 10 meaning very similar. There were 81 possible stimulus pairs. On the average, each pair was shown about 22 times during the course of the similarity sessions. The first day was considered practice, and data from that day were excluded from the analysis. Subjects responded on a keyboard with a row of buttons labeled from 1 to 10. There was a 2-s pause between trials and a 30-s pause between blocks. Each session consisted of 360 trials in four blocks of 90 trials.

Results and Theoretical Analyses

Identification

Confusion matrices (raw frequencies) for both subjects are presented in Table 1. The top number in each cell is the observed frequency. The bottom two numbers are predicted frequencies of models that are described below. Overall accuracy was 47.9% for Subject 1 and 36.9% for Subject 2.

Three classes of models were fit to the identification data: a number of models derived from GRT, the biased-choice model, and several MDS-choice models. The GRT and MDS-choice models all assumed a two dimensional perceptual representation.[3] The biased-choice model was included because it has been the best identification model for many years and so can serve as a benchmark for how well the other models accounted for the data (e.g., Smith, 1980; Townsend, 1971; Townsend & Ashby, 1982).

The GRT models that were tested are summarized in Table 2. Note that the different versions make different assumptions about perceptual independence (PI), perceptual separability (PS), and decisional separability (DS). The most restrictive model, GRT(PI, PS, DS), assumes all three of these conditions. Figure 3 shows contours of equal likelihood and decision bounds that might be predicted by GRT(PI, PS, DS). Decisional separability holds because each decision bound is parallel to a coordinate axis. Perceptual separability holds because the means fall in a rectangular formation and because the width of the ellipses is the same within every column and every row. Perceptual independence holds because the major and minor axes of each ellipse are parallel to the coordinate axes. For one stimulus, both means can arbitrarily be set to 0 and both variances to 1.0, and so GRT(PI, PS, DS) has only 12 free parameters. Of these, 4 determine the distribution means, 4 determine the variances, and 4 determine the decision bounds. We do not expect good fits from GRT(PI, PS, DS), because a preliminary analysis using techniques developed by Ashby (1988) turned up evidence that perceptual independence and separability were violated in the data of both subjects. Thus, more general versions of GRT are needed.

The other models described in Table 2 are all more general versions of GRT(PI, PS, DS). The most general model, GRT(●), does not assume perceptual independence, perceptual separability, or decisional separability. In all models not assuming decisional separability, the decision bounds were staircased, with two steps on each dimension.

Note that two of the models, GRT(PI, PS_S, DS_S) and GRT(PS_S, DS_S), assume separability only on the size dimension. The possibility of an asymmetric perceptual separability arises directly from an examination of Figure 1. Note that when judging size, the orientation of the line is irrelevant and so perceptual separability is feasible on this dimension. However, when judging orientation, circle size is relevant because the line is longer when the circle is large. Because of this, orientation judgments might be easier when the circle is large than when it is small. If so, we expect perceptual separability to fail on this dimension.

The eight GRT models described in Table 2 have a hierarchical structure in the sense that some contain others as a special case. This structure is illustrated in Figure 5. Any pair of models connected by arrows in Figure 5 are nested, with the upper model containing the lower as a special case.

Previous attempts to fit GRT models to identification data were restricted to versions that assume both PI and DS (Ashby & Perrin, 1988). Models that violate either of these assumptions are computationally more difficult to fit. However, we developed a fitting procedure that can accurately and reliably estimate the parameters of the most general versions of GRT.

[3] Although the stimuli have two obvious physical dimensions, models that assume three or more perceptual dimensions could be tested. However, the two dimensional models have more psychological plausibility, and in addition, they provide excellent fits to the data.

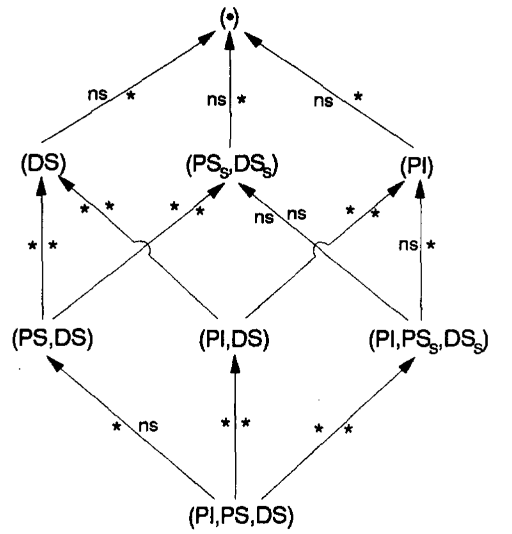

Figure 5. Hierarchical relation between the general recognition theory models fit to the data of Experiment 1. (PI = perceptual independence; PS = perceptual separability; DS = decisional separability.)

This procedure is described in the Appendix (along with a brief description of a method for fitting the biased-choice model).

Three versions of the MDS-choice model were fit: (a) the exponential–city-block model, (b) the Gaussian–Euclidean model, and (c) the exponential–Euclidean model. Recall that in general, Model (a) or (b) fits best when the dimensions are separable, and Model (c) fits best when the dimensions are integral. In addition, the Gaussian similarity function tends to fit best when the analysis is performed on the data of individual, well-practiced subjects. When the subjects have had little practice or when the data is aggregated over a number of subjects, the exponential similarity function tends to fit best (Shepard, 1986; Takane & Shibayama, in press). Because the dimensions of the Figure 1 stimuli are typically assumed to be separable and the data are from well-practiced individual subjects, we expect the Gaussian–Euclidean model to provide the best fit to the identification data.

All of these models were fit to the identification confusion matrices using an iterative search routine that minimized the sum of squared errors (*SSE*)

$$SSE = \sum_{i=1}^{9} \sum_{j=1}^{9} [f(R_j|S_i) - f_p(R_j|S_i)]^2$$

where $f(R_j|S_i)$ is the observed frequency with which Response R_j was given on trials when Stimulus S_i was presented and $f_p(R_j|S_i)$ is the corresponding theoretical prediction. Because many of the models are related in a nested fashion, it is possible to test whether the extra parameters of a more general model lead to a significant improvement in fit over the more restricted version. Let SSE_r and SSE_g refer to the *SSE* of a restricted and more general model, respectively, where the restricted model is a special case of the more general. Let df_r and df_g refer to the degrees of freedom associated with each model. Then under the null hypothesis that the restricted model is correct, the statistic

$$F_{obs} = \frac{(SSE_r - SSE_g)/(df_r - df_g)}{SSE_g/df_g}$$

has an approximate *F* distribution with $df_r - df_g$ degrees of freedom in the numerator and df_g degrees of freedom in the denominator[4] (e.g., Khuri & Cornell, 1987).

The SSE for the best fitting version of each model, along with its percentage of variance accounted for, is shown in Table 3. First, note that as predicted, the Gaussian–Euclidean version was the best fitting MDS-choice model. In fact, for both subjects, the Gaussian–Euclidean MDS-choice model

[4] Using a new algorithm of Wickens (in press), we repeated this analysis using the method of maximum likelihood. In this case, the difference in fit values of two nested models has a chi-square distribution, with degrees of freedom equal to the difference in the number of free parameters. In all cases, this second method of analysis led to exactly the same conclusions.

Table 3
Model Fits to Identification Data

Model	No. of parameters	Subject 1		Subject 2	
		SSE	% variance	*SSE*	% variance
Biased choice	44	224	99.4	1,948	92.7
EX/CB MDS	24	1,080	97.0	3,094	87.3
G/E MDS	24	316	99.1	2,291	91.2
EX/E MDS	24	1,254	96.5	3,421	86.0
GRT(PI, PS, DS)	12	1,673	95.3	4,398	82.1
GRT(PI, PS_s, DS_s)	28	208	99.4	817	96.7
GRT(PI, DS)	36	528	98.6	579	97.7
GRT(PS, DS)	21	1,188	96.7	3,394	86.2
GRT(PI)	44	179	99.5	195	99.2
GRT(PS_s, DS_s)	37	177	99.5	627	97.5
GRT(DS)	45	137	99.7	188	99.3
GRT(●)	53	100	99.7	78	99.7

Note. SSE = sum of squared errors; MDS = multidimensional scaling; GRT = general recognition theory; PI = perceptual independence; PS = perceptual separability; DS = decisional separability. "EX/CB MDS" refers to the exponential/city-block MDS-choice model. "G/E MDS" refers to the Gaussian/Euclidean MDS-choice model, and "EX/E MDS" refers to the exponential/Euclidean MDS-choice model.

was found not to fit significantly worse than the more general biased-choice model: for Subject 1, $F(20, 28) = .57$; for Subject 2, $F(20, 28) = .25$; in both cases, $p > .25$.

Second, note that GRT(PI, PS_S, DS_S) fit the data of both subjects better than the biased-choice model, despite the fact that it has 16 fewer free parameters. Third, although GRT(PI, PS_S, DS_S) has 4 more parameters than the MDS-choice models, it provides a better fit in the sense that generalizing the MDS-choice models by postulating extra stimulus dimensions, a different similarity function, a different distance metric, or attention weights will not allow the MDS-choice model to fit the data of either subject better than GRT(PI, PS_S, DS_S). This is because the MDS-choice model is a special case of the biased-choice model and so it can never have a lower *SSE* than the biased-choice model. Because GRT(PI, PS_S, DS_S) fits better than the biased-choice model, it must necessarily fit better than these generalized MDS-choice models.

Fourth, note that overall, the GRT models do an excellent job of accounting for the data of both subjects. In fact, the only versions that perform poorly are GRT(PI, PS, DS) and GRT(PS, DS). These are the only two versions that assume perceptual separability on both stimulus dimensions, and so their poor fit is strong evidence for at least an asymmetric perceptual integrality.

The relation between the various models is described in Figure 5. An asterisk on the left side of an arrow means that the more general model fitted significantly better (i.e., $p < .05$) than the more restricted model for Subject 1 and an asterisk on the right side means the difference was significant for Subject 2. An *ns* indicates a nonsignificant difference. The most efficient model provides the best fit with the fewest number of parameters. To find the most efficient GRT model, one begins at the bottom of Figure 5 and works upward. Whenever an asterisk is encountered, one continues upward; *ns* means that extra parameters have been added without a significant improvement in fit.

For Subject 1, note that GRT(PI, PS_S, DS_S) fits significantly better than GRT(PI, PS, DS) but that further generalizations provide no improvement. Another path moves through GRT(PI, DS) and ends at GRT(PI). However, because GRT(PI) contains GRT(PI, PS_S, DS_S) as a special case, relaxing the assumption of asymmetric separability provides no significant improvement in fit, and so GRT(PI, PS_S, DS_S) is a more efficient model for the data than GRT(PI). A third path proceeds through GRT(PS, DS) and ends at GRT(DS). Unfortunately, however, GRT(PI, PS_S, DS_S) is not a special case of GRT(DS) and so we have no way to compare these two models statistically. GRT(DS) provides a better fit, but at the cost of 17 additional parameters. If the two models were nested, the improvement in fit would be nonsignificant ($p > .25$), which suggests that GRT(PI, PS_S, DS_S) may be the more efficient model.

For Subject 2, the interpretation of the model fits is straightforward. With two exceptions, every generalization led to a significant improvement in fit. Thus, the most efficient model for describing the data of Subject 2 is GRT(●).

On the basis of these results, we conclude that PS and DS are violated for Subject 2 on both dimensions and for Subject 1 only on the orientation dimension. In addition, it appears that PI holds for Subject 1 but not for Subject 2. Although these results support our intuitive rationale for an asymmetric violation of separability, they contradict current ideas about these stimulus dimensions. The GRT fits to Nosofsky's (1985b) identification confusion matrices to be reported below allow an opportunity to substantiate these results.

The parameter estimates of the best fitting models can be effectively studied by examining their predicted contours of equal likelihood. Figure 6 shows the contours and decision bounds predicted by GRT(PI, PS_S, DS_S) for Subject 1 and by GRT(●) for Subject 2. Note that in both cases, PS is strongly violated on the orientation dimension, especially for the middle level of orientation. In particular, as size increases, perceived orientation increases. The Subject 1 model assumes PS on the size dimension, but in the case of Subject 2, note that the violations of PS on the size dimension are small. In summary, both subjects exhibited substantial violations of PS and DS on the orientation dimension, and Subject 2 exhibited small but significant violations of PS and DS on the size

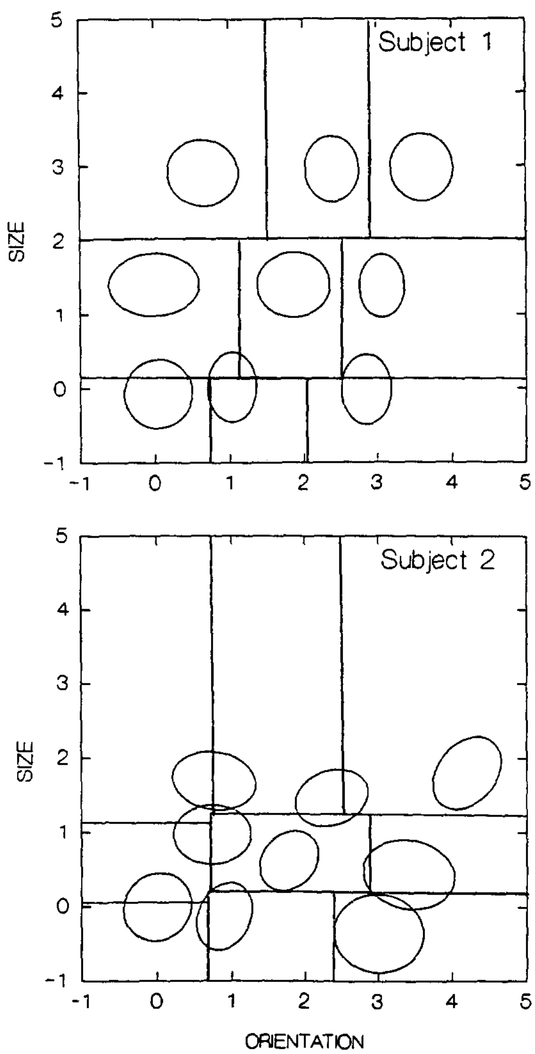

Figure 6. Decision bounds and contours of equal likelihood that describe the parameter estimates of GRT(PI, PS_S, DS_S) for Subject 1 and GRT(●) for Subject 2. (GRT = general recognition theory.)

dimension. In addition, Subject 2 also displayed moderate violations of PI.

Table 1 also contains the predictions of the best fitting biased-choice model and the best fitting versions of GRT, namely, GRT(PI, PS_S, DS_S) for Subject 1 and GRT(●) for Subject 2. Note that for Subject 1, neither model has any large prediction errors, but for Subject 2, the biased-choice model poorly predicts a number of cells in the confusion matrix. Specifically, it predicts that subjects will never respond R_6 on trials when Stimulus S_9 is presented, even though this was the most common error on S_9 trials. Similarly, it incorrectly predicts almost no R_2 errors on S_5 trials. It is interesting to note that many of the largest errors of prediction involve pairs of stimuli that differ on only one stimulus dimension.

Similarity Judgments

The data from this part of the experiment consisted of ordinal ratings, ranging from 1 to 10, with 10 indicating the greatest similarity. Subjects were instructed to judge the similarity between the stimulus on the left and the stimulus on the right. Specifically, they were not instructed to consider one of the stimuli as the standard or the referent. Because of this, all data analyses ignored presentation order. Thus, we assume $s(S_i, S_j) = s(S_j, S_i)$.

All models described above assume that ratings of judged similarity agree only monotonically with perceived similarity. Thus, we are only interested in the ordinal agreement between the model predictions and the observed similarity ratings. The first step, therefore, is to rank order the stimulus pairs according to their judged similarity. Next, predicted similarities are generated from the various models, these are rank ordered, and finally, the observed rank orders are compared with the predicted rank orders.

Rank ordering the stimulus pairs according to their judged similarity turns out not to be a trivial problem. For each stimulus pair, each subject made approximately 44 similarity judgments. Because it is not meaningful to compute means for ordinal data, and because an examination of the medians indicated a large number of ties, we developed an alternative method for rank ordering the judged similarities. Let R_{AB} represent the subject's rating of the similarity of Stimuli S_A and S_B and let $Pr\ (R_{AB} \leq i)$ denote the proportion of times that the subject's similarity rating for the (S_A, S_B) pair was less than or equal to i. Then we assumed that the pair (S_A, S_B) was judged by the subject to be at least as similar as the pair (S_C, S_D) if

$$\sum_{i=1}^{10} Pr\ (R_{AB} \leq i) \leq \sum_{i=1}^{10} \mathrm{Pr}\ (\mathrm{R}_{CD} \leq i). \tag{11}$$

Note that the maximum value of the sum on the left side of Equation 11 is 10.0 (when all proportions equal 1.0) and the minimum value is 1.0. (Because the rating scale ranged from 1 to 10, it is always true that $Pr[R_{AB} \leq 10] = 1$). The maximum value is achieved when all similarity ratings equal 1; that is, when the stimulus pair is maximally dissimilar and the minimum value is achieved when all ratings equal 10 (i.e., when the stimulus pair is maximally similar). Thus, Equation 11 establishes a weak order on the similarity judgments, which guarantees that the stimulus pairs can be rank ordered according to their judged similarity (e.g., Krantz, Luce, Suppes, & Tversky, 1971). The empirical rank orderings are presented in Table 4. The most similar pair is ranked 1, and the least similar pair is ranked 45.

Table 4
Observed Similarity Rankings

Stimulus no.	Stimulus no.								
	1	2	3	4	5	6	7	8	9
					Subject 1				
1	15	6	10	22	27	31	45	40	39
2	—	11	14	28	21	23	43	37	41
3	—	—	3	32	34	33	44	42	38
4	—	—	—	13	17	16	19	36	35
5	—	—	—	—	4	2	29	26	25
6	—	—	—	—	—	18	20	24	30
7	—	—	—	—	—	—	1	5	12
8	—	—	—	—	—	—	—	9	8
9	—	—	—	—	—	—	—	—	7
					Subject 2				
1	6	12	18	29	23	33	42	40	44
2	—	7	13	30	27	35	38	43	41
3	—	—	3	32	25	22	45	39	37
4	—	—	—	8	10	24	17	31	34
5	—	—	—	—	4	9	26	21	19
6	—	—	—	—	—	2	36	16	20
7	—	—	—	—	—	—	5	15	28
8	—	—	—	—	—	—	—	14	11
9	—	—	—	—	—	—	—	—	1

MDS Analysis

Before examining the ability of the various models to predict the similarity data from the results of the identification condition, it is instructive to submit the similarity data to a standard MDS analysis. Performing a nonmetric MDS (e.g., Kruskal, 1964a, 1964b) on the Table 4 similarities is equivalent to estimating a pair of coordinate values for each stimulus such that the rank ordering of the distances between stimulus points agrees as closely as possible with the rank ordering of the stimuli by judged similarity. In essence, MDS programs attempt to account for the similarity data by fitting a model with 14 free parameters to the Table 4 similarity matrix.[5] In contrast, the identification models have no free parameters in this application because their parameter values were all fixed when they were fit to the identification data. Because of the many free parameters of the MDS-similarity model, this approach should account for the similarity data substantially better than any of the identification models.

Thus, an MDS-similarity analysis has a number of advantages. First, the resulting goodness-of-fit value provides a realistic upper bound for the identification models. Second,

[5] Although nine stimuli have 18 coordinate values in a two dimensional representation, without loss of generality, 4 coordinate values can be fixed.

if one assumes that the MDS model is correct, then the MDS solution obtained from fitting the MDS-choice model to the identification data can be compared with the solution obtained from fitting the MDS-similarity model to the similarity data. This exercise should provide useful insights into the relation between identification and judged similarity.

The MDS-similarity model was fit to the Table 4 similarity matrix using SYSTAT's MDS routine (Wilkinson, 1989). Two separate models were fit to each subject's data; one was based on city-block distance and the other on Euclidean distance (i.e., the simple Euclidean MDS model). The predicted similarities were rank ordered, and then the rank order correlation (i.e., Kendall's tau correlation) between the observed and predicted rankings were computed. The results are shown in the top of Table 5.

There are two points of interest about these values. First, although the differences are not statistically significant ($p > .10$, in each case), note that for both subjects the city-block metric provides a better fit than the Euclidean metric. This finding agrees with previous research that has found the city-block metric to provide better fits when the stimulus dimensions are separable (e.g., Attneave, 1950; Garner, 1974). Second, note that the MDS-similarity model is unable to account for a large amount of variability in the data of both subjects. This is especially true for Subject 1. The fit values in Table 5 seem especially poor when one recalls that the best fitting identification models accounted for 99.7% of the variance in the identification data collected from these same two subjects. There are two possibilities: (a) The MDS model of similarity is incorrect. (b) Judged similarity data is noisier than identification data.

There is a good deal of evidence that MDS is an incorrect theory of perceived similarity. First, a number of researchers have found extensive violations of interdimensional additivity, a property assumed by both the city-block and Euclidean MDS models (Krantz & Tversky 1975; Nygren, 1979; Tversky & Krantz, 1969). Second, Tversky (1977; see also Krumhansl, 1978) forcefully argued that perceived similarity also violates the distance axioms, thereby falsifying all models that represent perceived dissimilarity by the distance between stimulus points. On the other hand, MDS solutions often have psychological validity (e.g., Cliff, 1973; Jones & Young, 1972; Shepard, 1963; Shepard & Chipman, 1970). Therefore, even if the model is incorrect, it should account for a substantial amount of variability in the present data set.

An examination of the raw (i.e., trial by trial) data supports the hypothesis that rated similarity is inherently noisy. For example, on trials when the pair (S_1, S_1) was presented, Subject 1 gave similarity responses that ranged from a low of 1 to a high of 10. Subject 2's responses ranged from 4 to 10. When the most dissimilar pair, (S_1, S_9), was presented, the responses ranged from 1 to 7 for Subject 1 and from 1 to 8 for Subject 2. This variation is much greater than that which occurred during the identification condition. For example, Subject 1 never confused S_1 and S_9, and Subject 2 only confused this pair once in 258 trials. Perhaps the similarity judgments were noisier than the identification responses because the identification task has more ecological validity than the similarity judgment task. Outside of the laboratory, most subjects will have less experience producing a numerical estimate of the similarity of a pair of stimuli than they will identifying or categorizing the same stimuli (e.g., Nosofsky, 1985b).

The MDS solutions for each subject are shown in Figure 7. At the top are the perceptual representations according to the MDS-similarity model with Euclidean distance and at the bottom are the representations according to the Gaussian–Euclidean MDS-choice model (i.e., when fit to the identification data). Note the striking difference between the similarity and identification representations. In particular, interpoint distances along the orientation dimension are much smaller in the similarity representation. In fact, for Subject 1, all stimuli of the same size have almost identical coordinates, regardless of orientation. This result is strong evidence that subjects allocated attention differently in the two tasks. Specifically, it appears that both subjects divided attention equally between the two dimensions in the identification task, but that when judging similarity, they both allocated more attention to the size dimension.

What could cause such differences in selective attention? One possibility is that on identification trials, the subject must pay equal attention to both stimulus dimensions or accuracy will suffer. On similarity trials, however, there is no objectively correct response, and so it makes sense that subjects will adopt a strategy that conserves mental effort. Introspective reports of our subjects indicated that the stimulus size differences were more salient than the orientation differences. Therefore, it seems plausible that when judging similarity, subjects allocated more attention to the salient size dimension. This hypothesis seems even more reasonable in light of the 400-ms stimulus exposure durations and the poststimulus mask. With limited stimulus information, unmotivated subjects may be especially likely to selectively attend to salient stimulus dimensions.

Table 5
Kendall's Tau Rank-Order Correlations Between Observed and Predicted Similarity Rankings

Model	Subject 1	Subject 2
MDS-similarity		
City-block distance	.772	.864
Euclidean distance	.743	.860
Identification		
Biased-choice	.529	.549
GRT	.568	.575
G–E MDS-choice	.541	.593
Attention		
Ignoring orientation		
GRT	.748	.592
G–E MDS-choice	.667	.634
Attention parameter		
GRT	.683	.749
G–E MDS-choice	.669	.679

Note. MDS = multidimensional scaling; GRT = general recognition theory; G–E = Gaussian–Euclidean. GRT(PI, PS_s, DS_s) for Subject 1 and GRT(●) for Subject 2.

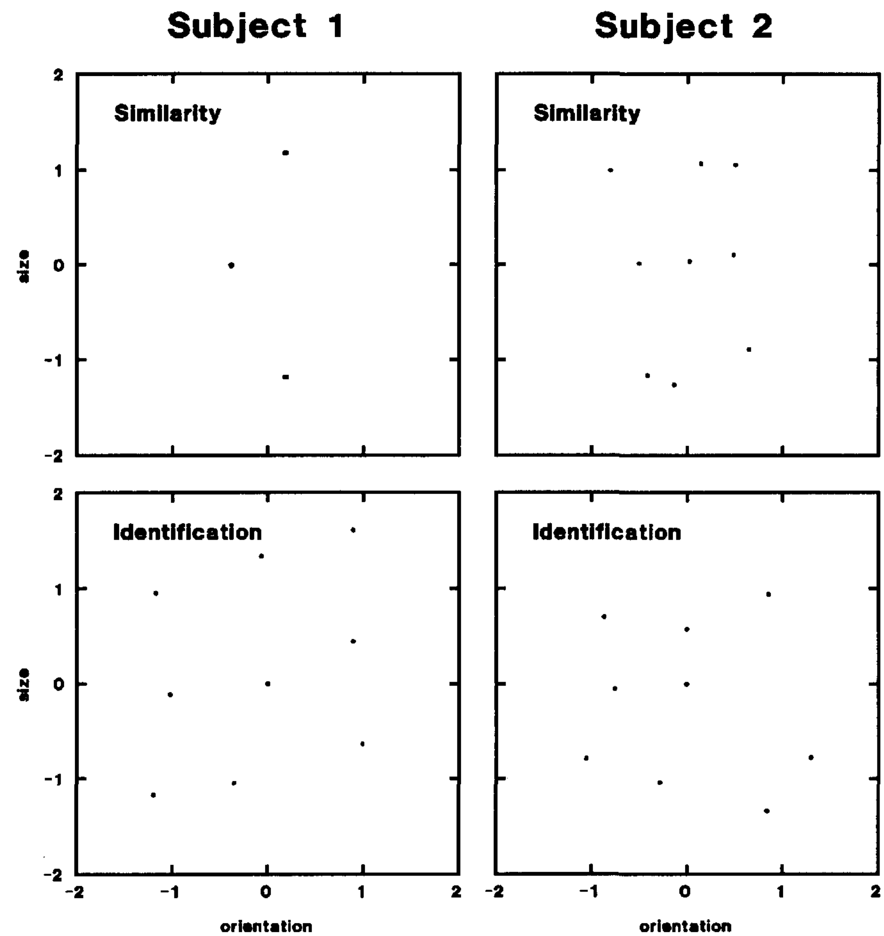

Figure 7. MDS representations (Euclidean distance) of the stimuli used in Experiment 1, as derived separately from the identification and similarity data.

Predictions of Similarity Models

As a first test of the similarity models, predicted similarity judgments were derived directly from the parameter estimates obtained when the various models were fit to the identification confusion matrices. During this procedure, no new parameters were estimated. For the biased-choice model, the predicted similarity between S_i and S_j is simply $s(S_i,S_j) = \eta_{ij}$. Recall that the biased-choice model sets $\eta_{ii} = 1$, for all values of *i*.

For the MDS-choice models, $s(S_i, S_j) = \exp(^{-}d_{ij}^{\alpha})$, where d_{ij} is the distance between the perceptual representations of Stimuli S_i and S_j, and $\alpha = 1$ or $\alpha = 2$, depending on the version of the model. The predicted similarities depend on whether the Euclidean or city-block distance metric is used, but they do not depend on the form of the similarity function (i.e., on α). Because the Gaussian similarity function is a monotonic transformation of the exponential similarity function, and because we are only interested in the rank order predictions of the various models, the two similarity functions make identical predictions (assuming they start with the same set of distances).

The similarity predictions of the GRT models are generated from Equation 6. The first step in this procedure is to reset the decision bounds estimated from the confusion matrices so that all response biases are eliminated. This was done under a local constraint of decisional separability. Globally, the unbiased decision bounds were step functions on each dimension. Unlike the biased-choice model and the MDS-choice models, the GRT models are not constrained to make symmetric similarity predictions. Most GRT models predict $s(S_i, S_j) \neq s(S_j, S_i)$. However, because the instructions ignored presentation order, the predictions of the GRT models for the similarity between Stimulus S_i and Stimulus S_j were taken as the average of the $s(S_i, S_j)$ and the $s(S_j, S_i)$ predictions.

Table 5 (under Identification Models) shows the rank order (Kendall's tau) correlations between the observed similarity rankings and the predicted rankings derived from the biased choice model, the Gaussian–Euclidean MDS-choice model, and the GRT model that best accounted for each subject's identification data, namely, GRT(PI, PS_S, DS_S) for Subject 1 and GRT(●) for Subject 2. Three aspects of Table 5 stand out. First, all rank order correlations are significantly greater than 0 ($p < .001$). In fact, all correlations are greater than .5. This seems quite good considering the fact that no new parameters were estimated during this phase of the data analysis. Second, note that for each subject, the difference between the smallest and largest correlation is less than .05. None of these differences is statistically significant (all $ps > .25$), and thus none of the models perform significantly better than any of their competitors.[6] Unlike the identification condition, all models predicted the similarity data approximately equally well. Third, all identification models fit substantially worse than the MDS-similarity model.

The poor performance of the identification models relative to the MDS-similarity model is not surprising given the potential shift in selective attention between the two tasks. The identification models should perform better if we allow them to allocate attention differently in the two tasks. We did this in two different ways. First, we assumed subjects completely ignored the orientation dimension. In this case, the MDS-choice models assume that the similarity between a pair of stimuli is completely determined by (the absolute value of) the difference between their coordinates on the size dimension. The GRT models assume that the relevant statistic is the overlap of their marginal perceptual distributions on the size dimension. To the extent that the Figure 6 MDS-similarity solution is valid, this approach should work particularly well for Subject 1.

The second way we allowed the identification models to allocate attention differently in the two tasks was by constructing versions of the models that had one free attention parameter. Specifically, this parameter measured the proportion of attention allocated to the size dimension in the similarity condition. In the GRT models, the attention parameter is a constant, w, that multiplies all variances on the size dimension (Ashby & Perrin, 1988). A value of $w < 1$ indicates that more attention was allocated to the size dimension than to the orientation dimension. A fitting procedure determined the value of this parameter that maximized the rank order correlation between the model's predictions and the observed similarity judgments.

The resulting rank order correlations are shown at the bottom of Table 5. For Subject 2, note that both models with a free attention parameter provided better fits than their counterparts that assumed all attention was focused on the size dimension. This agrees with the MDS solution shown in Figure 7. Also note that the GRT model performed better than the MDS-choice model. For Subject 1, the best fit is by the version of GRT(PI, PS_S, DS_S) that assumed all attention was focused on the size dimension. Note that this model actually predicts the similarity data better than the simple Euclidean MDS-similarity model, even though the GRT model has no free parameters in this application and the MDS-similarity model has 14. This is a striking result. It is only possible if (a) there is a close relation between similarity and identification and (b) distributional overlap is a better measure of perceived similarity than simple Euclidean distance.

Nosofsky's (1985b, 1986) Identification–Categorization Experiment

This section focuses on the identification–categorization relationship. An ideal application replicates Experiment 1, except that during Phase 2, subjects make categorization responses rather than similarity judgments. Our goal was to predict performance in the categorization conditions on the basis of performance in the identification task. Fortunately, Nosofsky (1985b, 1986) ran exactly this experiment.

Method

The stimuli Nosofsky used were similar to those in Figure 1. However, instead of a 3 × 3 stimulus set, he used 16 stimuli arranged in a 4 × 4 configuration. In addition, the bottom horizontal line of each stimulus was absent. As in Experiment 1, data were collected from 2 subjects. Each subject made about 210 identification responses to each stimulus. (See Table 1 of Nosofsky, 1985b, for the confusion matrices.)

In the categorization phase, 4 of the 16 stimuli were assigned to Category 1 and 4 were assigned to Category 2. The following four conditions, illustrated in Figure 8, were included: (a) Dimensional: Category 1 contains small circles and Category 2 contains large circles; (b) Criss-cross: Category 2 contains exemplars with small sizes and small orientations and exemplars with large sizes and large orientations, whereas Category 1 contains exemplars with small sizes and large orientations and exemplars with large sizes and small orientations; (c) Interior–exterior: Category 1 contains exemplars with intermediate values on each dimension, whereas Category 2 contains exemplars with an extreme value on one of the two dimensions; and (d) Diagonal: a diagonal bound with a slope of approximately −1 separates the two categories. The exemplars of Category 1 lie below the bound and the exemplars of Category 2 lie above the bound.

Each categorization condition consisted of a learning phase followed by a transfer phase. During the learning phase, only the 8 stimuli that had been explicitly assigned to one of the two categories were presented (with feedback). During the transfer phase, all 16 stimuli were each presented about 220 times, but feedback was given only on trials when one of the eight training exemplars was presented. The data of interest are the categorization confusion matrices collected during the transfer phase. (See Table 3 of Nosofsky, 1986, for the confusion matrices.)

Nosofsky (1986) showed that by allowing for shifts in selective attention between the two tasks, the GCM did a reasonable job of predicting the categorization data from the fits to the identification confusion matrices. In the dimensional and diagonal conditions, the GCM did an excellent job, accounting for more than 98% of the variance in the data of both subjects. In the criss-cross and interior–exterior conditions, however, the fits were less satisfactory. For example, in the interior–exterior condition, the GCM accounted for only 85% and 75% of the variance in the data of Subjects 1 and 2, respectively.

Results and Discussion

Identification Condition

As in Experiment 1, the first step of the data analysis was to fit the various models to the identification confusion mat-

[6] None of the other identification models listed in Table 3 performed significantly better. GRT(PI, PS_S, DS_S) yielded the highest rank order correlation for Subject 1 and GRT(PI, PS, DS) yielded the highest correlation for Subject 2 (i.e., $\tau = .622$).

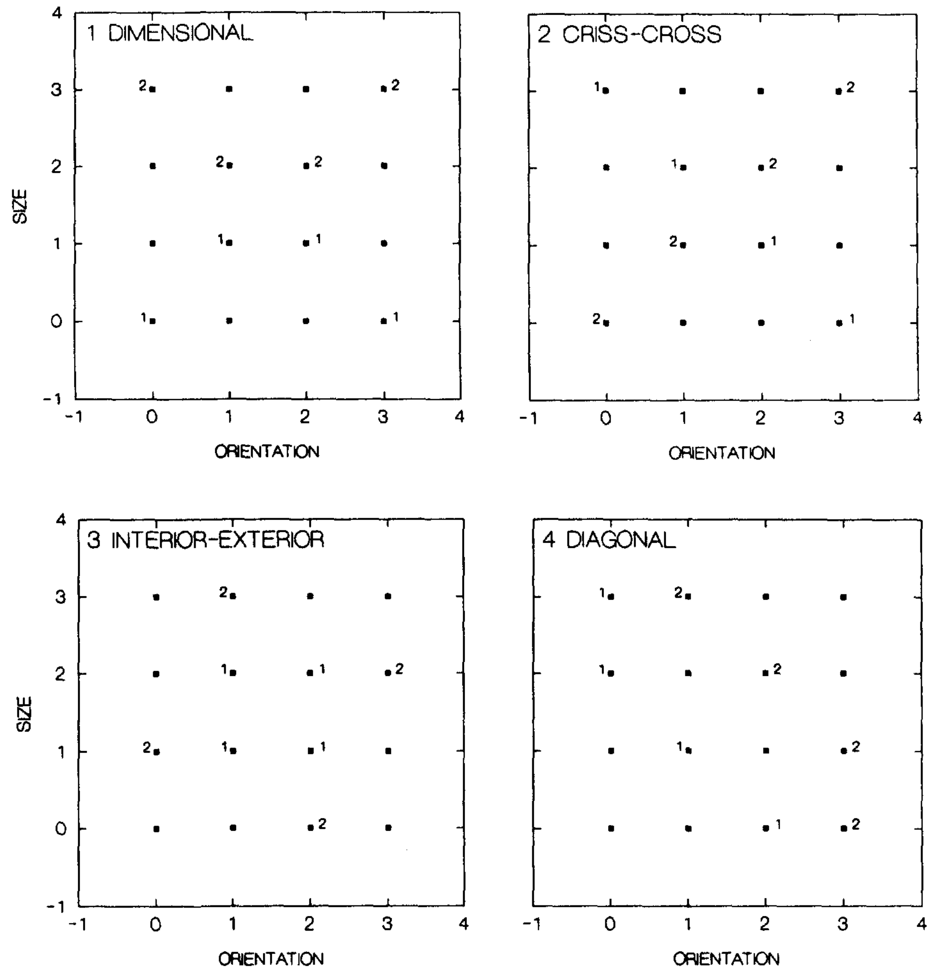

Figure 8. The four experimental conditions of Nosofsky's (1986) categorization experiment.

rices. Because Nosofsky's (1985b, 1986) experiment involved 7 more stimuli than the identification-similarity experiment reported above, all of the models had many more free parameters in this application. For example, with 16 stimuli the biased-choice model has 135 free parameters, whereas with 9 stimuli it has only 44. Because this data set is so large, we fit only a subset of the GRT models described in Figure 5. The nested relation between the various GRT models is illustrated in Figure 9.

The fits of the various models to the identification confusion matrices are described in Table 6. Note that the pattern of results is very similar to Experiment 1. The best fitting MDS-choice model again assumed a Gaussian similarity function and a Euclidean distance metric, and again the fit of this model was not significantly worse than the fit of the biased-choice model. In addition, as before, the best fitting GRT models substantially outperformed the biased-choice model. Because of the larger stimulus ensemble of this experiment, however, the GRT models have many fewer parameters than the biased-choice model, and so interpretation of these results is simpler than in Experiment 1. Specifically, note that the *SSE* associated with GRT(●) is about 4 times smaller than

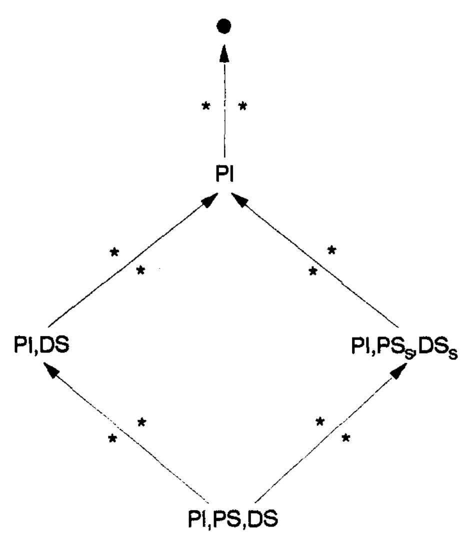

Figure 9. Hierarchical relation between the general recognition theory models fit to the identification data of Nosofsky (1985b). (PI = perceptual independence; PS = perceptual separability; DS = decisional separability.)

Table 6
Model Fits to Nosofsky's (1985) Identification Data

Model	No. of parameters	Subject 1		Subject 2	
		SSE	% variance	*SSE*	% variance
Biased-choice	135	821	99.5	1,544	98.4
G-E MDS	45	1,096	99.3	2,147	97.8
GRT(PI, PS, DS)	18	4,676	97.0	8,572	91.2
GRT(PI, PS_s, DS_s)	51	1,478	99.0	2,922	97.0
GRT(PI, DS)	66	1,228	99.2	4,020	95.9
GRT(PI)	84	282	99.8	1,340	98.6
GRT(●)	100	219	99.9	911	99.1

Note. *SSE* = sum of squared errors; G-E = Gaussian-Euclidean; MDS = multidimensional scaling; GRT = general recognition theory; PI = perceptual independence; PS = perceptual separability; DS = decisional separability.

the *SSE* associated with the biased-choice model for Subject 1 and about 40% smaller for Subject 2, even though GRT(●) has 35 fewer parameters than the biased-choice model. In fact, note that GRT(PI) has 51 fewer parameters than the biased-choice model and yet GRT(PI) fits better for both subjects. The excellent fits of the GRT models in these two experiments provide additional support for GRT, and they pose a significant challenge to the biased-choice model and the multidimensional scaling approach to modeling identification.

As Figure 9 indicates, each time the GRT model is generalized, a significant improvement in fit occurs ($p < .01$). This suggests that both subjects violated PS and DS on both stimulus dimensions and that PI was violated for at least some of the stimuli. An examination of the parameter estimates from the best fitting models indicated that for Subject 1, violations of PI and PS were small. However, Subject 1 did exhibit a substantial violation of DS on the orientation dimension. Subject 2 displayed larger violations of PI and PS but smaller violations of DS. In addition, in agreement with Experiment 1, both subjects showed a slight tendency for perceived orientation to increase with perceived size.

Categorization Conditions

Nosofsky (1986) reported the results of fitting the GCM with a Gaussian similarity function and the Euclidean distance metric to his categorization data. Three parameters were estimated: a response bias β_A, an attention parameter w, and a discriminability parameter c (see Equations 8–10; for more details see Nosofsky, 1986). Of these parameters, Nosofsky emphasized the role of attention in accounting for the identification–categorization relation.

Allocating more attention to a stimulus dimension has the effect of stretching the perceptual space along that dimension. This operation facilitates categorization only if it decreases between-category similarity. In the dimensional condition, and to a lesser extent in the diagonal condition, a shift in selective attention should have precisely this effect. However, in the criss-cross and interior–exterior conditions, it is unclear why stretching one of the perceptual dimensions should improve categorization performance. As predicted by this hypothesis, the GCM performed best in the dimensional condition, next best in the diagonal condition, and worst in the criss-cross and interior-exterior conditions.

According to GRT, however, the most important difference between identification and categorization is that the two require very different sets of decision bounds. Shifts in selective attention may occur, but these often play a minor role in the identification–categorization relation. It should be instructive then to see how well we can account for the categorization data without appealing to the concept of selective attention.

We began by assuming that the perceptual representations that are estimated when GRT(●) was fit to the identification confusion matrices accurately describe the percepts during the categorization conditions. Next, we assumed that during the training phase, subjects learned to assign categorization labels in an optimal fashion. That is, by the end of the learning phase and all through the transfer phase, subjects used the decision bound that maximized categorization accuracy. This is a rather naive assumption, which is, at best, only approximately correct. However, because it specifies decision bounds without the need to estimate any free parameters, it represents a convenient benchmark against which to test the other models. We call this the optimal GRT(●) model.

Starting with this optimal model, we also constructed a two parameter GRT(●) categorization model. Because the assumption that subjects exactly use the optimal bound is surely incorrect, most desirable are parameters that add flexibility to the decision bound. The first parameter is a response bias. In the optimal model, Response R_A is given whenever the likelihood ratio $l(x, y) = f_A(x, y)/f_B(x, y) > 1.0$. With a response bias, Response R_A is given whenever $l(x, y) > \beta$, for some constant β. When $\beta < 1$, the subject is biased toward Response R_A and when $\beta > 1$, the subject is biased toward Response R_B. A response bias directly shifts the decision bound. Often the biased decision bound (i.e., $\beta \neq 1$) is nearly parallel to the unbiased bound (i.e., $\beta = 1$). The second parameter represents the effects of the extra experience with the eight stimuli presented during the learning phase. In GRT, such extra experience is assumed to affect the perceptual variability along both stimulus dimensions. Thus, the second parameter, γ, is a constant that multiplies both variances in each of the eight training stimuli. The Parameter γ affects the decision bound in a manner similar to the response bias, except that the $\gamma \neq 1$ bound is less likely to be parallel to the $\gamma = 1$ bound.

Table 7
Model Fits to Nosofsky's (1986) Categorization Data

Model	No. of parameters	Subject 1 SSE	Subject 1 % variance	Subject 2 SSE	Subject 2 % variance
Dimensional condition					
Optimal GRT(●)	0	.008	99.72	.158	92.84
Optimal GRT(PS)	0	.006	99.80	.118	94.68
GRT(●)	2	.007	99.78	.070	96.98
GRT(PS)	2	.004	99.88	.066	97.14
GCM	3	.002	99.93	.017	99.21
Criss-cross condition					
Optimal GRT(●)	0	.041	97.86	.185	86.29
Optimal GRT(PS)	0	.052	97.21	.090	93.20
GRT(●)	2	.040	97.94	.167	86.56
GRT(PS)	2	.050	97.30	.074	93.98
GCM	3	.095	94.73	.159	86.97
Interior–exterior condition					
Optimal GRT(●)	0	.325	72.84	.318	68.88
Optimal GRT(PS)	0	.188	83.65	.209	80.84
GRT(●)	2	.157	88.75	.263	75.12
GRT(PS)	2	.077	92.72	.117	88.86
GCM	3	.126	84.52	.208	74.84
Diagonal condition					
Optimal GRT(●)	0	.078	97.93	.048	98.11
Optimal GRT(PS)	0	.056	98.54	.088	96.32
GRT(●)	2	.097	95.87	.033	98.63
GRT(PS)	2	.019	99.14	.032	98.68
GCM	3	.034	98.31	.043	98.19

Note. SSE = sum of squared errors; GRT = general recognition theory; PS = perceptual separability; GCM = generalized context model.

In addition to these two GRT models, a second but similar pair was constructed. The only difference was that the second pair assumed PS on both dimensions. The rationale for testing these two models is that the categorization conditions described in Figure 8 emphasize the dimensional structure of the stimuli. The principal exception is the diagonal condition, but note that even this condition can be described by a dimensional rule with one exception in each category (Nosofsky, 1986). This raises the interesting hypothesis that the dimensional structure of the stimuli may be more apparent in the categorization conditions than in the identification conditions.

Therefore, we constructed an optimal GRT(PS) model, which is identical to the optimal GRT(●) model except that perceptual separability is satisfied.[7] A similar relation exists between the GRT(PS) categorization model and the GRT(●) categorization model. Table 7 presents the results of the categorization fits for the four GRT models[8] and for the GCM. The fits of the GCM are from Nosofsky (1986).

First, note that the GRT(PS) models generally performed better than the GRT(●) models. The optimal GRT(PS) model fits better than the optimal GRT(●) model in six of the eight cases, and the two-parameter GRT(PS) model fits better than the two-parameter GRT(●) model in six of eight cases (in one additional case, they fit equally well). This is tentative support for the hypothesis that the specific nature of the categories chosen by Nosofsky (1986) caused the subjects to more strongly perceive the dimensional structure of the stimuli.

Second, note that even though it has no free parameters, the optimal GRT(PS) model does surprisingly well when compared with the three parameter GCM. In fact, with respect to SSE, it actually outperforms the GCM in two cases, and in one other case, it performs only slightly worse. With respect to the percentage of variance statistic, it performs even better, outperforming the GCM in half the cases. Thus, given the identification confusion matrices, the optimal GRT(PS) model is able to predict a priori the results of the categorization conditions almost as well as the three parameter GCM.

[7] This was done by averaging the means and variances of the GRT(●) identification model within each of the stimulus rows and columns illustrated in Figure 1.

[8] Following Nosofsky (1986), all GRT models were fit to the data using a maximum likelihood criterion (i.e., the fits minimized minus log likelihood). However, because of rounding error, the GRT models predicted zero entries in some of the cells of the categorization confusion matrices, and thus the absolute value of the final minus log likelihood statistic is meaningless. For this reason, those values are not reported here. The *SSE* and percentage of variance accounted for statistics reported in Table 8 were computed from the maximum likelihood fits (again following Nosofsky, 1986) and so should be interpreted with caution.

This result is important for several reasons. First, it supports the validity of the GRT(●) model of the identification data. If our parameter estimation procedure had led to incorrect estimates of the perceptual distributions, it is unlikely that the categorization data could have been predicted so accurately. Second, it reemphasizes the close relation between identification and categorization noted by Nosofsky (1986). Third, it affirms the important role of the decision bound in the identification–categorization relation.

Table 7 indicates that, for both subjects, the optimal GRT models did worst in the interior–exterior condition. An examination of the categorization confusion matrices from this condition indicates one reason (see Table 3 of Nosofsky, 1986). During the transfer phase, both subjects were correct on Stimulus 14 only half the time, even though Stimulus 14 was a training exemplar for Category 2. The optimal model predicts above chance performance on all training exemplars.

Finally, note that the two parameter GRT(PS) model fits better than the three parameter GCM for both subjects in the criss-cross, the interior–exterior, and the diagonal conditions. In the dimensional condition, both models account for about 99.9% of the variance in the data of Subject 1, and the GCM fits better for Subject 2. Thus, with the possible exception of the dimensional condition, it is not necessary to appeal to shifts in selective attention to account for the identification–categorization relation. A simpler model that emphasizes the different decision bounds required in the two tasks is adequate.

In an effort to improve the performance of the GCM in the categorization conditions, Nosofsky (1986) hypothesized that certain stimuli, which have not previously been assigned to a particular category, are nevertheless implicitly categorized in the memory representation. In other words, in addition to assigning a category membership to the training stimuli, the subject is assumed to infer the category membership of each stimulus not presented during training. Nosofsky (1986) called this the *augmented generalized context model* (AGCM). Many versions can be formulated depending on which category each inferred exemplar is assigned. Although technically the AGCM has the same number of free parameters as the GCM, determining the best partition of the stimulus ensemble involves a process much like parameter estimation. In the criss-cross, interior–exterior, and diagonal conditions, the AGCM performed substantially better than the GCM, accounting for at least 88% of the variance in each case. In the dimensional condition, the two models performed about equally well. More general versions of the GRT model could also be developed. Perhaps the most obvious method of doing this is to add extra parameters that allow the decision bound even more flexibility. We constructed and tested such models and found that they accounted for the categorization confusion matrices as well or better than the AGCM. These analyses are not reported here because our goal is not to obtain the absolute smallest *SSE* but to illustrate the predictive validity of decision bounds. The results in Table 7 achieve this goal.

Summary and Conclusion

The goal of this article was to study the relationship of identification to similarity judgment and categorization. The theoretical analysis focused on GRT but the GRT models were also compared with the MDS-choice models. The strategy was to fit the various models to the identification confusion data and to use the resulting parameter estimates to predict performance in the similarity judgment and categorization conditions.

The identification–similarity experiment reported above involved a total of nine stimuli. This number is small enough so that it is feasible to fit very general versions of GRT, yet large enough so that the parameters of these general models are identifiable. The simpler versions of GRT assumed PI and DS. As in past applications (Ashby & Perrin, 1988) these models fit about as well as the biased-choice model and the best fitting MDS-choice model. The more general versions of GRT allowed violations of PI and DS. These models have not previously been fit to identification data. For both subjects in Experiment 1, some version of these more general GRT models fit the empirical confusion matrices better than the biased-choice model or any version of the MDS-choice model. For Subject 2, this difference was substantial: the *SSE* for GRT(●) was about 25 times smaller than the *SSE* of the biased-choice model. The results of fitting the various models to the identification confusion matrices reported by Nosofsky (1985b) were similar to the results of Experiment 1. Specifically, GRT(●) again substantially outperformed the biased-choice model. This result is especially significant because in this application, GRT(●) has 35 fewer parameters than the biased-choice model. To our knowledge, GRT is therefore the first model to outperform the biased-choice model in an identification experiment.

The parameter estimates from the best fitting models suggest at least an asymmetric perceptual integrality, namely, that perceived orientation depends on size. In addition, in both experiments, evidence was obtained for a corresponding violation of DS on the orientation dimension. These findings contradict the conventional thinking about these stimulus dimensions, but they explain several apparently puzzling results. For example, using a speeded classification task, Garner and Felfoldy (1970) found a redundancy gain when subjects classified on the orientation dimension but no redundancy gain when they classified on the size dimension. Despite this evidence of asymmetric separability, Garner and Felfoldy concluded that size and orientation are mutually separable.

In addition to the violations of separability, the model that best fit the data of Subject 2 in both experiments also exhibited violations of PI. (Subject 1 of Nosofsky, 1985b, also violated PI, but only slightly). In both experiments, the biased-choice model fit the data of Subject 1 better than the data of Subject 2. Although it may be coincidence, this raises the possibility that a good fit of the choice model requires separability, PI, or both.

An analysis of the similarity data supported the hypothesis that subjects allocated attention to the two stimulus dimensions differently in the identification and similarity judgment tasks. Specifically, in identification, they allocated approximately equal amounts of attention to the two dimensions, but when judging similarity they allocated almost all attention to the size dimension. These selective attention differences may occur because feedback is provided in identification tasks but not in similarity experiments. The feedback motivates subjects to maximize accuracy, which requires allocating equal amounts of attention to the two dimensions. When judging similarity, however, there is no objectively correct response, and so subjects may be motivated to conserve energy, especially when stimulus information is limited. One way of conserving energy is by selectively attending to the more salient stimulus dimension.

The model that best predicted the similarity judgments was a GRT model with one free attention parameter. In fact, for Subject 1, this model accounted for the data about as well as the MDS-similarity model from the SYSTAT statistical package with 14 free parameters.

The analysis of Nosofsky's (1986) categorization data yielded some support for the hypothesis that the nature of the particular categories chosen by Nosofsky caused the subjects to more strongly perceive the dimensional structure of the stimuli. It was also shown that the categorization data could be predicted successfully from the identification confusions without appealing to the notion of selective attention. A simpler model that emphasizes the different decision bounds used in identification and categorization was adequate.

These studies point to the possibly important role of selective attention in the identification–similarity relationship, and they underscore the importance of the decision bound when trying to predict categorization performance from identification confusions. They also establish GRT as a powerful alternative to the identification, categorization, and similarity models that are based on the biased-choice model and on the multidimensional scaling approach to modeling similarity.

References

Ashby, F. G. (1988). Estimating the parameters of multidimensional signal detection theory from simultaneous ratings on separate stimulus components. *Perception & Psychophysics, 44,* 195–204.

Ashby, F. G. (1989). Stochastic general recognition theory. In D. Vickers & P. L. Smith (Eds.), *Human information processing: Measures, mechanisms, and models* (pp. 435–457). New York: Elsevier Science.

Ashby, F. G. (in press). Multivariate probability distributions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition.* Hillsdale, NJ: Erlbaum.

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

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

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

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–179.

Attneave, F. (1950). Dimensions of similarity. *American Journal of Psychology, 63,* 516–556.

Burns, B., Shepp, B. E., McDonough, D., & Wiener-Ehrlich, W. K. (1978). The relation between stimulus analyzability and perceived dimensional structure. In G. H. Bower (Ed.), *The psychology of learning and motivation.* (Vol. 12, 77–115). San Diego, CA: Academic Press.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35,* 283–319.

Carroll, J. D., & Chang, J. J. (1972, March). *IDOSCAL (individual differences in orientation scaling): A generalization of INDSCAL allowing IDIOsyncratic reference systems as well as analytic approximation to INDSCAL.* Paper presented at the meeting of the Psychometric Society, Princeton, NJ.

Cliff, N. (1973). Scaling. *Annual Review of Psychology, 24,* 473–506.

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

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.

Getty, D. J., Swets, J. B., & Swets, J. A. (1980). Use of perceptual dimensions in signal identification. In R. S. Nickerson (Ed.), *Attention and performance* (Vol. 8, pp. 361–380). Hillsdale, NJ: Erlbaum.

Getty, D. J., Swets, J. A., Swets, J. B., & Green, D. M. (1979). On the prediction of confusion matrices from similarity judgments. *Perception & Psychophysics, 26,* 1–19.

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

Handel, S., & Imai, S. (1972). The free classification of analyzable and unanalyzable stimuli. *Perception & Psychophysics, 12,* 108–116.

Horan, C. B. (1969). Multidimensional scaling: Combining observations when individuals have different perceptual structures. *Psychometrika, 34,* 139–165.

Hyman, R., & Well, A. (1967). Judgments of similarity and spatial models. *Perception & Psychophysics, 2,* 233–248.

Hyman, R., & Well, A. (1968). Perceptual separability and spatial models. *Perception & Psychophysics, 3,* 161–165.

Jones, L. E., & Young, F. W. (1972). Structure of a social environment: Longitudinal individual differences scaling on an intact group. *Journal of Personality and Social Psychology, 24,* 108–121.

Kadlec, H., & Townsend, J. T. (in press). Signal detection analysis of dimensional interactions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition.* Hillsdale, NJ: Erlbaum.

Khuri, A. I., & Cornell, J. A. (1987). *Response surfaces: Designs and analyses.* New York: Marcel Dekker.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). San Diego, CA: Academic Press.

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

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review, 85,* 445–463.

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29,* 1–27.

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika, 29,* 115–129.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103–189). New York: Wiley.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

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

Nosofsky, R. M. (1985a). Luce's choice model and Thurstone's categorical judgment model compared: Kornbrot's data revisited. *Perception & Psychophysics, 37,* 89–91.

Nosofsky, R. M. (1985b). 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, and Cognition, 13,* 87–108.

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

Nygren, T. E. (1979). A theoretical framework for testing the additive difference model for dissimilarities data: Representing gambles as multidimensional stimuli. *Journal of Mathematical Psychology, 20,* 53–77.

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. (1958a). Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review, 65,* 242–256.

Shepard, R. N. (1958b). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology, 55,* 509–523.

Shepard, R. N. (1963). Analysis of proximities as a technique for the study of information processing in man. *Human Factors, 5,* 33–48.

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: Comment on Ennis (1988). *Journal of Experimental Psychology, General, 117,* 415–416.

Shepard, R. N., & Chang, J. J. (1963). Stimulus generalization in the learning of classifications. *Journal of Experimental Psychology, 65,* 94–102.

Shepard, R. N., & Chipman, S. (1970). Second-order isomorphism of internalized representations: Shapes of states. *Cognitive Psychology, 1,* 1–17.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75* (13, Whole No. 517).

Smith, J. E. K. (1980). Models of identification. In R. S. Nickerson (Ed.), *Attention and performance VIII.* Hillsdale, NJ: Erlbaum.

Smith, J. E. K. (1982). Recognition models evaluated: A commentary on Keren and Baggen. *Perception & Psychophysics, 31,* 183–189.

Takane, Y., & Shibayama, T. (1986). Comparison of models of stimulus recognition data. In J. de Leeuw, W. Heiser, J. Meulman, & F. Critchley (Eds.), *Multidimensional data analysis.* Leiden, The Netherlands: DSWO Press.

Takane, Y., & Shibayama, T. (in press). Structures in stimulus identification data. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition.* Hillsdale, NJ: Erlbaum.

Tanner, W. P. (1956). Theory of recognition. *Journal of the Acoustical Society of America, 28,* 882–888.

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: Wiley.

Townsend, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics, 9,* 40–50.

Townsend, J. T., & Ashby, F. G. (1982). Experimental test of contemporary mathematical models of visual letter recognition. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 834–864.

Townsend, J. T., Hu, G. G., & Ashby, F. G. (1981). Perceptual sampling of orthogonal straight line features. *Psychological Research, 43,* 259–275.

Tucker, L. R. (1972). Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika, 37,* 3–28.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.

Tversky, A., & Krantz, D. H. (1969). Similarity of faces: A test of interdimensional additivity. *Perception & Psychophysics, 5,* 124–128.

Wickens, T. D. (in press). Maximum-likelihood estimation of a multivariate Gaussian rating model with excluded data. *Journal of Mathematical Psychology.*

Wickens, T. D., & Olzak, L. A. (in press). Three views of association in concurrent detection ratings. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition.* Hillsdale, NJ: Erlbaum.

Wilkinson, L. (1989). *SYSTAT: The system for statistics.* Evanston, IL: SYSTAT.

(Appendix follows on next page)

Appendix

Model Fitting

GRT Models

This appendix describes the techniques used to overcome two problems in fitting the GRT models. The first problem is to compute response probabilities from the bivariate normal distribution when the correlation term (ρ) is nonzero. The second problem is to search efficiently through the parameter space in such a way that local minima are avoided.

Correlation problem. Consider a model that assumes that a certain perceptual distribution is bivariate normal with mean vector μ and covariance matrix Σ. Generating predictions from this model requires multiple integrations of the bivariate normal density function. If the correlation coefficient is zero, the multiple integral can often be easily decomposed into a product of single integrals, and in this way, it can be quickly evaluated. However, if the correlation coefficient is nonzero, the problem is more difficult.

For example, consider the following integral

$$P(X_1 \leq x_{o1}, X_2 \leq x_{o2}) = \int_{-\infty}^{x_{o2}} \int_{-\infty}^{x_{o1}} N(\mu, \Sigma) dx_1 \, dx_2 \quad \text{(A1)}$$

where $N(\mu, \Sigma)$ is the bivariate normal probability density function. To begin, we use Cholesky factorization (e.g., Ashby, in press) to find a lower triangular matrix A such that

$$\mathbf{X} = \mathrm{A}\mathbf{Z} + \mu$$

where the random vector $\mathbf{X}' = (X_1, X_2)$ has a bivariate normal distribution with mean vector μ and covariance matrix Σ, and $\mathbf{Z}' = (Z_1, Z_2)$ has a bivariate Z distribution (i.e., bivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}$).

Note that Equation A1 can be rewritten as

$$P(X_1 \leq X_{o1}, X_2 \leq X_{o2})$$

$$= P\{[1\ 0]\mathbf{X} \leq x_{o1}, [0\ 1]\mathbf{X} \leq x_{o2}\}$$

$$= P\{[1\ 0](A\mathbf{Z} + \mu) \leq x_{o1}, [0\ 1](A\mathbf{Z} + \mu) \leq x_{o2}\}$$

$$= P\{[1\ 0]A\mathbf{Z} \leq x_{o1} - [1\ 0]\mu, [0\ 1]A\mathbf{Z} \leq x_{o2} - [0\ 1]\mu\}$$

$$= P(a_{11}Z_1 \leq x_{o1} - \mu_1, a_{21}Z_1 + a_{22}Z_2 \leq x_{o2} - \mu_2). \quad \text{(A2)}$$

Probabilities such as Equation A2 can be numerically computed quickly and accurately by using a table that contains areas under the Z distribution. In our experience, extremely accurate estimates can be obtained from a list of only 25 of these values (along with linear interpolation when appropriate). Let $\Phi(z) = P(Z \leq z)$; that is, Φ is the cumulative Z distribution function. Now, for each of the 25 Z values, define

$$G(z_i) = \Phi\left(\frac{z_i + z_{i+1}}{2}\right) - \Phi\left(\frac{z_i + z_{i-1}}{2}\right).$$

The 25 $G(z_i)$ are used to approximate the Equation A2 probability in the following manner. First, step through all 25 Z_2 values for each value of Z_1 that is less than or equal to $(x_{o1} - \mu_1)/a_{11}$. For each of these (Z_1, Z_2) pairs, check whether $a_{21}Z_1 + a_{22}Z_2 \leq x_{o2} - \mu_2$. If this inequality is satisfied, compute the product $G(Z_1)G(Z_2)$. The sum of all such products approximates the Equation A2 probability and therefore also the Equation A1 integral.

Parameter estimation problem. Because many of the GRT models have a large number of parameters, the models are highly susceptible to local minima; thus, a new method was needed to overcome this problem. We used a method developed by Ennis (personal communication, February, 1990), which proceeds as follows. First, fit a model in which the only free parameters are the distribution means; the variances are all set to 1, and the correlation coefficients are all set to 0. Next, fit a model in which the means and variances are free parameters. During this step, the correlation coefficients are constrained to equal 0. In addition, the final estimates from the fit of the first model (i.e., the estimates of the means) are used as the initial estimates of the second model. Finally, during the last step all parameters are free to vary. The estimates of the means and variances obtained from the second step are used as the initial estimates during this final step. Ennis has found that this procedure reliably locates the global minimum.

Choice Models

Finding the best fits for the choice models is straightforward. For the MDS-choice models a one-step procedure that used the same minimization routine described above produced reliable parameter estimates. To fit the biased-choice model, we first used the method outlined in Smith (1982, Appendix B) to get the maximum likelihood estimates of the model parameters. These estimates were then used as the "initial guesses" for the *SSE* minimization routine.

Received June 5, 1990
Revision received November 28, 1990
Accepted December 14, 1990 ■