Journal of Experimental Psychology: General
1986, Vol. 115, No. 1, 39–57

0096-3445/86/$00.75

Attention, Similarity, and the Identification-Categorization Relationship

Robert M. Nosofsky
Indiana University

A unified quantitative approach to modeling subjects' identification and categorization of multidimensional perceptual stimuli is proposed and tested. Two subjects identified and categorized the same set of perceptually confusable stimuli varying on separable dimensions. The identification data were modeled using Shepard's (1957) multidimensional scaling-choice framework. This framework was then extended to model the subjects' categorization performance. The categorization model, which generalizes the context theory of classification developed by Medin and Schaffer (1978), assumes that subjects store category exemplars in memory. Classification decisions are based on the similarity of stimuli to the stored exemplars. It is assumed that the same multidimensional perceptual representation underlies performance in both the identification and categorization paradigms. However, because of the influence of selective attention, similarity relationships change systematically across the two paradigms. Some support was gained for the hypothesis that subjects distribute attention among component dimensions so as to optimize categorization performance. Evidence was also obtained that subjects may have augmented their category representations with inferred exemplars. Implications of the results for theories of multidimensional scaling and categorization are discussed.

In their 1956 classic, *A Study of Thinking,* Bruner, Goodnow, and Austin marveled at the capacity of people to discriminate stimuli and to *identify* them as unique items. At the same time they stressed the importance of *categorization,* the process by which discriminably different things are classified into groups and are thereby rendered equivalent. In one sense the processes of identification and categorization seem diametrically opposed, the former dealing with the particular and the latter with the general. Yet similar principles may underlie subjects' identification and categorization of multidimensional stimuli, and performance in these tasks may be highly related. Indeed, the present research renews the issue explored previously by Shepard, Hovland, and Jenkins (1961) and Shepard and Chang (1963)—namely, Do the principles of stimulus generalization underlying identification performance also underlie categorization performance? Furthermore, given knowledge of performance in an identification paradigm, can one predict performance in a categorization paradigm using the same set of stimuli?

This article is based on a PhD dissertation submitted to Harvard University.

This work was supported by Grants BNS 80-26656 from the National Science Foundation and MH 37208 from the National Institute of Mental Health to Harvard University, and by BRSG Grant S07 RR07031 from the Biomedical Research Support Grant Program at Indiana University.

I would like to thank my teachers, William Estes and R. Duncan Luce, for their guidance and encouragement during my graduate training, and for numerous criticisms, suggestions, and invaluable discussions regarding this research. My thanks also to Richard Shiffrin, Roger Shepard, and an anonymous reviewer for their criticisms of an earlier version of this article.

Correspondence concerning this article should be addressed to Robert M. Nosofsky, Department of Psychology, Indiana University, Bloomington, Indiana 47405.

A Unified Approach to Modeling Identification and Categorization

The term *identification paradigm* is used in this article to refer to a choice experiment in which there are n distinct stimuli and each stimulus is assigned a unique response. The data in an identification paradigm are summarized in an $n \times n$ confusion matrix, where cell (i, j) of the matrix gives the frequency with which Stimulus i was identified as Stimulus j. In a categorization paradigm the n stimuli are partitioned into $m < n$ groups, each group assigned a distinct response. The data in a categorization paradigm are summarized in an $n \times m$ confusion matrix, where cell (i, j) of the matrix gives the frequency with which Stimulus i was classified in Category j. The identification and categorization models studied in this article are designed to account for the data observed in these confusion matrices.

In this section a unified quantitative approach is proposed for modeling subjects' identification and categorization of multidimensional perceptual stimuli. This approach integrates well-known theories in the areas of choice and similarity so as to form a new composite model. At the heart of the approach is the assumption that subjects store individual category exemplars in memory, with classification decisions based on the similarity of stimuli to the stored exemplars (Medin & Schaffer, 1978).

Although similarity is basic for determining identification and categorization performance, it is not treated as a primitive element. The reason is that stimulus similarity is context-dependent, a point made clear by investigators such as Tversky (1977) and Tversky and Gati (1978). The key to understanding the identification-categorization relationship is to understand the manner in which similarity varies across different contexts. Although similarity is not invariant, it is presumed to change in constrained and systematic ways. Thus, a crucial move in this research will be to employ a theory of similarity in which these context-dependent changes can be modeled.

Identification Model

The present approach takes as a starting point the *similarity choice model* for stimulus identification (Luce, 1963; Shepard, 1957). Researchers have had a great deal of success in fitting the choice model to identification confusion data (e.g., Smith, 1980; Townsend, 1971; Townsend & Ashby, 1982; Townsend & Landon, 1982). According to the model, the probability that Stimulus i leads to Response j in an identification experiment, $P(R_j|S_i)$, is given by

$$P(R_j|S_i) = \frac{b_j \eta_{ij}}{\sum_{k=1}^{n} b_k \eta_{ik}}, \tag{1}$$

where $0 \leq b_j \leq 1$, $\Sigma\, b_j = 1$, $\eta_{ij} = \eta_{ji}$, and $\eta_{ii} = 1$. The b_j parameters are interpreted as response bias parameters and the η_{ij} parameters as similarity measures on the stimuli S_i and S_j.

In Shepard's (1957) original formulation of the model, the similarity parameters were given an explicit interpretation in terms of distances in a psychological space. He assumed that

$$\eta_{ij} = f(d_{ij}) \tag{2}$$

where f is some monotonically decreasing function and where the d_{ij}'s are distances that satisfy the metric axioms. To reduce the number of parameters to be estimated, Shepard suggested that the stimuli be represented as points in a low-dimensional psychological space. The d_{ij}'s could then be derived by computing the distances between the points in the space. The configuration of points that achieved the best account of the identification data would then be taken as the *multidimensional scaling* (MDS) solution for the stimulus set. I will refer to Equation 1 with the assumption that the similarity parameters are functionally related to distances in a multidimensional psychological space as the *MDS-choice* model.

The MDS-choice model provided excellent accounts of data in a series of identification learning experiments reported by Shepard (1958a). Nosofsky (1985b) found that the model provided an impressive account of a set of absolute identification data reported by Kornbrot (1978). Lockhead and his associates (Lockhead, 1970, 1972; Monahan & Lockhead, 1977) conducted numerous studies yielding results consistent with the model. In general, the pattern of identification errors observed in these studies reflected the form of the psychological space in which the stimuli were embedded. Stimuli close together in the space, and therefore similar to one another, were confused more often than stimuli far apart in the space.

To implement the MDS-choice model (Equations 1 and 2), two decisions are needed. First, what is the distance function for computing interstimulus distance relationships in the psychological space? Second, what is the function f that relates stimulus similarity to psychological distance?

In the present study the distance function was assumed to take the form of the Minkowski r-metric, in which the distance between the points x_i and x_j is given by

$$d_{ij} = [\sum_{k=1}^{N} |x_{ik} - x_{jk}|^r]^{1/r} \tag{3}$$

where $r \geq 1$, N is the number of dimensions composing the stimuli, and x_{ik} is the psychological value of Stimulus i on dimension k. Previous research suggests that the value of r that provides the best account of psychological distance relationships depends on the type of dimensions that compose the stimuli. The traditional view is that the value $r = 2$ (the Euclidean metric) is appropriate for integral-dimension stimuli and the value $r = 1$ (the city-block metric) for separable-dimension stimuli (e.g., Garner, 1974; Shepard, 1964; Torgerson, 1958). Integral dimensions are those that combine into relatively unanalyzable, integral wholes, whereas separable dimensions are highly analyzable and remain psychologically distinct when in combination. Most conclusions regarding the appropriate r-metric have been based on studies using direct judgments of similarity. In contrast, the present study tested which r-metric provides the best account of identification confusion data. The meaningfulness of defining similarity in terms of "direct" ratings or judgments will be questioned.

Two functions for relating stimulus similarity to psychological distance were considered. The first function was an exponential decay function:

$$\eta_{ij} = e^{-d_{ij}}. \tag{4a}$$

The second function was Gaussian:

$$\eta_{ij} = e^{-d_{ij}^2} \tag{4b}$$

The choice of these two functions was based on previous theoretical and empirical considerations (Nosofsky, 1985b; Shepard, 1958a, 1958b) and on empirical results observed in the present study.

Categorization Model

The categorization model proposed is a generalization of the context theory of classification developed by Medin and Schaffer (1978). The context theory has provided good accounts of data in numerous categorization experiments (Busemeyer, Dewey, & Medin, 1984; Medin, Altom, Edelson, & Freko, 1982; Medin, Altom, & Murphy, 1984; Medin, Dewey, & Murphy, 1983; Medin & Schaffer, 1978; Medin & Smith, 1981). According to the theory, the probability that Stimulus S_i is classified in Category C_J, $P(R_J|S_i)$, is given by

$$P(R_J|S_i) = \frac{b_J \sum_{j \epsilon C_J} \eta_{ij}}{\sum_{K=1}^{m} (b_K \sum_{k \epsilon C_K} \eta_{ik})}. \tag{5}$$

Uppercase letters are used here and throughout the rest of the article to index categories and categorization responses, whereas lower case letters are used to index individual stimuli and identification responses. The parameter b_J represents the bias for making category response R_J. As before, the symbol η_{ij} denotes the similarity between Stimuli S_i and S_j. The index $j \epsilon C_J$ is intended to read "all j such that S_j is a member of C_J."

As is evident, the context model response rule (Equation 5) bears a striking structural resemblance to the choice model for stimulus identification (Equation 1). Indeed, the two response rules can be linked in a simple way. The one-to-one mapping of stimuli onto responses in identification is transformed into a many-to-one mapping of stimuli onto responses in categorization. A natural starting hypothesis for a quantitative model relating the two paradigms was proposed by Shepard, Hovland, and Jenkins (1961) and Shepard and Chang (1963): To predict categorization performance from identification performance, one should simply cumulate over all stimulus-response cells in the identification matrix that would map onto a given stimulus-response cell in the categorization matrix. Stated another way, all interitem confusions in the identification paradigm that are within-class confusions would result in correct categorization responses. Only between-class confusions would result in categorization errors. I will refer to this hypothesis, illustrated schematically in Figure 1, as the *mapping hypothesis*. The mapping hypothesis formalizes the idea that the principles of stimulus generalization underlying identification will also underlie categorization. The response rule of Medin and Schaffer's context model (Equation 5) arises essentially by combining the mapping hypothesis with the assumption that the choice model accurately characterizes performance in identification paradigms (see Nosofsky, 1984a, Equation 4). The only difference is that the identification response bias parameters in Equation 1 are replaced by categorization response bias parameters in Equation 5.

Although the mapping hypothesis is structurally compelling, the present approach does not assume a *direct* mapping relation between identification and categorization performance. In particular, the η_{ij} similarity parameters in Equations 1 and 5 are not assumed to be invariant across the identification and categorization paradigms.

It is assumed that the same basic multidimensional perceptual representation underlies performance in both the identification and categorization paradigms. However, a selective attention process is assumed to operate on this perceptual representation that can lead to systematic changes in the structure of the psychological space and associated changes in interstimulus similarity relations (Shepard, 1964). Selective attention is modeled by differential weighting of the component dimensions in the psychological space, as in the INDSCAL approach to multidimensional scaling (Carroll & Chang, 1970; Carroll & Wish, 1974). In geometric terms, the weights act to stretch or shrink the psychological space along its coordinate axes. The selective attention process is formalized in the model by augmenting the Minkowski r-metric formula as follows:

$$d_{ij} = c[\sum_{k=1}^{N} w_k|x_{ik} - x_{jk}|^r]^{1/r}, \quad (6)$$

where $0 \leq c < \infty$, $0 \leq w_k \leq 1$, and $\sum w_k = 1$. The parameter c is a scale parameter reflecting overall discriminability in the psychological space. The scale parameter would be expected to increase, for example, with increases in stimulus exposure duration, or as subjects gained increased experience with the stimuli (Nosofsky, 1985a). The scale parameter is also needed to model factors associated with resource sharing among the psychological dimensions, a point to be clarified later. The w_k parameters in Equation 6 are the attention weight parameters.

As a working hypothesis, it is assumed that subjects will distribute attention among the component dimensions so as to optimize performance in a given categorization paradigm. That is, it is assumed that the w_k parameters will tend toward those values

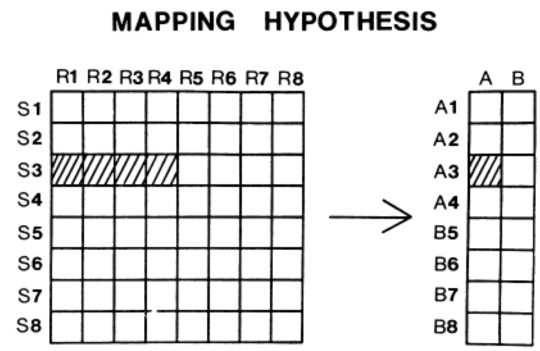

Figure 1. Left panel, An 8 × 8 stimulus-response (S-R) confusion matrix for an identification experiment; Right panel, An 8 × 2 S-R confusion matrix for a categorization experiment. (The same stimuli are used as in the identification task.) Stimuli 1–4 are assigned to Category A, and Stimuli 5–8 are assigned to Category B. According to the mapping hypothesis, one predicts the probability that Stimulus 3 is classified in Category A by summing over the probabilities that Stimulus 3 is identified as either Stimulus 1, 2, 3, or 4 in the identification task.

that maximize the average percentage of correct categorizations. The distribution of attention weights that optimizes performance will depend on the particular category structure under investigation. The notion that subjects may optimally weight component dimensions in tasks of stimulus categorization was suggested previously by Reed (1972), although an exemplar-based generalization model of the form studied here was not tested. Shepard et al. (1961, p. 42) advanced a related idea using an information-theoretic approach.

It is useful to provide an illustration of the way in which selective attention can influence stimulus similarity and the identification–categorization relationship. In Figure 2, panel A, eight stimuli are shown that vary along three binary-valued dimensions: color (black or white), shape (triangles or circles), and size (large or small). The stimuli are represented by the vertices of a cube, each face of the cube corresponding to a value along one of the dimensions. Figure 2, panel B, illustrates the situation in which subjects begin to attend selectively to the color dimension. The psychological space is stretched along the color dimension and shrunk along the size and shape dimensions. Note that by attending selectively to color, the black stimuli are rendered more similar to one another, and less similar to the white stimuli. The situation illustrated in Figure 2, panel B, would be suboptimal in an identification task because subjects would confuse stimuli of the same color with one another. Suppose, however, that subjects were required to classify the black stimuli into one category and the white stimuli into a second category. Then attending selectively to the color dimension would benefit performance, because there would be few between-class confusions and all within-class confusions result in correct categorization responses. By attending selectively to color, subjects would be maximizing within-category similarity and minimizing between-category similarity—they would be optimizing similarity relations for the given categorization problem.

The attention-optimization hypothesis has some support. Getty, Swets, Swets, and Green (1979) predicted subjects' confusion errors in an identification task from their similarity ratings of the same stimuli. First, they applied a multidimensional scaling

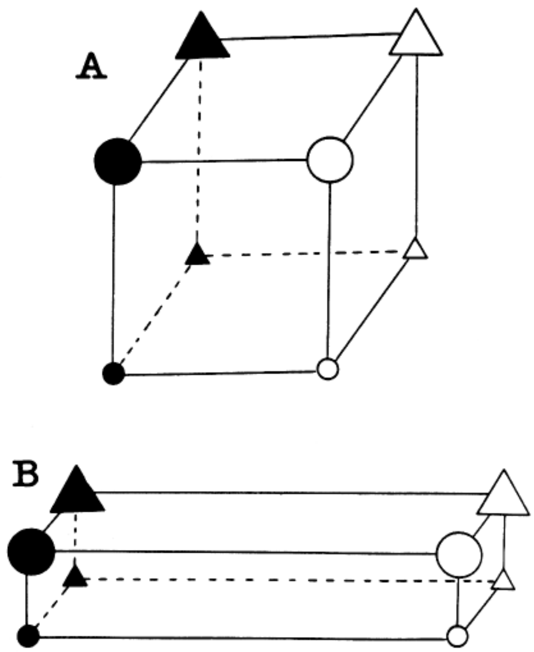

Figure 2. Schematic illustration of the attention-optimization hypothesis.

procedure to the similarity judgment data to construct a psychological space and obtain the locations of the stimuli in that space. Then, they used this scaling solution in conjunction with the MDS-choice model to predict subjects' performance in an identification paradigm. A weighted Euclidean metric was used for computing psychological distance relationships. Some support was gained for the hypothesis that subjects weighted the component dimensions so as to optimize identification performance (see also Getty, Swets, & Swets, 1980).

In a previous analysis, Nosofsky (1984a) showed that the attention-optimization hypothesis could account well for a set of categorization data reported by Shepard et al. (1961). Subjects learned to identify and categorize sets of eight stimuli that varied along three binary-valued dimensions (like those shown in Figure 2). Six different types of category structures were studied. Shepard et al. (1961) demonstrated convincingly that one could not predict subjects' categorization performance from their identification performance by directly applying the mapping hypothesis. They suggested that an additional process of selective attention intervened between the identification and categorization tasks, a process distinct from that of pure stimulus generalization. Nosofsky (1984a) formalized this idea about selective attention in terms of Equation 6 and noted that the pattern of results observed in Shepard et al.'s study provided support for an *indirect* mapping relation between identification and categorization performance, in which similarity relationships changed systematically across the two paradigms. One could account for the results within the framework of an exemplar-based generalization model by assuming that subjects distributed attention so as to optimize performance in each given categorization problem.

The data collected by Shepard et al. (1961) were obtained in a dynamic learning situation in which performance changed dramatically over the course of a session. In contrast, the present model is a static one, intended to account for categorization performance at a given stage of learning or under experimental conditions in which performance is stable. Thus, an important goal in the present research is to study the identification–categorization relationship under fairly steady-state performance conditions. This will allow for a more appropriate test of the categorization model and the attention-optimization hypothesis. A second goal is to study categorization performance in some depth at the level of individual subjects. Some researchers have suggested that the success of the exemplar-based context model may be an artifact of averaging over different subjects' responses (Martin & Caramazza, 1980). A successful application of the model to individual subject data would lead one to question this interpretation.

Before turning to the empirical portion of this article, the relationship between the present model and Medin and Schaffer's (1978) context theory should be clarified. In addition to suggesting Equation 5 as a categorization decision rule, Medin and Schaffer proposed a rule for computing interstimulus similarity. The stimuli used in their experiments varied along binary-valued separable dimensions. The similarity between Stimuli S_i and S_j was given by the following multiplicative rule:

$$\eta_{ij} = \prod_{k=1}^{N} s_k, \tag{7}$$

where $s_k = p_k$ $(0 \leq p_k \leq 1)$ if $x_{ik} \neq x_{jk}$; and $s_k = 1$ if $x_{ik} = x_{jk}$. That is, if Stimuli S_i and S_j mismatched on dimension k, then s_k was set equal to some parameter p_k; and if S_i and S_j matched on dimension k, then s_k was set equal to 1. The multiplicative rule is a crucial feature of the context theory differentiating it from some alternative categorization theories (Medin & Smith, 1981; Smith & Medin, 1981). A virtue of the multiplicative rule is that it is sensitive to correlational structure (see Medin, 1983).

As noted previously by Nosofsky (1984a), the multiplicative rule is a special case of the multidimensional scaling approach to modeling stimulus similarity. An interdimensional multiplicative similarity rule arises if

$$d_{ij} = c\Big[\sum_{k=1}^{N} w_k|x_{ik} - x_{jk}|^r\Big]^{1/r} \tag{8a}$$

and

$$\eta_{ij} = e^{-d_{ij}^r}, \tag{8b}$$

because

$$\eta_{ij} = e^{-(c[\sum_{k=1}^{N} w_k|x_{ik}-x_{jk}|^r]^{1/r})^r} = e^{-c^r \sum_{k=1}^{N} w_k|x_{ik}-x_{jk}|^r} = \prod_{k=1}^{N} e^{-c^r w_k|x_{ik}-x_{jk}|^r} = \prod_{k=1}^{N} s_k, \tag{9}$$

where $s_k = \exp(-c^r w_k|x_{ik} - x_{jk}|^r)$. So, for example, an interdimensional multiplicative similarity rule would arise if the city-block metric described psychological distance relationships and an exponential decay function related stimulus similarity to psychological distance. For binary-valued stimulus dimensions, Equation 9 reduces to the multiplicative rule proposed by Medin and Schaffer.

For obvious reasons then, the categorization model proposed in this article will be referred to as the *generalized context model* (GCM). The multidimensional scaling approach adopted by the GCM offers several advantages for purposes of studying subjects' categorization performance. First, whereas Medin and his associates have in effect limited their tests of the context theory to stimuli varying along binary-valued dimensions, the present interpretation allows for a straightforward extension of the model to stimuli varying along multivalued continuous dimensions. Once a multidimensional scaling solution for the stimulus set is derived, the similarity between any two stimuli will be a function of their distance in the psychological space. Second, the present approach has the advantage that the model-fitting process is less post hoc. In previous applications of the context model, researchers had to estimate best-fitting similarity parameters for each of the dimensions along which the stimuli are presumed to vary (see Equation 7). In contrast, once a multidimensional scaling solution is derived, the possible similarity relationships are more highly constrained, thereby yielding a more rigorous test of the theory under consideration. Another advantage of this approach is that it removes some of the arbitrariness from the theoretical analysis. In a great deal of research on categorization, the experimenters specify a set of physical dimensions that define the stimuli and then assume that the psychological dimensions match this physical specification. As noted by investigators such as Lockhead and King (1977) and Cheng and Pachella (1984), discrepancies between the physically specified dimensions and the underlying psychological ones can lead to erroneous conclusions and interpretations. The advantage of a multidimensional scaling approach is that the psychological dimensions are revealed to the experimenter rather than assumed a priori.

Overview of Theoretical Goals

In summary, the central goal in this research is to account quantitatively for subjects' identification and categorization of multidimensional perceptual stimuli, and to characterize performance relationships between these two paradigms. In the experiments to be reported, subjects are required to identify and categorize the same set of separable-dimension stimuli. To obtain fairly stable performance data, perceptually confusable stimuli and short exposure durations are used. The identification data are analyzed using Shepard's (1957) MDS-choice model. The multidimensional scaling solution that provides a best account of the identification data within this modeling approach is taken as the underlying perceptual representation for the stimulus set. This perceptual representation is then used in conjunction with the GCM (Equations 5, 4, and 6) to predict subjects' performance in various categorization paradigms. The hypothesis is then tested that subjects distribute attention among component dimensions so as to optimize categorization performance.

Method

Subjects

Two subjects, one male and one female, served as paid observers. Subject 1, the male, was highly experienced in auditory psychophysical experiments.

Apparatus

A Tektronix 604 monitor, interfaced with a PDP 11/10 computer, was used to present the stimuli.

Stimuli were semicircles that varied in size (four levels) and angle of orientation of a radial line drawn from the center of the semicircle to the rim (four levels). The four sizes (length of radius) were .478, .500, .522, and .544 cm; the four angles were 50°, 53°, 56°, and 59°. The dimension values were combined orthogonally to yield a 16-member stimulus set. Previous research indicates that stimuli like these are composed of separable dimensions (Garner & Felfoldy, 1970; Shepard, 1964).

The stimuli were constructed by illuminating points on the Tektronix screen. The center of each semicircle was located at the center of the screen. All stimuli were of equal luminance. A poststimulus pattern mask was used that consisted of a grid of points centered at the center of the screen. Details of the procedure for generating the stimuli are provided in the work of Nosofsky (1984b).

Although the stimuli were constructed from discrete collections of points, the grain of the screen was fine enough that they appeared as continuous images (except for the pattern mask). The stimuli appeared green on a black background. Subjects sat approximately 2 feet from the screen in a dimly lit room.

Procedure

Identification conditions. On any given trial in the identification conditions a fixation dot appeared on the center of the screen for 500 ms. A randomly selected stimulus was then presented immediately for 150 ms and was followed by the pattern mask. Subjects were required to enter their response within 10 s of stimulus offset. Immediately following the response, the correct answer was presented on the screen for 1 s. There was a 500-ms intertrial interval.

In Condition AS the subjects identified both the size and angle of the stimulus. The size and angle dimension values were each given the labels 1 (smallest size, lowest angle) through 4 (largest size, highest angle). Subjects entered their responses by pressing one of 16 buttons arranged in a 4 × 4 grid. To enter size i and angle j, a subject pressed the button in row i and column j. Feedback on each trial consisted of a pair of numbers presented on the center of the screen, the number on the left corresponding to size and the number on the right to angle.

In Condition A subjects identified only the angle of the radial line. The response was entered by pressing one of the four buttons in row 1 of the grid. In Condition S subjects identified only the size of the semicircle. The response was entered by pressing one of the four buttons in column 1 of the grid.

An experimental session was organized into 12 blocks of 100 trials each. In each session, Condition AS was tested on Blocks 1–4 and Blocks 7–10. Conditions A and S were tested on Blocks 5–6 and 11–12, in alternating order each day. Each subject completed eight identification sessions, plus some additional sessions to be discussed shortly. The identification condition was preceded by approximately 2,650 trials of practice for each subject, using a slightly longer exposure duration (250 ms).

Following each block subjects were presented with a summary of their performance. The summary included two 4 × 4 confusion matrixes, one for size and one for angle. For example, row i and column j of the size matrix gave the frequency with which size i was presented and the subject responded with size j. Subjects were also presented with a summary of their overall percentage correct scores on each level of size and each level of angle.

The subjects were tested individually in 2-hour sessions. They were encouraged to respond accurately and to take rest breaks during the testing. In Condition AS the subjects were instructed to attend equally to both dimensions and not to favor one dimension over the other. In Conditions A and S the subjects were instructed to attend to only the relevant dimension.

Categorization conditions. Following the identification condition, each subject participated in four categorization conditions. Each condition consisted of a learning phase followed by a transfer phase. In the learning phase, four stimuli were assigned as exemplars of Category 1, and four other stimuli were assigned as exemplars of Category 2. The remaining eight stimuli were not used. The only other procedural difference between the categorization condition learning phase and the identification condition involved the stimulus-response mapping. In the categorization learning phase the subjects classified the stimuli in either Category 1 or 2. The subjects entered their responses by pressing one of two buttons on the response grid. Feedback at the end of each trial indicated whether the stimulus was assigned to Category 1 or Category 2. Summary feedback at the end of each block consisted of a 2×2 confusion matrix. The entry in row i and column j gave the frequency with which a stimulus from Category i was presented and the subject responded Category j. Percentage correct scores for each category were also presented.

Subjects completed approximately 1,200 trials per session in each of the learning conditions. The learning phase continued until a subject scored above chance on each category exemplar for the final 600 trials of a session.

In the transfer phase all 16 stimulus set members were presented. Subjects continued to receive trial-by-trial feedback for the stimuli that served as category exemplars during the learning phase. No feedback was given on those trials in which unassigned transfer stimuli were presented. The same end-of-block feedback was presented in the transfer phase as in the learning phase. (Transfer stimulus presentations were not included in the end-of-block feedback.)

The four categorization conditions differed only in the structure of the categories that were used. The category structures are shown schematically in Figure 3. In these grids the rows correspond to levels of size (top = largest, bottom = smallest), and the columns correspond to levels of angle (leftmost = lowest, rightmost = highest). Cells in the grid that are marked with a 1 represent stimuli assigned to Category 1, whereas cells with a 2 represent stimuli assigned to Category 2. Cells that have no number were unassigned transfer stimuli.

Each of the category structures can be described by a fairly simple rule. In the "dimensional" categorization, small stimuli are assigned to Category 1, and large stimuli are assigned to Category 2. The "criss cross" categorization can be described by a biconditional rule: Small stimuli with low angles and large stimuli with high angles are assigned to Category 2, and the reverse for Category 1. In the "interior–exterior" categorization, stimuli that have an extreme value (either 1 or 4) on either dimension are assigned to Category 2, whereas stimuli with intermediate values on both dimensions are assigned to Category 1. The diagonal categorization can be described as a rule-plus-exception structure: Stimuli with low angles are assigned to Category 1, and stimuli with high angles to Category 2, with one exception in each category.

Approximately 3,500 trials were conducted for each subject in each of the transfer conditions. The order of administration of the conditions was criss-cross, dimensional, interior–exterior, and diagonal for Subject 1; and dimensional, criss-cross, interior–exterior, and diagonal for Subject 2. Following each condition, subjects were tested in an additional session of identification Condition AS. This was done to assess any changes in sensitivity that might have accompanied subjects' increased experience with the stimuli. Approximately 1,000 trials were conducted for each subject in each of the additional identification sessions.

Results

Identification Condition Theoretical Analysis

The first step in the analysis was to fit the MDS-choice model (Equations 1, 3, and 4) to the Condition AS identification data. Because the analysis was lengthy, and the general approach is a classic one, the major results are simply summarized. A fuller treatment of the identification data can be found in separate reports (Nosofsky, 1984b, in press).

1. The data obtained for both subjects in Identification Condition AS are presented in Table 1.

2. The MDS-choice model yielded its best fits to the identification data by assuming a Gaussian function for relating similarity to psychological distance and a Euclidean metric for describing psychological distance relationships. These same functions will be assumed to operate in the categorization conditions. It is interesting to note that the combination of a Gaussian similarity function and a Euclidean distance metric yields an interdimensional multiplicative similarity rule (see Equation 9).

The support for the Euclidean metric contrasts with the widely held view that for separable-dimension stimuli, values of r less than or equal to 1 in the Minkowski r-metric formula provide the best account of psychological distance relationships. In the present analyses these values did dramatically worse than $r = 2$. Possible reasons for the discrepancies between the present results and earlier conclusions are considered in the General Discussion section.

3. The Gaussian-Euclidean MDS-choice model provided excellent fits to the identification data, accounting for 99% of the response variance in the data of both subjects. Scatterplots of the observed confusion frequencies against the predicted confusion frequencies are shown in Figure 4. The excellent fits that were obtained provide support for the choice model and for the multidimensional scaling approach to modeling similarity.

4. By fitting the MDS-choice model to the identification data, a two-dimensional scaling solution was derived for the stimulus set. The maximum-likelihood coordinate parameters are reported in Table 2 and are shown graphically in Figure 5. The gridlike regularity evident in these plots reflects the physical structure of the stimulus set. The psychological dimensions are interpreted as corresponding to the physical dimensions of size and angle. Now that a multidimensional scaling solution has been derived, the GCM can be applied to account for performance in each of the categorization conditions.

Categorization Conditions Theoretical Analysis

The generalized context model. Since there are two categories and the stimuli are two-dimensional, the GCM can be summarized as follows. The probability that a subject classifies Stimulus S_i into Category C_1, $P(R_1|S_i)$, is given by

$$P(R_1|S_i) = \frac{b_1 \sum_{j \epsilon C_1} \eta_{ij}}{b_1 \sum_{j \epsilon C_1} \eta_{ij} + (1 - b_1) \sum_{k \epsilon C_2} \eta_{ik}}, \tag{10}$$

where $0 \leq b_1 \leq 1$. The similarity between stimuli S_i and S_j is given by

$$\begin{aligned} \eta_{ij} &= e^{-[c\sqrt{w_1(x_{i1} - x_{j1})^2 + (1 - w_1)(x_{i2} - x_{j2})^2}]^2} \\ &= e^{-c^2[w_1(x_{i1} - x_{j1})^2 + (1 - w_1)(x_{i2} - x_{j2})^2]}, \end{aligned} \tag{11}$$

where $0 \leq c < \infty$ and $0 \leq w_1 \leq 1$. The x_{ik} coordinate values are given by the multidimensional scaling solution for the stimulus set. The parameters in the model are the bias parameter b_1, the scale parameter c, and the attention weight parameter w_1.

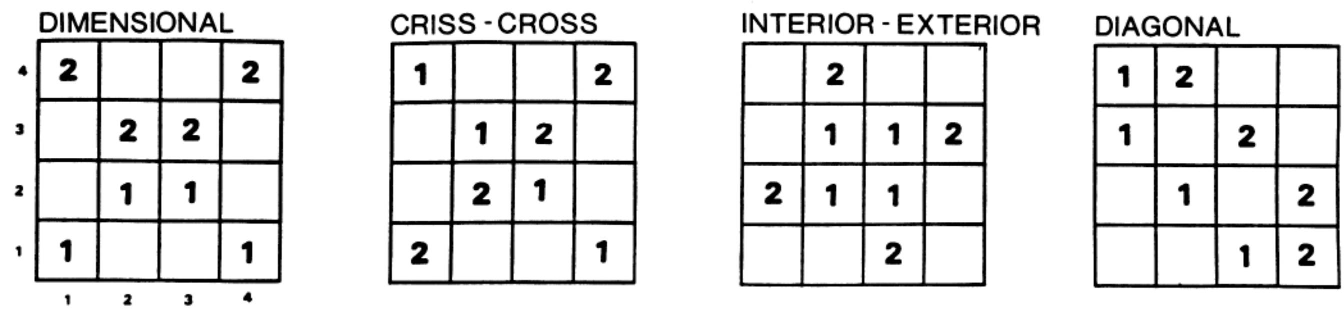

Figure 3. Schematic illustration of the four category structures. (Rows = size; columns = angle.)

Fits of the GCM to the categorization data. The data obtained in the four categorization transfer conditions are presented in Table 3. The table shows the frequency with which each stimulus was classified in Category 1 or Category 2. To aid in the interpretation of the transfer data, the response proportions are also summarized in the spatial layouts in Figure 6. The value on the top right-hand side of each cell is the observed proportion of times that the stimulus was classified in Category 1.

The GCM was fitted to the categorization data using a maximum-likelihood criterion. Note that for each categorization fit, 3 parameters were estimated to account for 16 data values that were free to vary. The theoretical response proportions for each

Table 1
Condition AS Confusion Data

Si	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
							Subject 1									
1 (1, 1)[a]	250	94	0	0	192	42	0	0	7	5	0	0	0	0	0	0
2 (2, 1)	50	239	90	3	50	102	49	4	3	4	0	0	0	0	0	0
3 (3, 1)	0	59	258	93	3	20	113	37	0	3	4	2	0	0	0	0
4 (4, 1)	0	11	137	288	0	3	60	79	0	1	6	5	0	0	0	0
5 (1, 2)	29	16	0	0	341	75	7	0	122	27	2	0	6	1	0	0
6 (2, 2)	2	41	36	1	70	193	97	5	30	67	35	2	0	0	1	1
7 (3, 2)	0	7	48	22	5	40	226	104	0	17	61	20	0	0	0	0
8 (4, 2)	0	1	26	86	0	10	110	271	0	2	39	55	0	0	2	0
9 (1, 3)	0	0	0	0	58	18	0	0	351	80	4	0	82	20	1	0
10 (2, 3)	1	0	1	0	13	55	40	6	77	255	95	9	16	45	19	0
11 (3, 3)	0	0	2	1	1	11	73	64	1	58	213	113	0	10	52	8
12 (4, 3)	0	0	0	1	0	1	32	137	0	8	120	255	0	2	19	32
13 (1, 4)	0	0	0	0	1	2	1	0	111	23	1	0	309	96	6	1
14 (2, 4)	0	0	0	0	1	1	4	0	31	95	51	1	49	246	96	3
15 (3, 4)	0	0	0	0	0	0	8	5	1	18	105	78	2	50	234	84
16 (4, 4)	0	0	0	0	0	0	1	2	0	2	56	186	0	2	101	249
							Subject 2									
1 (1, 1)	328	39	4	0	130	30	4	0	17	6	3	0	1	0	1	0
2 (2, 1)	149	145	42	5	92	83	22	3	8	30	6	1	1	2	1	0
3 (3, 1)	24	110	167	64	8	70	87	29	1	16	24	5	0	1	0	0
4 (4, 1)	1	19	116	185	0	21	89	85	0	11	26	19	0	1	5	0
5 (1, 2)	92	22	3	1	215	42	7	1	124	23	0	1	7	3	0	0
6 (2, 2)	35	56	26	4	75	153	45	9	39	99	25	0	3	13	3	0
7 (3, 2)	3	16	40	36	6	61	144	87	3	54	71	37	0	5	8	3
8 (4, 2)	0	1	28	54	2	9	75	198	1	10	74	91	0	1	12	5
9 (1, 3)	9	5	0	0	109	27	2	1	244	45	2	1	91	29	2	0
10 (2, 3)	5	5	3	1	37	92	24	6	72	181	37	7	23	71	18	0
11 (3, 3)	1	2	8	4	1	19	71	67	5	90	144	76	1	25	69	14
12 (4, 3)	0	0	3	9	0	1	24	95	0	7	106	170	0	6	64	50
13 (1, 4)	0	0	0	0	21	6	0	0	142	57	3	0	239	99	5	0
14 (2, 4)	0	0	1	0	5	16	9	3	31	118	48	11	34	205	103	20
15 (3, 4)	0	1	0	1	0	5	7	15	0	25	79	95	2	36	162	119
16 (4, 4)	0	0	0	2	0	1	6	16	1	3	54	106	0	3	93	308

Note. Rows correspond to stimuli and columns correspond to responses.
[a] Angle, Size

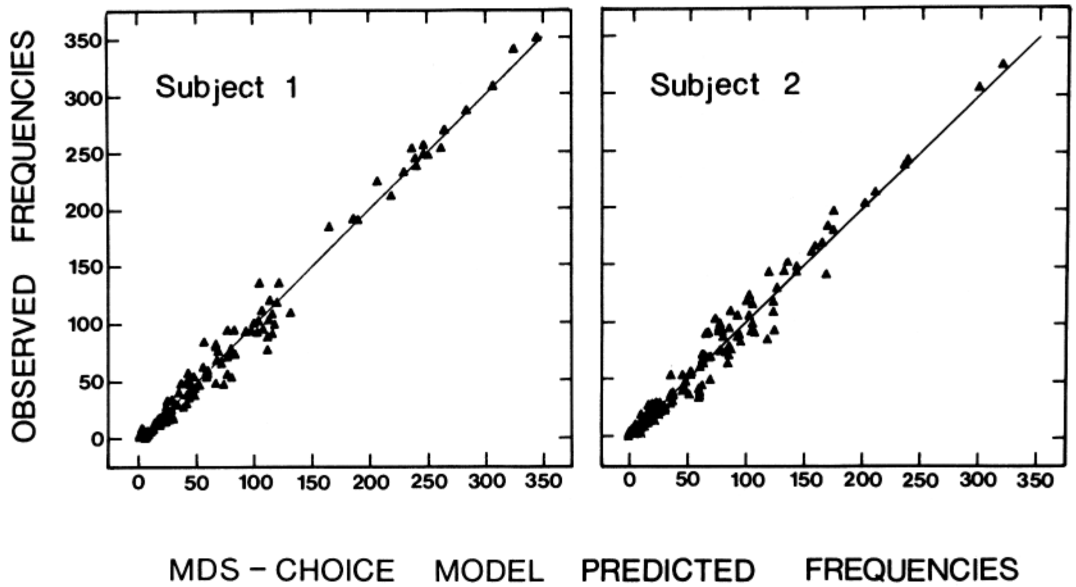

Figure 4. Observed confusion frequencies for condition AS plotted against the predicted confusion frequencies (MDS-choice model).

categorization condition are compared to the observed proportions in the spatial layouts in Figure 6, with the summary fits reported in Table 4. Figure 7 plots the predicted and observed response proportions for all the categorization conditions taken together. The model accounts for 96.6% of the variance in Subject 1's collapsed data and for 93.7% of the variance in Subject 2's collapsed data.

Although the GCM accounts fairly well for the collapsed data, there are some discrepancies in some of the individual conditions that make the model seem suspect. It is particularly the results for the interior-exterior categorization that lead one to question the simple GCM. Because of these discrepancies, I was led to consider an augmented version of the original model. As will be seen, this augmented model does considerably better than the original one for some of the conditions. Since the simple GCM may be lacking in some important respects, I postpone examination of the best-fitting parameters, and, instead, turn directly to a presentation of the augmented model.

Table 2
Maximum-Likelihood Coordinate Parameters for Gaussian-Euclidean MDS-Choice Model

	Subject 1		Subject 2	
S_i	Dimension 1	Dimension 2	Dimension 1	Dimension 2
1	−1.855	−1.532	−1.356	−1.430
2	−0.687	−1.617	−0.492	−1.211
3	0.436	−1.633	0.349	−1.235
4	1.331	−1.647	1.088	−1.121
5	−1.615	−0.469	−1.413	−0.528
6	−0.531	−0.558	−0.412	−0.425
7	0.500	−0.590	0.518	−0.382
8	1.373	−0.535	1.275	−0.215
9	−1.522	0.657	−1.477	0.302
10	−0.395	0.518	−0.388	0.294
11	0.648	0.469	0.607	0.368
12	1.513	0.481	1.317	0.539
13	−1.427	1.770	−1.389	1.189
14	−0.301	1.639	−0.313	1.153
15	0.767	1.541	0.673	1.187
16	1.764	1.512	1.414	1.518

Note. MDS = multidimensional scaling; S_i = Stimulus i.

The Augmented GCM

In the GCM it is assumed that subjects' categorization of a given stimulus is determined by its similarity to the stored category exemplars. The exemplars that are stored in memory are assumed to be precisely those stimuli that were assigned by the experimenter to one or the other category. Consider the following

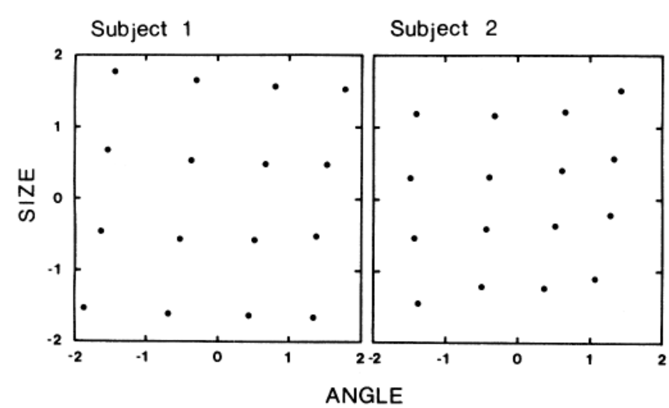

Figure 5. Multidimensional scaling (MDS) solution derived by fitting the Gaussian-Euclidean MDS-choice model to the subjects' condition AS identification data. (Note that for an unweighted Euclidean metric the orientation of the solution is arbitrary. The orientations shown here are those that provided the best overall account of the categorization data and the most easily interpretable set of GCM parameter estimates.)

hypothesis. Suppose that instead of storing only assigned stimuli in memory, subjects augment their memory representations with unassigned stimuli. In particular, the category representation consists not only of experimenter-assigned exemplars but also of sets of stimuli that are *inferred* to be members of the category. The basic spirit of the GCM is preserved in the sense that categorization is still determined by similarity of stimuli to stored category exemplars. The sets of stored exemplars, however, are now assumed to include inferred exemplars, rather than simply experimenter-assigned exemplars. The process by which such memory augmentation may take place is discussed shortly. This notion of inferred sets of exemplars is similar in certain respects to previous ideas advanced by Garner (1974, Chapter 1).

Formally, in the augmented GCM, the probability that a subject classifies Stimulus S_i into Category C_J, $P(R_J|S_i)$, is given by

$$P(R_J|S_i) = \frac{b_J \sum_{j \in IC_J} \eta_{ij}}{\sum_{K=1}^{m} (b_K \sum_{k \in IC_K} \eta_{ik})} \qquad (12)$$

where IC_J is the set of all stimuli that are inferred to belong to Category J. The η_{ij} values are computed as before. For starting purposes, I assume that each of the stimuli in the 16-member set is inferred to belong in either Category 1 or Category 2. That is, I assume that the subject partitions the stimulus set into two mutually exclusive and jointly exhaustive subsets. Now to fit the augmented GCM to the categorization data, one needs to know which particular partition the subject adopted. Unfortunately, this adds a major unknown to the modeling enterprise, meaning that the power of the basic theory is reduced. For present purposes, the adopted partition is conceptualized as being an additional free "parameter" that needs to be estimated. Later, I consider possible determinants of this parameter.

To simplify the analysis, I assume that the subject partitions all experimenter-assigned exemplars into their correct category. Since there are eight remaining exemplars, there are $2^8 = 256$ possible partitions for each categorization condition. The partition that the subject adopts is not a free parameter in the usual sense. It is perhaps more appropriately described as a *qualitative* parameter. The partition parameter does not have the properties

Table 3
Categorization Transfer Data Summaries

	Dimensional			Criss-cross			Interior–exterior			Diagonal		
S_i	C1	C2	P (C1|S_i)	C1	C2	P (C1|S_i)	C1	C2	P (C1|S_i)	C1	C2	P (C1|S_i)
						Subject 1						
1	213	4	.98	7	203	.03	19	238	.07	226	0	1.00
2	253	1	1.00	27	187	.13	78	162	.32	231	20	.92
3	192	2	.99	183	61	.75	76	181	.30	165	69	.71
4	218	1	1.00	206	9	.96	65	219	.23	92	168	.35
5	185	57	.76	58	162	.26	36	216	.14	214	6	.97
6	193	47	.80	73	151	.33	179	72	.71	206	67	.75
7	187	40	.82	152	54	.74	161	65	.71	109	151	.42
8	162	36	.82	187	47	.80	99	159	.38	44	212	.17
9	24	194	.11	193	21	.90	60	189	.24	208	20	.91
10	33	198	.14	147	64	.70	206	62	.77	108	135	.44
11	31	190	.14	46	155	.23	171	75	.70	31	264	.11
12	40	181	.18	66	154	.30	101	150	.40	12	245	.05
13	0	204	.00	212	4	.98	32	238	.12	209	41	.84
14	0	235	.00	149	44	.77	128	126	.50	71	191	.27
15	0	220	.00	35	214	.14	116	157	.42	13	211	.06
16	0	258	.00	13	216	.06	39	223	.15	3	258	.01
						Subject 2						
1	196	7	.97	30	190	.14	14	132	.10	216	3	.99
2	185	4	.98	95	161	.37	48	81	.37	199	15	.93
3	214	9	.96	164	26	.86	54	116	.32	139	58	.71
4	197	8	.96	192	14	.93	24	118	.17	41	195	.17
5	155	27	.85	101	139	.42	38	106	.26	203	5	.98
6	150	35	.81	88	128	.41	95	40	.70	193	36	.84
7	152	55	.73	155	70	.69	89	55	.62	53	178	.23
8	165	51	.76	169	62	.73	41	99	.29	14	215	.06
9	59	120	.33	176	51	.78	61	83	.42	219	12	.95
10	86	116	.43	131	118	.53	131	16	.89	151	67	.69
11	57	135	.30	62	152	.29	122	45	.73	36	194	.16
12	58	170	.25	75	147	.34	33	98	.25	6	237	.02
13	9	193	.04	199	23	.90	40	106	.27	189	24	.89
14	11	171	.06	122	101	.55	70	71	.50	75	143	.34
15	7	195	.03	22	199	.10	34	101	.25	6	233	.03
16	4	199	.02	18	219	.08	15	124	.11	3	241	.01

Note. C1 = Category 1; C2 = Category 2; P (C1|S_i) = Probability of a Category 1 response given presentation of Stimulus i.

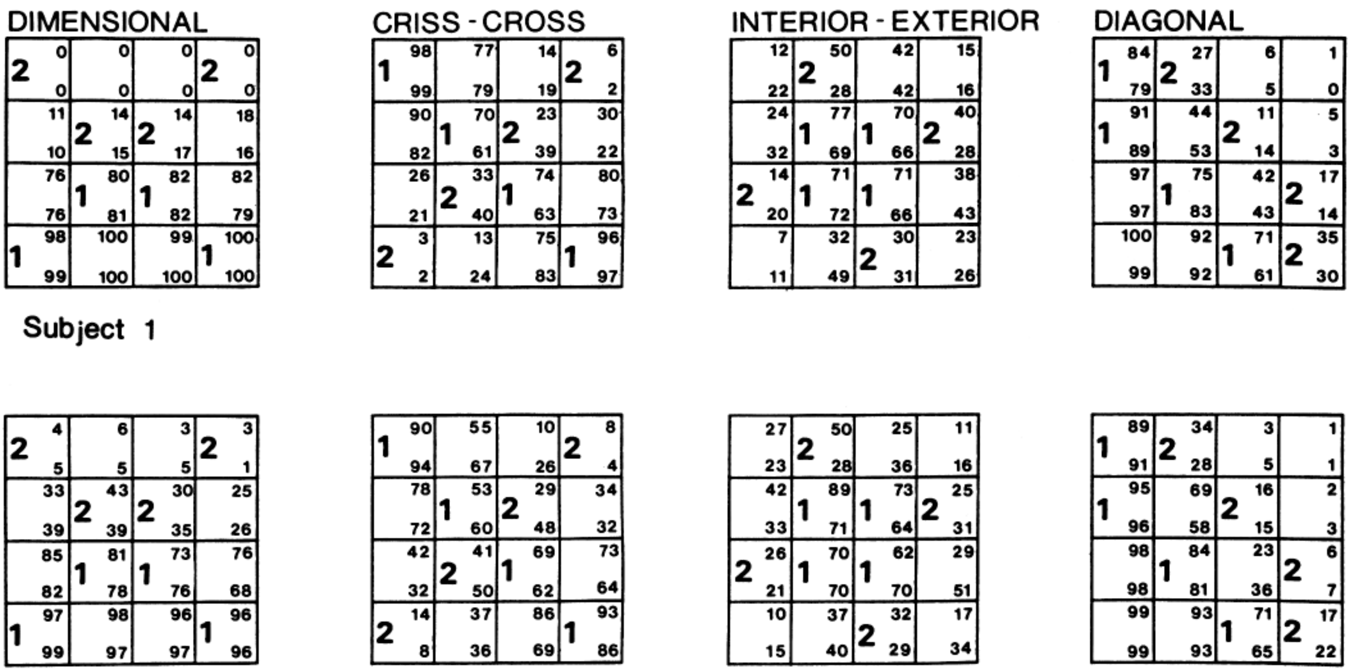

Figure 6. Observed (top right) and predicted (bottom right) classification proportions for each categorization condition. (Values are the proportions of times that the stimuli were classified in Category 1.)

of a continuously varying quantitative parameter for which confidence intervals can be derived. From a formal point of view, each partition is in reality a new model, although the "model space" is a highly constrained one.

Fits of the Augmented GCM to the Categorization Data

The augmented GCM was fitted to the categorization data in systematic fashion. For each of the 256 partitions, a search was carried out to find the GCM parameters (c, w_1, and b_1) that provided a best account of the data. The best-fitting partitions and a comparison of the theoretical and observed proportions for each condition are summarized in the spatial layouts in Figure 8. Cells with large boldface numbers represent experimenter-assigned exemplars and cells with small boldface numbers represent inferred exemplars. In some cases more than one partition provided a good account of the same categorization condition data. The summary fits and best-fitting parameters for all partitions that were competitive are reported in Table 5.

Comparing the summary fits in Table 5 with those in Table 4, it is clear that the augmented model provides a better overall account of the categorization data than the simple GCM. The improvement is substantial for the criss-cross and interior–exterior categorization conditions. Because of the qualitative nature of the partition parameter, standard statistical tests are not appropriate. However, computer simulation suggested that the im-

Table 4
GCM Summary Fits

	Subject 1			Subject 2		
Condition	−ln L	SSE	% Var	−ln L	SSE	% Var
Dimensional	6.179	.002	99.93	11.427	.017	99.21
Criss-cross	60.896	.095	94.73	91.485	.159	86.97
Interior–exterior	75.644	.126	84.52	72.022	.208	74.84
Diagonal	24.046	.034	98.31	25.576	.043	98.19

Note. GCM = generalized context model; −ln L = − log likelihood; SSE = sum of squared deviations between observed and predicted categorization probabilities; % Var = percent variance accounted for.

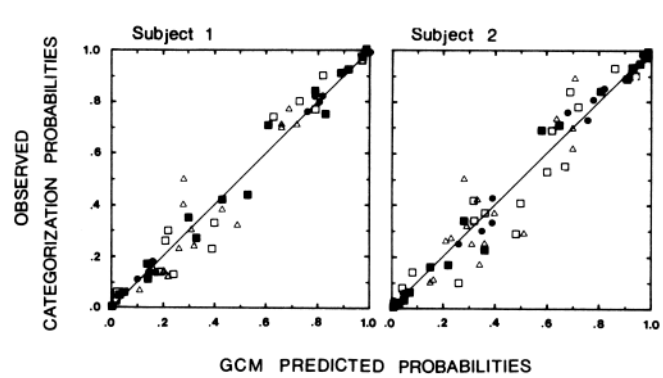

Figure 7. Observed classification proportions plotted against predicted generalized context model (GCM) classification proportions for all the categorization conditions taken together. (Dot = dimensional; hollow square = criss-cross; triangle = interior–exterior; solid square = diagonal.)

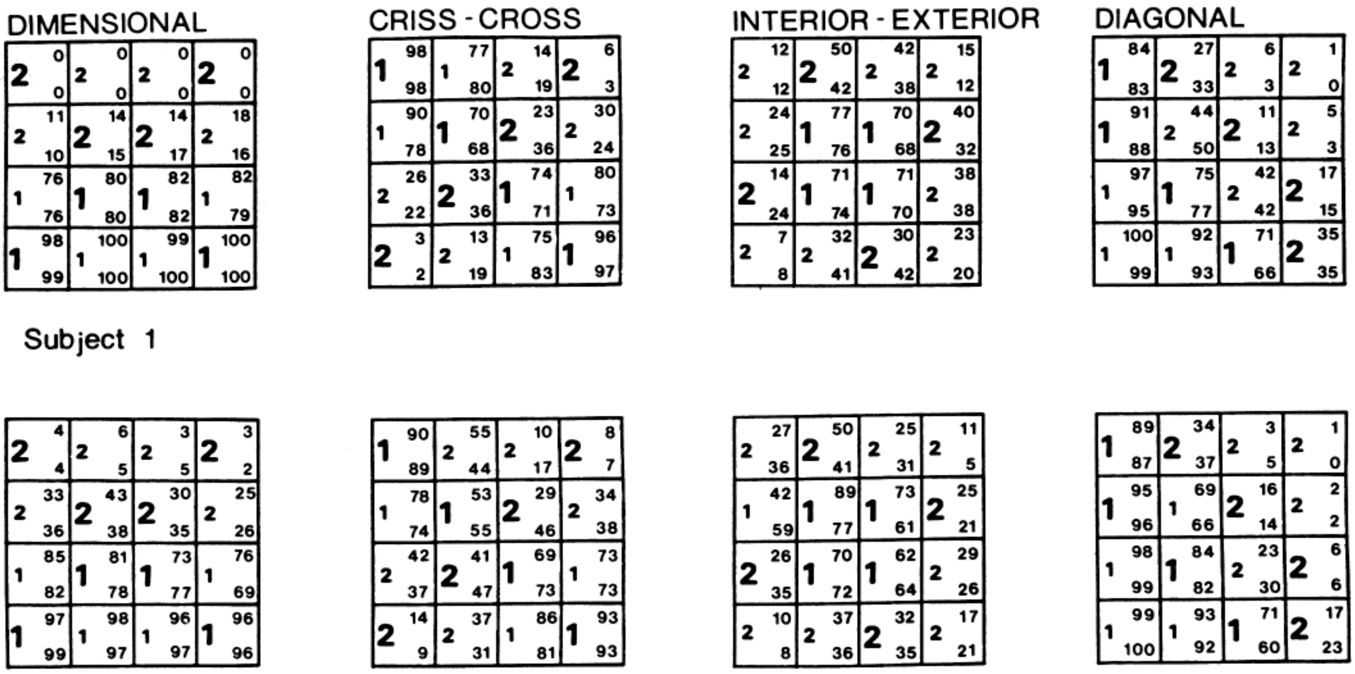

Figure 8. Augmented generalized context model fitted to the categorization data. (Large boldface numbers on the left side of each cell represent experimenter-assigned exemplars and small boldface numbers on the left represent inferred category exemplars. Observed [top right] and predicted [bottom right] values are the proportions of times the stimuli were classified in Category 1.)

provements in fit yielded by the partition parameter for the interior–exterior and criss-cross category structures were statistically significant. The augmented GCM theoretical proportions are plotted against the observed proportions for all the categorization conditions taken together in Figure 9. The model accounts for 98.4% of the variance in Subject 1's collapsed data, and for 97.1% of the variance in Subject 2's collapsed data.

Examination of the Augmented GCM Parameters

Selective attention and categorization performance. The parameter of greatest interest in the present investigation is w_1. According to the theory, the relationship between identification and categorization performance may be understood in terms of selective attention to the component dimensions that compose the stimuli. Deviations of the attention weight parameter away from $w_1 = 1/2$ provide evidence of differential selective attention, and, therefore, of changes in interstimulus similarity relations across the identification and categorization paradigms. The results of likelihood ratio tests of the null hypothesis $w_1 = 1/2$ are summarized in Table 5 (asterisks). There are numerous cases in which the attention weight parameter deviates significantly from 1/2.

A more interesting question concerns the manner in which the attention weight parameter varies. I suggested earlier that subjects may weight component dimensions so as to optimize performance in a given categorization paradigm. In the present context, it is natural to assume that subjects try to maximize average percentage correct. In Figure 10, theoretical performance gradients are plotted for the best-fitting partitions in each of the categorization conditions. These gradients were generated by holding fixed the best-fitting values of c and b_1, and then varying w_1 from 0 to 1.0 in increments of .05. For each value of w_1, the percentage of correct categorizations predicted by the augmented GCM was computed. These percentage correct computations were carried out only over the original training exemplars because there were no experimentally defined correct answers for the unassigned stimuli. The locus on each gradient that is marked with a solid circle is the observed value of w_1. That is, this is the value of w_1 that, in conjunction with the other augmented GCM parameters, provided a maximum-likelihood fit to the categorization data.

The results for Subject 1 support the attention-optimization hypothesis. Full attention is given to the size dimension in the dimensional categorization, approximately equal attention is given to size and angle in the criss-cross categorization, and there is a tendency to weight angle more than size in the diagonal categorization. Each of these results is in accord with the optimization prediction. The subject appears to be operating in suboptimal fashion only in the interior–exterior categorization, although the performance gradient here is extremely flat.

The results for Subject 2 are more equivocal. Support for the optimization prediction comes from the dimensional categorization, in which the subject has focused almost all attention on the size dimension. The subject is also virtually at optimum in the diagonal categorization. For the criss-cross and interior–exterior categorizations, however, the subject does not appear to be optimizing.

The memory-augmentation process and the partition parameter. A major question raised by the present research regards the generality and underlying basis of the posited memory-augmentation process by which subjects added inferred exemplars to their category representations. Several aspects of the present experimental conditions may have made such a process particularly likely to occur. First, subjects were very experienced with the entire stimulus set that was used in the categorization conditions. Categorization learning was preceded by an identification paradigm in which all stimulus set members were used. Therefore, subjects knew the full range of stimulus possibilities. Another important aspect of the present experiment is that the stimuli were perceptually confusable. In hindsight, it seems obvious that the stimulus-response mapping learned by the subject under these conditions might not match exactly the one defined by the experimenter. Assume that on a given trial Stimulus S_i is presented and feedback for Category C_K is provided. If the subject encoded S_i as S_j (because of perceptual confusability), then a mapping between S_j and C_K would be reinforced. In a natural way, then, stimuli similar to the original training exemplars would also become part of a subject's category representation. General inspection of the partitions in Figure 8 reveals that the inferred exemplars tend to be more similar to members of their own category than to members of the opposite category, as predicted by the perceptual-confusability hypothesis.

Another important point is that the unassigned stimuli were presented repeatedly during the transfer phase. It is plausible that given multiple presentations, subjects are led to make initial decisions about category membership and these initial decisions influence subsequent ones. Consider, for example, the following

Table 5
Augmented GCM Parameters and Fits

Partition	Parameters			Summary fits		
	c	w_1	b_1	−ln L	SSE	% Var
		Subject 1				
Dimensional						
D1 1111111122222222	1.099	.000**	.444	6.616	.002	99.92
Criss-cross						
CC1 2211221111221122	1.381	.578	.510	42.266	.060	96.69
Interior–exterior						
IE1 2222211221122222	1.327	.637**	.642	31.307	.052	93.62
Diagonal						
DG1 1112112212221222	1.163	.601*	.641	16.461	.013	99.36
DG2 1112111212221222	1.252	.727**	.494	32.160	.036	98.22
		Subject 2				
Dimensional						
D1 1111111122222222	1.142	.089*	.563	10.833	.015	99.31
Criss-cross						
CC1 2211221111221222	1.552	.664**	.594	35.701	.067	94.52
CC2 2111121111211122	1.708	.295**	.269	44.077	.074	93.92
CC3 2111221111211122	1.511	.491	.349	77.223	.132	89.19
CC4 2111221111221122	1.404	.430	.427	77.234	.133	89.07
CC5 2211221111221122	1.356	.542	.516	87.121	.153	87.47
Interior–exterior						
IE1 2222211211122222	1.758	.661**	.561	35.494	.094	88.69
IE2 2222211221122222	1.525	.482	.639	54.226	.147	82.24
Diagonal						
DG1 1112112211221222	1.493	.558	.500	21.543	.025	98.98
DG2 1112112212221222	1.489	.758**	.653	33.093	.049	97.96

Note. GCM = generalized context model; −ln L = − log likelihood; SSE = sum of squared deviations between observed and predicted categorization probabilities; % Var = percentage of variance accounted for.
Partitions are shown in sequential format and can be decoded with the following key:

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

* Value of w_1 is significantly different from .5 ($p < .05$).
** Value of w_1 is significantly different from .5 ($p < .01$).

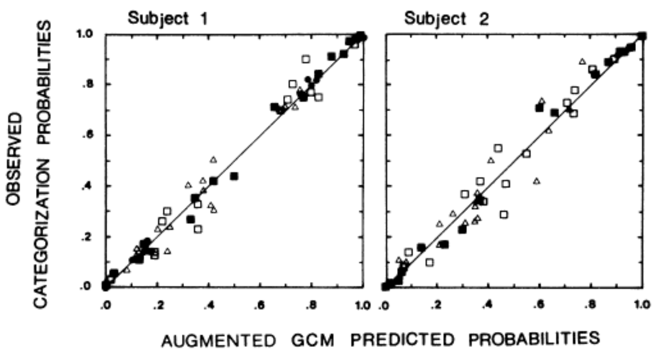

Figure 9. Observed classification proportions plotted against predicted (augmented generalized context model; GCM) categorization proportions for all the categorization conditions taken together. (Dot = dimensional; hollow square = criss-cross; triangle = interior-exterior; solid square = diagonal.)

process model. The initial category representation consists of only experimenter-assigned exemplars so that subjects' initial choices are governed by the simple GCM. Once an unassigned stimulus is classified in a given category some number of times, it is inferred to belong to that category. The category representation is then augmented to include this stimulus, and subsequent categorization decisions are governed by the augmented memory set. This process is very much in the spirit of the simple GCM, because memory augmentation is determined by the same principles of stimulus generalization that determine subjects' choices. Nosofsky (1984b) implemented this process as a computer simulation that was generally quite successful at predicting the best-fitting category partitions.

The ideas discussed thus far emphasize similarity-based determinants of the memory-augmentation process. An alternative idea is that subjects adopted "rules" by which to partition the stimulus set members. A rule-based approach might predict the category partitioning according to some criterion of "economy of description" or "simplicity of organization." General inspection of the partitions in Figure 8 reveals that they are highly organized (particularly those for Subject 1), although this is a difficult concept to formalize. The present experiment was not designed to distinguish between a similarity-based account and a rule-based account of the memory-augmention process, and both remain viable alternatives. Future research will need to explore in greater depth the nature and underlying basis of the memory-augmentation process. To the extent that it cannot be explained as a plausible outgrowth of exemplar-based generalization, there is the suggestion that additional processes of abstraction beyond those discussed in this article may mediate identification and categorization performance.

Theoretical Analysis of the Conditions A and S Data

Although Conditions A and S were referred to as identification conditions (see the Method section), they are more appropriately viewed as categorization conditions because they involved a many-to-one mapping of stimuli onto responses. The GCM (Equations 5, 4b, and 11) can be applied directly to analyze these data by simply letting C_J be the set of all stimuli having dimension level J. (In Condition A the relevant dimension is angle and in Condition S the relevant dimension is size.) Again, one needs to estimate the scale parameter c and the attention weight parameter w_1. Instead of estimating a single bias parameter, three bias parameters need to be estimated, with $b_4 = 1 - b_1 - b_2 - b_3$. Because there are 16 stimuli, and each stimulus can be classified in one of four categories, there are $16(4 - 1) = 48$ degrees of freedom in the data that are being accounted for by 5 free parameters.

The cumulative stimulus-response confusion matrixes for Conditions A and S are summarized for each subject in Table

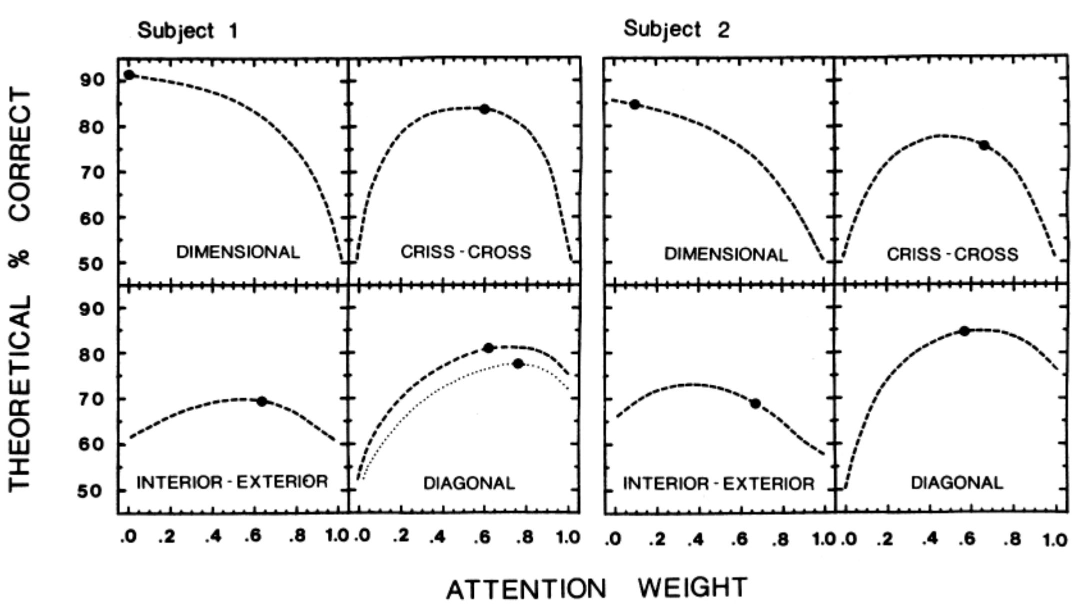

Figure 10. Augmented generalized context model optimization gradients for the attention weight parameter (w_1). (The dotted gradient for Subject 1's diagonal categorization is for partition DG2–see Table 5.)

6. The GCM was fitted to these confusion matrixes using a maximum-likelihood criterion. The best-fitting parameters and summary fits for each condition are presented in Table 7. Scatterplots of the predicted and observed categorization proportions are shown in Figure 11. With the exception of Subject 2's performance in Condition A, the GCM accounts quite well for these data.

The value of the attention weight parameter follows a systematic and easily interpretable pattern. For Subject 1, there is almost exclusive weighting of the size dimension in the size identification condition and almost exclusive weighting of the angle dimension in the angle identification condition. Subject 2 shows a similar pattern, although the tendency to weight size in Condition S is not as extreme as for Subject 1.

The discussion thus far has not considered the value of the scale parameter c. Let $\eta_{ij}^{(I)}$ denote the similarity between S_i and S_j in Identification Condition AS, and let $\eta_{ij}^{(C)}$ denote the similarity between S_i and S_j in one of the categorization conditions. Assuming conditions of nondifferential selective attention, then, as the model is currently parametrized, we have

$$\eta_{ij}^{(I)} = e^{-[(x_{i1}-x_{j1})^2+(x_{i2}-x_{j2})^2]}; \quad (13a)$$

$$\eta_{ij}^{(C)} = e^{-c^2[1/2(x_{i1}-x_{j1})^2+1/2(x_{i2}-x_{j2})^2]}. \quad (13b)$$

If similarity were invariant across the identification and categorization paradigms, as might be expected under conditions of nondifferential selective attention, then we would have $c = \sqrt{2}$. General inspection of the scale parameter estimates in Table 5 for the criss-cross, interior–exterior, and diagonal categorizations reveals no systematic deviations from this value, with Subject 1's estimates tending to be slightly lower and Subject 2's slightly higher. In the conditions in which there is the most evidence of differential selective attention (A, S, and the dimensional categorization), however, the value of c is consistently lower than $\sqrt{2}$ (see Tables 5 and 7). This pattern is probably not coincidental. Under the current parameterization, it is assumed that the attention weights add up to 1. In other words, as a subject ignores information from one dimension, there is a concomitant gain in the information extracted from the other dimension. The lowered scale parameter estimates would make sense if this "gain–loss"

Table 6
Stimulus-Response Confusion Matrixes for Conditions S and A

	Subject 1				Subject 2			
S_i	C1	C2	C3	C4	C1	C2	C3	C4
				Condition S				
1	60	34	1	0	75	32	4	0
2	72	36	1	0	61	29	4	1
3	75	21	0	0	87	35	2	0
4	79	20	0	0	60	31	4	0
5	2	77	20	0	28	55	26	1
6	10	74	24	0	26	47	27	1
7	9	58	25	0	16	46	17	1
8	10	65	20	0	13	44	21	3
9	0	10	82	7	6	31	51	15
10	0	8	87	6	4	27	53	15
11	1	7	88	10	1	29	50	19
12	0	16	73	6	1	22	63	13
13	0	1	28	78	0	7	34	72
14	0	0	33	71	0	2	25	55
15	0	0	35	54	1	2	32	80
16	0	0	38	68	0	0	27	66
				Condition A				
1	64	24	3	0	88	11	0	0
2	13	70	15	1	42	53	6	0
3	3	39	58	18	4	48	40	9
4	0	7	53	65	0	8	50	44
5	72	21	0	0	66	25	2	0
6	12	49	23	0	17	58	17	1
7	1	28	54	21	1	22	69	20
8	0	2	52	47	0	4	36	55
9	60	30	2	0	68	45	3	0
10	17	62	32	2	12	53	39	3
11	2	17	46	21	0	15	56	28
12	0	4	39	54	0	0	27	65
13	62	26	1	0	44	49	13	0
14	8	63	26	1	4	30	44	9
15	1	11	74	23	0	5	37	57
16	0	3	20	59	0	0	7	90

Note. C1, C2, C3, and C4 refer to categories 1–4; S_i = stimulus i.

Table 7
Best-Fitting GCM Parameters and Summary Fits for Conditions A and S

	Parameters						Summary fits		
Condition	c	w_1	b_1	b_2	b_3	b_4	−ln L	SSE	% Var
				Subject 1					
Size(S)	1.191	.036	.146	.264	.417	.173	19.521	.069	98.76
Angle(A)	.987	.903	.227	.296	.284	.194	25.234	.113	97.66
				Subject 2					
Size(S)	1.275	.264	.252	.267	.260	.220	22.204	.074	97.84
Angle(A)	1.002	1.000	.197	.243	.274	.287	138.399	.751	81.52

Note. GCM = generalized context model; −ln L = − log likelihood; SSE = sum of squared deviations between observed and predicted probabilities; % Var = percentage of variance accounted for.

process were not fully compensatory. The specification of quantitative relationships between the scale parameter and the distribution of attention weights is an important question for future investigation.

General Discussion

The major achievement in this research was a demonstration that one could make excellent predictions of categorization performance given knowledge of performance in an identification paradigm. Moreover, the theoretical analysis assumed that essentially the same underlying process of exemplar-based generalization operated in both paradigms. Confusion errors in both the identification and categorization paradigms were assumed to be a direct reflection of the form of the multidimensional space in which the stimuli were embedded. Stimuli close together in the space, and therefore similar to one another, were confused more often than stimuli far apart in the space.

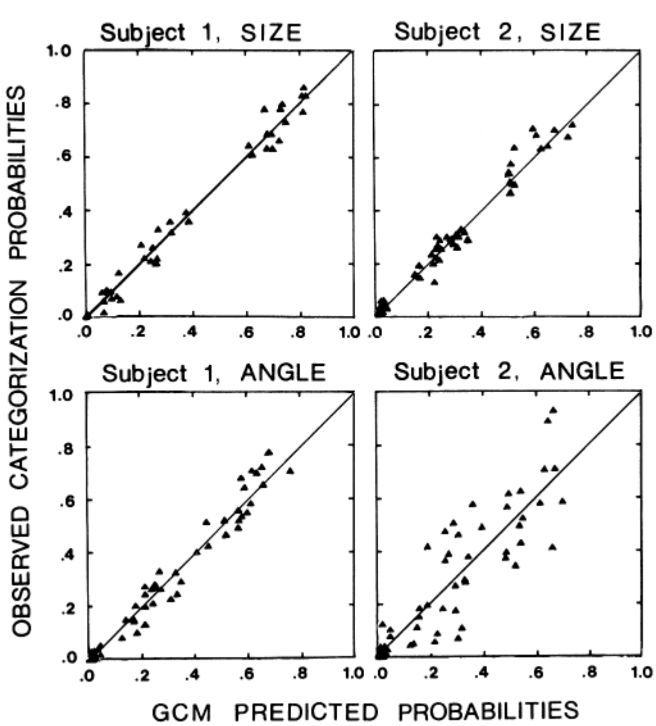

Figure 11. Observed classification proportions for Conditions A and S plotted against predicted generalized context model classification proportions.

In an earlier investigation, Shepard et al. (1961) rejected an exemplar-based generalization model for predicting categorization performance from identification performance. They showed that confusion errors in an identification task, which presumably gave an indication of interstimulus similarity relations, could not be used directly to generate successful categorization predictions. As an alternative, Shepard et al. suggested that subjects learned to attend selectively to the relevant stimulus dimensions in a given categorization problem and formulated rules about how values on these dimensions interacted.

The key difference between the exemplar-based generalization model investigated by Shepard et al. and the one advanced in this article regards the notion of similarity invariance. According to the present interpretation, similarity is not an invariant relation but a context-dependent one. In developing the present theory, I followed Shepard et al.'s lead by assuming that subjects attend selectively to relevant stimulus dimensions. Unlike Shepard et al., however, I maintained the assumption that categorization decisions are based on similarity to stored exemplars. Because of selective attention to component dimensions, there will be systematic changes in the structure of the psychological space in which the exemplars are embedded, and associated changes in similarity relations. With inclusion of assumptions about the role of selective attention in determining stimulus similarity, an exemplar-based generalization model for relating identification and categorization performance may be tenable.

Prospects for Generality

An important criticism of the present study concerns the issue of limits to generality of the results. Although a variety of category structures was sampled, the research examined the behavior of only two subjects classifying the same set of 16 artificially constructed stimuli. And the categorization tasks were introduced only after the subjects had extensive experience with the entire stimulus set.

The main goal of the present report was to set forth the basic theoretical approach and to illustrate a detailed quantitative application under simplified experimental conditions. That the approach holds promise of more general application is evidenced by several concurrent lines of research. First, note that Medin and Schaffer's context model is a special case of the present theory and has now been applied successfully to account for categorization performance using a variety of different stimuli including geometric forms, schematic faces, photographs, and random dot patterns. In an extension of previous work conducted by Shepard and Chang (1963), Nosofsky (1984b, 1985a) used the GCM to characterize accurately the relationship between subjects' identification and categorization of a set of stimuli varying along continuous integral dimensions. And in ongoing research in my laboratory, successful replications of the work reported herein are being achieved using large numbers of subjects who have relatively little experience with the stimulus set.

The present approach is also applicable in the domain of natural object classification. In a previous experiment using names of animals as stimuli, Sternberg and Gardner (1983, Experiment 1) presented subjects with a series of classification problems. In each problem, three animals were presented as a category and four additional transfer animals were presented. Subjects rank-ordered the transfer animals according to the degree to which they belonged in the category. Sternberg and Gardner used a multidimensional scaling solution for the animals to generate predictions of the ranking probabilities. First, they computed the centroid (or "prototype") for each category of animals in the multidimensional space. Next, they computed the distance between each of the transfer animals and the category centroid, transformed these distances into similarities, and then applied the MDS-choice model to predict the ranking probabilities (see Sternberg & Gardner, 1983, pp. 82–84, for a detailed description of the model-fitting procedure). This prototype model yielded excellent predictions of the subjects' classification rankings. The GCM can also be used to predict the classification rankings. Instead of computing the similarity between each of the transfer animals and the category centroid, one computes the sum of similarities between each of the transfer animals and the individual category exemplars. As it turns out, for the particular classification problems chosen by Sternberg and Gardner, the prototype model and exemplar-based GCM make virtually identical predictions. Currently, I am conducting a follow-up to Sternberg and Gardner's study that attempts to distinguish the predictions of the prototype and exemplar models in this natural object domain. Reports of this work will be forthcoming.

Tversky's Diagnosticity Principle

One of the keys to generating successful categorization predictions in the present study was the idea that selective attention modifies similarity relations across the identification and categorization paradigms. Tversky (1977) and Tversky and Gati (1978) have provided demonstrations of the context-dependent nature of similarity, and, more important, of the way in which an adopted classification can influence similarity. Indeed, the attention-optimization hypothesis studied in this article may be considered an explicit quantitative version of Tversky's (1977) *diagnosticity principle.* According to this principle, subjects attend to the features of a stimulus that have classificatory significance. Similarity relations, in turn, depend on which features are attended.

The notion that similarity is not an invariant relation because of the influence of selective attention has important bearing on numerous fundamental psychological theories. As examples, I now discuss implications of this notion of similarity noninvariance for multidimensional scaling theory and prototype models of categorization.

Attention, Similarity, and Multidimensional Scaling Theory

The analysis of the Condition AS identification data pointed toward a Euclidean metric for describing psychological distance relationships among a set of separable-dimension stimuli. This result is at odds with almost all previous conclusions regarding the appropriate distance metric for separable-dimension stimuli. Garner (1974), for example, summarized a great deal of research suggesting that the city-block metric provides a more adequate description than the Euclidean metric. Tversky and Gati (1982) advanced an idea that is even more at odds with the Euclidean assumption, namely, that no distance metric is appropriate for describing psychological distance relationships among separable-dimension stimuli. In a series of experiments that used highly analyzable stimuli, Tversky and Gati (1982) provided evidence of systematic violations of the triangle inequality, one of the basic metric axioms. It was pointed out by Tversky and Gati (1982) that a value of r less than 1 in the Minkowski r-metric formula could account for these violations. Note that a value of r less than 1 moves even further away from the Euclidean assumption (r = 2) than does the city-block metric (r = 1).

Much of the work on the appropriate psychological distance metric has used direct judgments of similarity. As suggested previously by other researchers, there may be complex cognitive processes and strategies involved in these judgment tasks (Krumhansl, 1978; Shepard, 1958a; Torgerson, 1965). Torgerson (1965, p. 383), for example, suggested that

> . . . as one adds more and more perceptual structure to a set of stimuli, the process underlying the judgments of similarity changes from what appears to be a rather basic perceptual one, to one which contains more and more cognitive features. And as the contribution of cognition goes up, the appropriateness of the multidimensional representation goes down.

Following this line of argument, I propose that a multidimensional scaling approach with the assumption of a Euclidean distance metric may provide an adequate *representational* model of the similarity structure of a given stimulus set. Operating on this similarity representation, however, may be rather complex attention and decision processes. It is important to determine whether the behavioral output in a given task is a direct reflection of the underlying similarity representation or of attention and decision processes that operate on this representation.

Note that for the identification and categorization models studied in this article, a combination of representation and process was specified. The perceptual representation was summarized by a multidimensional scaling solution for the stimulus set. Attention processes were posited to operate on this perceptual representation. These were described by attaching weights to the component dimensions. Finally, the decision rule was assumed to take the form of the similarity choice model. It was within the confines of this representation–process combination that support for the Euclidean metric was obtained.

Given this line of argument, a natural question arises regarding similarity judgment tasks: What processes may operate in these tasks that have led investigators to favor the city-block metric over the Euclidean one for separable-dimension stimuli? Sjoberg and Thorslund (1979) made a suggestion that is relevant here, namely, that in making similarity judgments, subjects carry out an active search for the ways in which stimuli are similar. This idea could be formalized by assuming that the similarity judgment is some transformation of a weighted Euclidean metric, in which subjects weight more heavily the dimensions along which a given pair of stimuli are more similar. For example, letting SJ_{ij} denote the similarity judgment for Stimuli S_i and S_j, one might write

$$SJ_{ij} = G[f(\sqrt{w_1(i,j)(x_{i1} - x_{j1})^2 + w_2(i,j)(x_{i2} - x_{j2})^2})], \quad (14)$$

where f is some monotonically decreasing function relating similarity to psychological distance, and G is some transformation of similarity onto the judgment scale. Note that the weight parameters in Equation 14 depend explicitly on the pair of stimuli being judged. Higher weight would be given to the dimension along which the stimuli are more similar.

Suppose that this model were correct. If a researcher analyzed a set of similarity judgment data—without taking into account the selective weighting of the component dimensions—the results would point away from a Euclidean metric and toward the city-block model. The reason is that the value of r in the Minkowski r-metric formula functions to weight differentially the component unidimensional distances in computing overall multidimensional distance (see Tversky & Gati, 1982, p. 150). For $r = 1$, the component distances are weighted equally. As the value of r increases above 1, the larger component distance is weighted more heavily, whereas as the value of r decreases below 1, the smaller component distance is weighted more heavily. In general, then, as one moves away from the Euclidean metric toward values of $r = 1$, smaller component distances are weighted more heavily than they were under the Euclidean model. Thus, the city-block metric may be mimicking the selective-attention similarity-judgment model formalized in Equation 14.

The Euclidean metric was supported in the identification and categorization studies reported in this article because within any given condition the weightings of the component dimensions were constant. The weightings changed only *across* conditions. In contrast, in a similarity judgment task the weights are not constant within a condition. They can vary systematically as a function of the pair of stimuli being judged. This seems sensible at an intuitive level. For example, in judging the similarity of Cuba and Red China one may focus on a political dimension, whereas in judging the similarity of Cuba and Jamaica one may more likely focus on dimensions of size, geography, and climate.

Tversky (1977) has provided interesting demonstrations of various other failures of the assumptions that underlie multidimensional scaling theory as an approach to modeling stimulus similarity. Although a systematic examination of these demonstrations goes beyond the scope of the present article, I note that many of the demonstrations may be tapping not the underlying similarity representation but attention and decision processes that operate on this representation.

Attention, Similarity, and Prototype Models

Among the most well-known findings in the categorization literature are protoypicality effects (e.g., Posner & Keele, 1968, 1970; Rosch, 1973; Rosch, Simpson, & Miller, 1976). These effects have led some researchers to posit that subjects abstract the central tendency, or prototype, of a set of category exemplars, and that categorization decisions are based on the similarity of stimuli to the abstracted prototypes. Assume that there are N_J exemplars in category C_J. Then, as formalized by Reed (1972), the prototype for category C_J is given by

$$\begin{aligned} P_j &= (p_{j1}, p_{j2}, \ldots, p_{jn}) \\ &= (\sum_{i=1}^{N_j} x_{i1}/N_j, \sum_{i=1}^{N_j} x_{i2}/N_j, \ldots, \sum_{i=1}^{N_j} x_{in}/N_j), \end{aligned} \quad (15)$$

where the summations in Equation 15 are being carried out over all stimuli that are members of category C_J. Thus, the category prototype is simply the centroid for all points in the multidimensional space that are associated with the category. According to prototype theory, the centroid *is* the category representation. Various decision rules can be formulated that operate on these representations. For example, one might compute weighted distances between test stimuli and the category prototypes, transform these distances into similarities, and then apply the similarity choice model.

Regardless of the decision rule, however, the prototype representation would be insufficient as a basis for categorization in the experiments reported here. Consider, for example, the criss-cross categorization structure. Note that the centroids for the Category 1 members and the Category 2 members virtually overlap. The same situation arises for the interior–exterior categorization. Obviously, if the category representations cannot be discriminated, it matters little what decision processes operate on them.

In the prototype model, a summary representation is formed by summing information independently over the component dimensions. Such a model will often have problems when the exemplars of a given category have correlated values on their component dimensions. The criss-cross categorization gives a simple illustration of this type of situation.

Although the present data favor the exemplar-based approach over the prototype model, some previous research points in the opposite direction. It is sometimes found, for example, that a category prototype is classified more accurately than the actual category exemplars, even though the prototype was never presented during training. Researchers have noted that this type of result is not necessarily inconsistent with an exemplar model (Hintzman & Ludlam, 1980; Medin & Schaffer, 1978). A category prototype would be expected to be fairly similar to almost all the category exemplars. In contrast, a particular exemplar may be highly similar only to itself. Thus, accurate categorization of a prototype does not imply that the prototype was abstracted and stored in memory.

Prototype theorists have tried to dispel this type of explanation by controlling for similarity (Homa, Sterling, & Trepel, 1981;

Posner & Keele, 1968, 1970; Reed, 1972). Posner and Keele (1968), for example, showed that a category prototype was classified more accurately than control patterns on a transfer test. Previous scaling work indicated that the control patterns and the prototype were equated in terms of perceived distance from the category exemplars. Posner and Keele (1968, pp. 361–362) argued that their results showed ". . . that the maximal generalization for multidimensional patterns of this sort occurs at the prototype even though other patterns are nearly the same average distance from the stored exemplars."

A conceptual problem with this argument is that it rests on the assumption that perceived distance is an invariant relation, independent of the experimental context. A major point advanced in this article is that perceived similarity (or distance) is *not* invariant across different paradigms. It was hypothesized that in any given choice context, subjects will distribute attention among the psychological dimensions that compose the stimuli so as to optimize performance and that this leads to systematic changes in similarity relations.

It is reasonable to argue that the category prototype may share with the stored exemplars more of the features that are useful for discriminating between classes than do the control patterns. If attention were focused on these criterial features, then from a psychological standpoint, the category prototype would be more similar than the control patterns to the stored category exemplars. This hypothesis is difficult to evaluate in many of the experiments that have been used to bolster prototype models because the psychological structure of the stimuli is often left unspecified. I suggest that convincing evidence of the idea that a prototype is abstracted and stored in memory will require a firm understanding of the psychological dimensions along which the stimuli are organized.

Ill-Defined and Well-Defined Categories

An interesting feature of the experiments used to test the generalized context model in this article is that the category structures were fairly "well defined." That is, the categories could be described in terms of simple rules. Medin and Schaffer (1978) originally proposed the context model with more "ill defined" category structures in mind. That the exemplar-based generalization model appears to provide accurate quantitative accounts of performance for both ill-defined and well-defined categories is a point in its favor. Of course, the research is still in its preliminary stages. More rigorous tests are needed and modifications of theory will undoubtedly continue to take place. Nevertheless, the present work suggests the intriguing possibility that an exemplar-based generalization model—that makes provision for the role of selective attention in determining similarity—may provide a unified framework by which to understand ill-defined and well-defined categorization and their relation to stimulus identification.

References

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking.* New York: Wiley.

Busemeyer, J. R., Dewey, G. I., & Medin, D. L. (1984). Evaluation of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 638–648.

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., & Wish, M. (1974). Models and methods for three-way multidimensional scaling. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2) San Francisco: W. H. Freeman.

Cheng, P. W., & Pachella, R. G. (1984). A psychophysical approach to dimensional separability. *Cognitive Psychology, 16,* 279–304.

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 VIII* (pp. 361–380). Hillsdale, NJ: Erlbaum.

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

Hintzman, D. L., & Ludlam, G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory & Cognition, 8,* 378–382.

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

Kornbrot, D. E. (1978). Theoretical and empirical comparison of Luce's choice model and logistic Thurstone model of categorical judgment. *Perception & Psychophysics, 24,* 193–208.

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.

Lockhead, G. R. (1970). Identification and the form of multidimensional discrimination space. *Journal of Experimental Psychology, 85,* 1–10.

Lockhead, G. R. (1972). Processing dimensional stimuli: A note. *Psychological Review, 79,* 410–419.

Lockhead, G. R., & King, M. C. (1977). Classifying integral stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 436–443.

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

Martin, R. C., & Caramazza, A. (1980). Classification in well-defined and ill-defined categories: Evidence for common processing strategies. *Journal of Experimental Psychology: General, 109,* 320–353.

Medin, D. L. (1983). Structural principles in categorization. In T. J. Tighe & B. E. Shepp (Eds.), *Perception, cognition, and development: Interactional analysis* (pp. 203–230). Hillsdale, N.J.: Erlbaum.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 37–50.

Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 3,* 333–352.

Medin, D. L., Dewey, G. I., & Murphy, T. D. (1983). Relationships between item and category learning: Evidence that abstraction is not automatic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 607–625.

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

Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning.

Journal of Experimental Psychology: Human Learning and Memory, 7, 241–253.
Monahan, J. S., & Lockhead, G. R. (1977). Identification of integral stimuli. *Journal of Experimental Psychology: General, 106,* 94–110.
Nosofsky, R. M. (1984a). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 104–114.
Nosofsky, R. M. (1984b). Attention, similarity, and the identification-categorization relationship. Unpublished doctoral dissertation, Harvard University.
Nosofsky, R. M. (1985a). *Attention and learning processes in the identification and categorization of integral stimuli.* Manuscript submitted for publication.
Nosofsky, R. M. (1985b). Luce's choice model and Thurstone's categorical judgment model compared: Kornbrot's data revisited. *Perception & Psychophysics, 37,* 89–91.
Nosofsky, R. M. (in press). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception & Psychophysics.*
Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.
Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83,* 304–308.
Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3,* 382–407.
Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language.* New York: Academic Press.
Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 491–502.
Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika, 22,* 325–345.
Shepard, R. N. (1958a). 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. (1958b). Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review, 65,* 242–256.
Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology, 1,* 54–87.
Shepard, R. N., & Chang, J. J. (1963). Stimulus generalization in the learning of classifications. *Journal of Experimental Psychology, 65,* 94–102.
Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75,* (13, Whole No. 517).
Sjoberg, L., & Thorslund, C. (1979). A classificatory theory of similarity. *Psychological Research, 40,* 223–247.
Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.
Smith, J. E. K. (1980). Models of identification. In R. Nickerson (Ed.), *Attention and performance VIII.* Hillsdale, NJ: Erlbaum.
Sternberg, R. J., & Gardner, M. K. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology: General, 112,* 80–116.
Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: Wiley.
Torgerson, W. S. (1965). Multidimensional scaling of similarity. *Psychometrika, 30,* 379–393.
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 models of visual letter recognition. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 834–864.
Townsend, J. T., & Landon, D. E. (1982). An experimental and theoretical investigation of the constant-ratio rule and other models of visual letter confusion. *Journal of Mathematical Psychology, 25,* 119–162.
Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.
Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization.* Hillsdale, NJ: Erlbaum.
Tversky, A. & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review, 89,* 123–154.

Received February 19, 1985
Revision received June 4, 1985 ■

New Look for the APA Journals in 1986

Beginning with this issue, the APA journals have a new look. All the journals are 8¼ × 11 inches—a little larger than the *American Psychologist* is now. This change in trim size will help reduce the costs of producing the journals, both because more type can be printed on the larger page (reducing the number of pages and amount of paper needed) and because the larger size allows for more efficient printing by many of the presses in use today. In addition, the type size of the text will be slightly smaller for most of the journals, which will contribute to the most efficient use of each printed page.

These changes are part of continuing efforts to keep the costs of producing the APA journals down, to offset the escalating costs of paper and mailing, and to minimize as much as possible increases in the prices of subscriptions to the APA journals.