

# Expanding the search for a linear separability constraint on category learning

MARK BLAIR and DON HOMA  
Arizona State University, Tempe, Arizona

Formal models of categorization make different predictions about the theoretical importance of linear separability. Prior research, most of which has failed to find support for a linear separability constraint on category learning, has been conducted using tasks that involve learning two categories with a small number of members. The present experiment used four categories with three or nine patterns per category that were either linearly separable or not linearly separable. With overall category structure equivalent across category types, the linearly separable categories were found to be easier to learn than the not linearly separable categories. An analysis of individual participants' data showed that there were more participants operating under a linear separability constraint when learning large categories than when learning small ones. Formal modeling showed that an exemplar model could not account for many of these data. These results are taken to support the existence of multiple processes in categorization.

Grouping objects into categories on the basis of their similarity is a primary cognitive task. To the extent that categories are not separable on the basis some combination of perceptual information, they might be expected to be harder to learn. Formal models of categorization differ in their predictions about whether categories that are not linearly separable (LS) can ever be learned fully. The main issue addressed in this paper is whether there is a linear separability constraint on category learning. Evidence for a linear separability constraint can be taken as evidence for certain kinds of categorization processes and, therefore, certain kinds of models. After defining linear separability, we will discuss some categorization processes that are constrained by linear separability and some that are not. We then will explain some of the complexities in determining what sort of data provide evidence of a linear separability constraint on category learning. Finally, we will review the range of categories learned in previous research and show how the range is expanded in the present attempt to find evidence for a linear separability constraint.

## Linear Separability

LS categories are those categories that can be separated in two-dimensional stimulus space by a linear discriminant

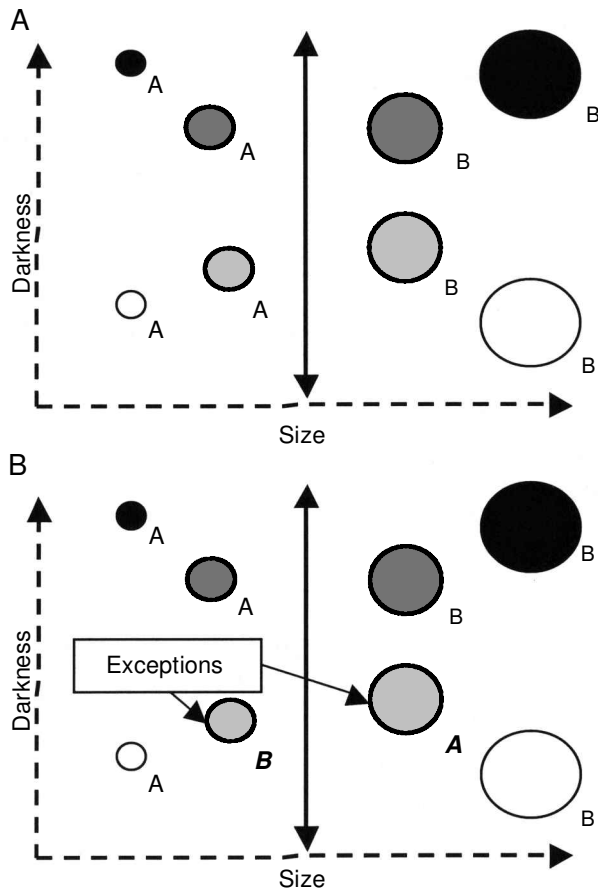
function or in  $N$ -dimensional stimulus space by a hyperplane. Figure 1A shows an example of LS categories in a simple category space with two dimensions. Stimuli—in this case, circles—vary in size across the horizontal axis and vary in darkness along the vertical axis. All the members of Category A are on one side of the line, and all the members of Category B on the other, so these categories are LS. Figure 1B shows a similar space, except that a member of Category A has been switched with a member of Category B. Now there is no line that can divide the category members into their correct categories; these categories are not linearly separable (NLS). Any attempt to separate the categories with a line will result in stimuli on the opposite side of the line from their fellow category members. These stimuli, which we will call *exceptions*, are very important for deciding whether or not there is a linear separability constraint on category learning.

## Categorization Processes: Some Constrained, Some Not

NLS categories should be impossible to learn with prototype processes alone. Prototype theory holds that categories are represented in memory by the central tendency of all the category members, a representation called the *prototype* (e.g., Posner & Keele, 1968; Reed, 1972). In order for a new exemplar to be judged a member of a particular category, it must be more similar to the prototype of that category than to the prototype of any other category. Figure 2 illustrates this idea. The distance ( $d_A$ ) between a new exemplar and the prototype of Category A ( $P_A$ ) is compared with the distance ( $d_B$ ) between the new exemplar and the prototype of Category B ( $P_B$ ). Because  $d_A < d_B$ , the new exemplar will more likely be classified into Category A. Between any two categories, the line connect-

---

Portions of this manuscript were presented at the Western Psychological Association meeting in Irvine, CA, April 1999. Portions of this manuscript were presented by the first author in partial fulfillment of the degree of Masters of Arts at Arizona State University. We thank Susan Somerville, Michael McBeath, Stephen Goldinger, Jacqueline Blair, Syd Norns, and two anonymous reviewers for their thoughtful comments on earlier drafts of this paper. Correspondence concerning this article should be addressed to M. Blair, Department of Psychology, Arizona State University, Tempe, AZ 85287-1104 (e-mail: mark.blair@asu.edu).



**Figure 1.** (A) An example of linearly separable categories. A linear function divides the categories accurately. (B) An example of not linearly separable categories. No line exists that could accurately partition these two categories. Using a line to divide these categories always leaves exception patterns on the side opposite their proper category.

ing the two prototypes has a perpendicular bisector that acts as a linear decision bound. On one side of the decision bound are the exemplars classified into Category A; on the other side of the decision bound are the exemplars classified into Category B. Exception patterns appear on the other side of the decision bound from their prototype and would, therefore, be misclassified. A prototype model thus predicts that there will be a linear separability constraint on category learning such that NLS categories can never be learned completely with a categorization process based only on prototypes, because the exception patterns would always be misclassified. Any categorization process that involves a weighted, additive combination of features makes the same prediction (Wattenmaker, Dewey, Murphy, & Medin, 1986). Contrasting the predictions of prototype theory with the predictions of Medin and Schaffer's (1978) context model, which was the first quantitative exemplar model of categorization, was the main purpose of the first search for a linear separability constraint (Medin & Schwanenflugel, 1981).

Exemplar models (e.g., Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986) determine a category choice on the basis of the weighted similarity of the current stimulus to other stimuli in memory. By using exemplar processes, the exception patterns in an NLS category can be learned if they are sufficiently similar to other patterns in memory associated with the correct category, even including past presentations of that identical stimulus. Other categorization processes—for example, connectionist networks using pairwise feature inputs (Gluck, 1991; Gluck & Bower, 1988; Shanks, 1991)—also predict that exception patterns are possible to learn.

Even though exemplar processes confer the ability to learn the exception patterns of NLS categories, actual predictions about performance are not always simple to intuit, and it is not accurate to say that exemplar models always predict that NLS categories and LS categories can be learned equally fast. Swapping a pattern from Category A and a pattern from Category B, as in Figure 1, creates NLS categories that are less structured than the original LS categories. The patterns that changed categories are now less similar to the other patterns in their category (lower within-category similarity), and more similar to the patterns in the other category (higher between-categories similarity). Category structure, commonly defined as the ratio of within-category similarity to between-categories similarity (e.g., Homa, Rhoads, & Chambliss, 1979; Smith, Murray, & Minda, 1997), is therefore dramatically reduced. Category structure is easily confounded with linear separability. Under an exemplar model, the exception patterns in this example would be less likely to be categorized correctly than the normal (nonexception) patterns. Equating category structure between LS categories and NLS categories is necessary to provide some control for this problem. Even then, the exemplar model's predictions of the exact degree to which the exceptions are harder to learn than the normal items depends, in complex ways, on the similarities between patterns and is difficult to determine without running simulations of a formal model.

To summarize, a participant using only prototype processes will be unable to learn the exception patterns in NLS categories, whereas a participant using only exemplar processes can. In order to have an informative comparison of the performance of participants on LS and NLS categories, category structure must be equated. The learning data from such an experiment should also be modeled formally by an exemplar model, to examine whether the data can be predicted with exemplar processes. These steps were taken in the study reported here.

### What Counts as Evidence for a Linear Separability Constraint?

Typically, the importance of a linear separability constraint on category learning has been evaluated on the basis of a comparison of performance between NLS and LS categories (Medin & Schwanenflugel, 1981; Wattenmaker et al., 1986). However, overall learning rate can be mis-

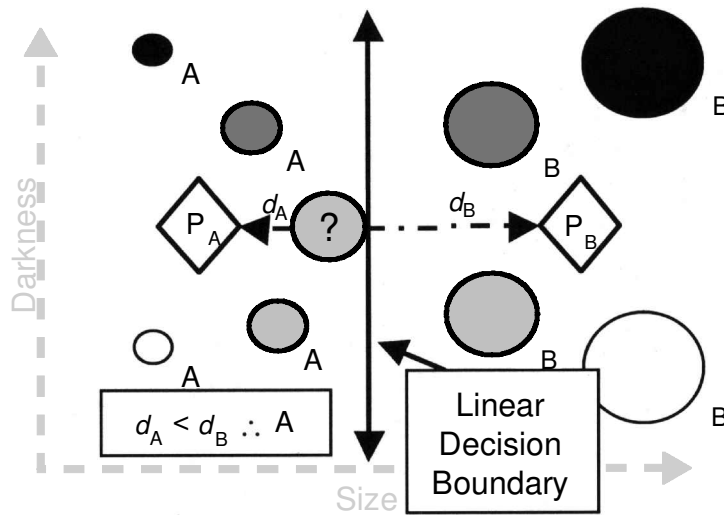


Figure 2. An illustration of prototype theory. Distance to the prototype is used to classify new exemplars, in effect creating a linear decision bound midway between the category prototypes.

leading (Smith et al., 1997). The reasoning for using this measure is that if exception patterns are impossible to learn, or even very difficult, there should be more errors in the learning of NLS categories than in the learning of LS categories. But, if obtained data show that NLS categories are harder to learn, it is still possible that an exemplar model fits the data, showing that processes that predict a linear separability constraint are not a necessary part of the explanation. Conversely, failure to obtain differences in performance between NLS and LS categories does not rule out the possibility that there is a linear separability constraint active. Consider an experiment in which one is comparing performance on LS and NLS categories with three members. Participants learning the LS categories categorize 66% of the patterns correctly. Participants learning NLS categories, with two normal patterns and one exception pattern, correctly classify 100% of the normal patterns and 0% of the exceptions. On the basis of a comparison of group means, we would be forced to conclude that there was no difference between the 2 participants' performance, even though 100% correct on normal patterns and 0% correct on the exceptions is the strongest possible indication of a linear separability constraint.

Averaging across participants within a single experimental condition creates the same problems of interpretation (Smith et al., 1997). Some participants may be learning both the normal patterns and the exception patterns reasonably well, which would suggest that there is not a linear separability constraint. Others may be learning normal patterns very well but consistently misclassifying the exceptions, which would suggest that there is a linear separability constraint. With aggregate data, important differences between participants are lost. Evaluating the data for a linear separability constraint requires that

performance on the normal and exception patterns be evaluated at the individual participant level.

**Expanding the Search**

In the original research on a linear separability constraint on category learning, which failed to find evidence for such a constraint, Medin and Schwanenflugel (1981) wrote: "To our knowledge, the present series of studies represents the first controlled comparisons of learning LS and NLS categories. Therefore, cautions concerning the generality of the results are even more pertinent than is usually the case" (p. 366). Despite this warning, further

**Table 1**  
Summary of Previous Research on Linear Separability

Experiment	Categories		Structure	
	No.	Size	LS	NLS
Medin and Schwanenflugel (1981)				
Experiment 1	2	4	1.36	1.32
Experiment 2	2	4	1.23	1.37
Experiment 3	2	4	1.25	1.25
Experiment 4	2	3	1.25	1.25
Wattenmaker, Dewey, Murphy, and Medin (1986)				
Experiments 1 and 2	2	4	1.23	1.37
Experiment 3	2	4	1.67	1.00
Smith, Murray, and Minda (1997)				
Experiment 1*	2	7	1.29	1.29
Experiment 1†	2	7	1.83	1.83

Note—Medin and Schwanenflugel (1981) used unique stimuli that were identical on their values on the three or four relevant dimensions for their Experiments 3 and 4. In one sense, category size was still limited to three or four patterns per category, but in another sense the category size was infinite. LS, linearly separable categories; NLS, not linearly separable categories. Category structure is the number of shared features within the category relative to the number of shared features between categories. \*Low structure condition. †Moderate structure condition.

research used learning tasks that were very similar to those of Medin and Schwanenflugel, as can be seen in Table 1. Considering all of the possible combinations of category size, category structure, and number of categories, the range that has been explored is quite limited, and concerns about the generality of the results are still pertinent (Smith & Minda, 1998; Smith et al., 1997).

Smith and colleagues (Smith et al., 1997) expanded on earlier research, using increased category size and structure. By using stimuli that varied on six binary-valued dimensions, rather than just three or four, they were able to increase the total number of possible patterns, which allowed them to use larger and more structured categories. They showed that participants learning these categories produced data that were fit better by a prototype model than by an exemplar model, and more of the participants' data showed evidence of an active linear separability constraint for their task than for Medin and Schwanenflugel's (1981) tasks. They argued that the small and poorly structured categories of previous research had elicited exemplar processes, including explicit memorization of whole exemplars, that do not operate under a linear separability constraint.

We take seriously the idea that small categories may elicit exemplar memorization strategies and inhibit prototype strategies. To avoid this potential problem, the present study used a different kind of stimuli that allowed us to use more categories and larger categories in the learning task. The stimuli used were abstract line shapes, also called random-dot polygons, that were distortions of one of four prototype shapes. Because a large number of unique patterns can be generated from a particular prototype, the limitations in category size and in the number of categories that are a result of using stimuli that vary on a small number of binary-valued dimensions were avoided (Minda & Smith, 2001; Smith & Minda, 1998). Increasing the number of categories also increases the total number of stimuli seen and thwarts the strategy of simply memorizing the members of one of the two categories. Using random-dot polygons also allows for increased generality, sorely needed in research on this topic.

In the present study, linear separability (LS and NLS) and category size (three and nine) were manipulated as between-subjects variables. Each participant placed the stimuli into one of four categories; for example, the participants in the NLS 3 condition learned patterns from four different NLS categories, each containing three category members. Generalizing from the results of Smith et al. (1997), we predicted that a linear separability constraint on category learning would be more evident in this study than in previous research. We also predicted that the impact of a linear separability constraint on participants' performance would be more pronounced for the larger categories (NLS 9).

In the NLS conditions, two thirds of the patterns, the normal patterns, were from the same prototype, whereas the remaining patterns, the exceptions, were generated from the other prototypes. Thus, in the NLS 3 condition, there were two normal patterns and one exception pattern

in each of the four categories, and in the NLS 9 condition, there were six normal patterns and three exception patterns in each of the four categories. In order to control category structure between the NLS and the LS categories, the patterns in the LS conditions were more distorted, making them less similar to each other. This step lowered the within-category similarity and allowed for equivalent structure.<sup>1</sup>

## METHOD

### Participants

The participants were 139 introductory psychology students from Arizona State University, who participated in partial fulfillment of a class requirement. The participants were assigned randomly to conditions. Thirteen subjects were excluded from the analysis because they made numerous responses with reaction times of less than 250 msec. This was taken to indicate that they were answering without attempting to complete the task as instructed.

### Stimuli

The stimuli were members of four abstract shape categories generated from a prototype shape. Prototypes were created by randomly placing nine dots on a  $50 \times 50$  grid, then connecting each point to the next with a line, until the figure became completely closed. Exemplars were created by changing the location of the nine points of the prototype. The new location of each point of the new distorted pattern was determined by sampling from a distribution whose mean and mode were the prototype's values for that point. These derivative shapes were high-level distortions when the standard deviation of the distribution was large and low-level distortions when it was small. On average, each point was moved 4.6 Euclidean units for high-level distortions and 1.2 units for low-level distortions (see Homa, 1978, for a full description).

### Multidimensional Scaling

To control category structure between the LS and the NLS categories at each level of category size, a multidimensional scaling (MDS) was done, and categories were built using the resulting normalized distance in psychological space (Kruskal, 1964; Shepard, 1962).

The participants were 15 introductory psychology students from Arizona State University, who participated in partial fulfillment of a class requirement. They were to judge the dissimilarity between all possible pairs of 24 stimuli on a scale from 1 to 9. These stimuli consisted of three high-level and three low-level distortions of each of the four prototypes used in the experiment. The specific patterns were among the experimental stimuli described above. The matrix of mean dissimilarity ratings was analyzed using an MDS technique. Using the three-dimensional solutions, the resulting mean normalized psychological distance between two low-level distortions from the same prototype was 0.28, and the distance between two high-level distortions from the same prototype was 1.20. The between-prototypes distance was averaged across the two distortion levels, which were comparable, and was 1.50. Category structure was then calculated, using the average of the within-category distances, including exemplar self-similarity of zero, over the average between-categories distances.<sup>2</sup> Table 2 shows the structural ratios for each experimental condition.

### Linear Separability

Linear separability was manipulated by using stimuli from one prototype per category in the LS conditions and by using stimuli from more than one prototype per category in the NLS conditions. For the LS categories, high-level distortions were used, and for the

**Table 2**  
Average Within-Category and Between-Categories MDS Distances of Exemplars in Each Condition

Distance	LS 3	NLS 3	LS 9	NLS 9
Within	0.80	0.73	1.07	0.88
Between	1.50	1.32	1.50	1.29
Ratio	0.53	0.55	0.71	0.68

Note—LS 3, three-member, linearly separable categories; NLS 3, three-member, not linearly separable categories; LS 9, nine-member, linearly separable categories; NLS 9, nine-member, not linearly separable categories

NLS categories, low distortions were used. The NLS 3 categories were formed by combining two patterns generated from the same prototype, with one pattern generated from a different prototype. The ratio of exceptions to normal patterns was kept constant across category sizes; so, in the NLS 9 condition, each category had six patterns generated from the same prototype and three exception patterns, one generated from each of the other prototypes. A different one of the prototypes was used for each of the three exception patterns in each NLS 9 category to avoid any confounding that might be due to subordinate categories that could arise from having a three-exemplar cluster within each larger NLS category. Therefore, each NLS 9 category contained distortions of all four prototypes. An example of one of the categories used is shown for each of the four conditions in Figure 3. It should be noted that the examples shown in this figure are significantly smaller than those viewed by the participants and, so, distinguishing between the low-level distortions was much easier for the participants than it will be for the

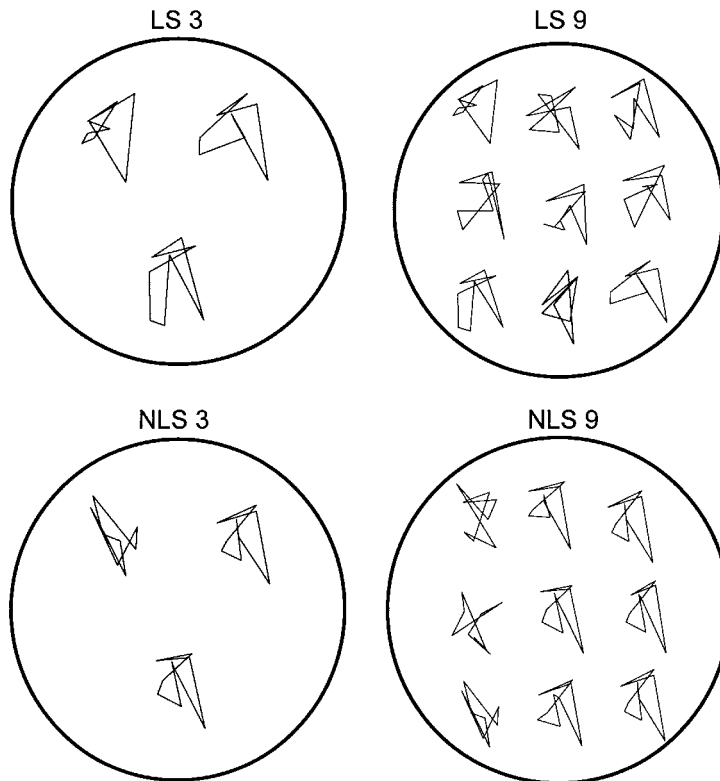
reader. Each of the 36 low-level distortions used in the NLS conditions was shown as a normal pattern for some participants and as an exception for others.

**Procedure**

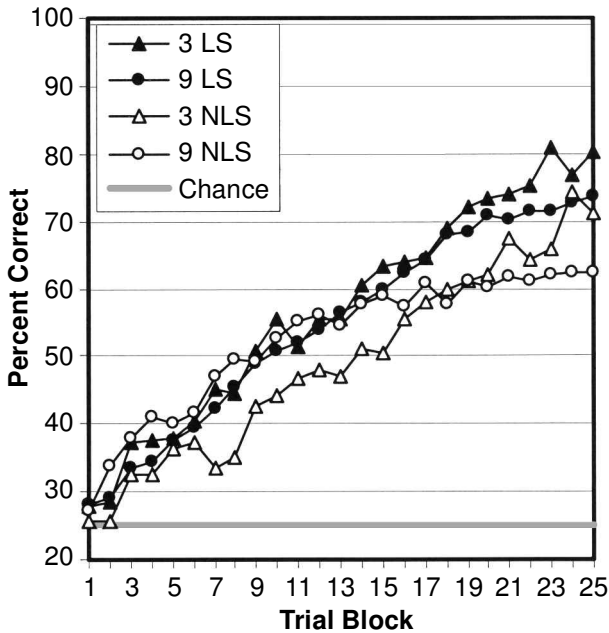
The participants performed the experiment on a computer in a sound-dampening chamber, seated approximately 16 in. from a 14-in. computer monitor. They responded by using the A, B, C, and D keys of a standard keyboard. The patterns were displayed one at a time in white on a black background and were approximately 5 × 5 in. in size. Learning was self-paced, although the participants were told to spend approximately 2–10 sec per pattern. After the participant responded with a choice of category, the correct category (A, B, C, or D) was displayed to the right of the pattern for 1 sec; then, a new pattern was shown. Each trial block consisted of a randomly ordered presentation of all the stimuli for that condition. Every participant performed 25 trial blocks, with no break between blocks. The participants in the category size 3 conditions made a total of 300 responses, which took about 20 min. The participants in the category size 9 conditions made a total of 900 responses, which took about 60 min.

**RESULTS**

Consistent with previous studies, we first will compare overall performance on LS and NLS categories. Then we will examine individual participants' performance on the normal items and exception items for signs of a linear separability constraint. Finally, we will compare these data with the predictions of an exemplar model.



**Figure 3.** Examples of the stimuli for Category A for each of the four conditions. The LS categories are constructed of high distortions of a single prototype. The NLS categories are constructed of low distortions of multiple prototypes.



**Figure 4.** The learning curves for each condition. 3 LS = three-member, linearly separable categories; 3 NLS = three-member, not linearly separable categories; 9 LS = nine-member, linearly separable categories; 9 NLS = nine-member, not linearly separable categories.

The data show that the LS categories were easier to learn than the NLS categories. The learning curves for each of the four conditions, plotted as percent correct, are shown in Figure 4. These percent correct data were analyzed by using a three-way analysis of variance (ANOVA), with category size (3, 9) and separability (LS, NLS) as between-subjects variables and trial block (1–25) as a within-subjects variable. There was a significant three-way interaction between trial block, category size, and separability [ $F(24,2928) = 1.913, p < .01$ ; using the Greenhouse–Geisser sphericity correction,  $F(9,147,1122.092) = 1.913, p = .0451$ ]. Separate analyses were then conducted for the category size 3 and the category size 9 conditions. The ANOVA for the category size 3 conditions revealed no significant interactions, but there was a significant main effect for separability [ $F(1,63) = 5.077, p < .05$ ], with better performance in the LS condition. The ANOVA for the category size 9 conditions showed a significant interaction between trial block and separability [ $F(24,1416) = 7.521, p < .001$ ]. As is shown in Figure 4, performance in the NLS 9 condition seemed to level off over the last half of the experiment, resulting in the poorest performance of all the groups on the final trial. All other conditions show relatively stable rates of learning throughout the 25 trial blocks.

For both the category size 3 and the category size 9 conditions, overall performance was better in the LS conditions than in the NLS conditions. This is the result that both Medin and Schwanenflugel (1981) and Smith et al. (1997)

failed to obtain. An increased number of categories, an increased length of training, an increased proportion of exception items, and the use of integral dimensional stimuli all distinguish this research from previous studies. These modifications all may have contributed to the performance advantage found in this study for the participants learning LS categories.

Although there was poorer performance on the NLS categories than on the LS categories, it is unclear from the preceding analysis, as was discussed in the introduction, whether the participants in the NLS condition were placing the exceptions into the category most similar to their respective prototypes, which would be expected if the participants were operating under a linear separability constraint. It is also unclear to what extent individual participants followed the aggregate pattern. In order to clarify these issues, the data from the participants in the NLS conditions were analyzed individually.

### Exception Pattern Classification by Participant Subgroups

In the following analysis for participants in the two NLS conditions, performance on the exception patterns is broken up into three groups: correct classifications to the feedback category, errors to the prototype category, and errors to the other two categories. The feedback category for a particular exception is the category in which the participant is told that the pattern belongs. The prototype category for a particular exception pattern is the category whose normal patterns are distorted from the same prototype as the exception. This is the category into which we would expect them to be classified if there was a linear separability constraint.

On the basis of the proportion of times in the last set of trial blocks that exception items were categorized by the participant into the prototype category, the feedback category, or the other two categories, the participants break down into four groups: random respondents, feedback respondents, split respondents, and prototype respondents. The random respondents had no discernible pattern to their categorization of the exception items and had very poor overall performance. The feedback respondents placed more exception patterns into the feedback category than into the prototype category. Split respondents placed exceptions into the prototype and feedback categories in roughly equal proportions. Prototype respondents placed more exception patterns into the prototype category than into the feedback category.

In the NLS 3 condition, there were 5 random respondents, 17 feedback respondents, 7 split respondents, and 3 prototype respondents. With the exception of the random respondents, the data were averaged across participants within each group and are graphed in Figure 5. This figure shows the proportions of trials on which the participants classified the exception items into the prototype (incorrect) category, the feedback (correct) category, and the remaining two categories, which are treated as a single group. The proportion of trials on which the participants

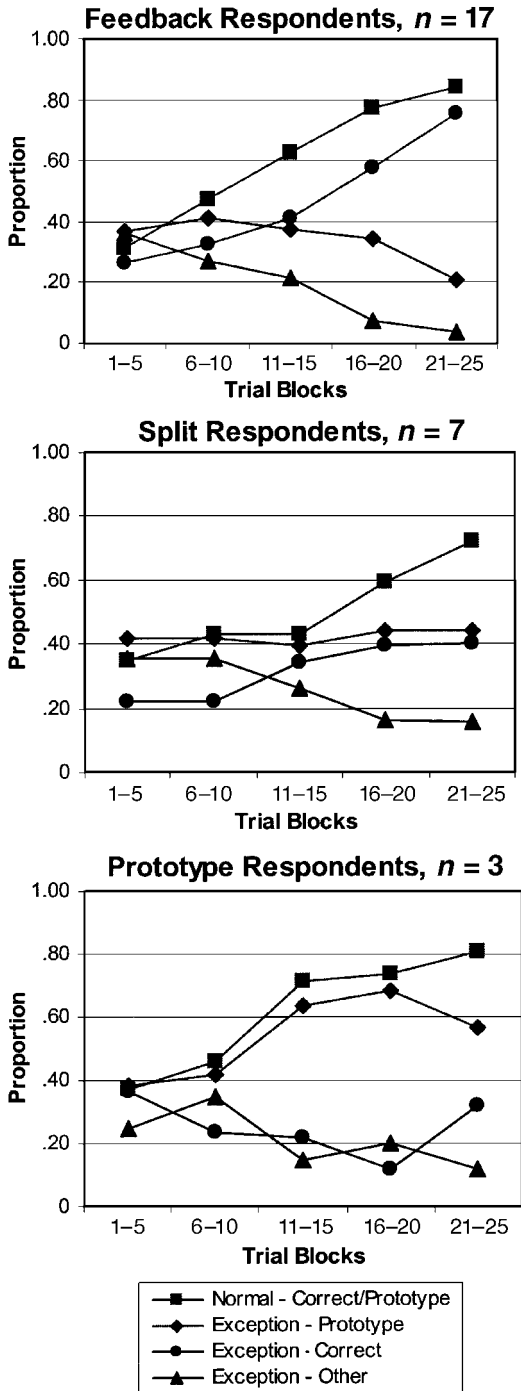


Figure 5. The performance profiles for participant subgroups in the NLS 3 condition. Random respondents ( $n = 5$ ) were not included.

classified the normal patterns into the prototype (correct) category is shown for comparison.

The error pattern of the feedback respondents in the NLS 3 condition shows no signs of a linear separability constraint. For these participants, more exceptions were

categorized into the prototype category than into the feedback category at first, but not by much and not for long. Over trial blocks 16–20 and 21–25, the total percentages correct (71%, 82%) for these participants learning the NLS 3 categories was better than those for the participants learning the LS 3 categories (69% and 77%, respectively). If all of the participants had performed like this, there would not have been an LS advantage in the category size 3 conditions. The 7 split respondents also began classifying the patterns into the prototype categories, although over the last half of the experiment, they were classifying nearly as many exceptions into the feedback category as into the prototype category. The 3 prototype respondents categorized the normal and the exception patterns into the prototype category roughly the same proportion of the time. This is consistent with performance under a linear separability constraint. During the last block of trials, they showed improvement in classifying the exception patterns, although they continued to classify many more exceptions into the prototype category than into the feedback category.

In the NLS 9 condition, there were 3 random respondents, 1 feedback respondent, 2 split respondents, and 25 prototype respondents. Excluding the random respondents, the data were averaged across participants within each group and are graphed in Figure 6.

The lone feedback respondent placed the highest proportion of exception patterns into the feedback category during the late stages of learning. This participant did, however, show an early tendency to categorize both normal and exception patterns into the prototype category, which was much stronger than the tendency shown by most of the 17 feedback respondents in the NLS 3 condition. The large early disparity between the proportion of exception patterns placed into the prototype category and the proportion placed into the feedback category indicates that a linear separability constraint was active even though this participant had some success overcoming it.

The 2 participants who fell into the split respondents category again began with the same strong tendency to categorize the normal and exception patterns into their prototype categories but ended their learning by using the prototype and the feedback categories equally for the exception patterns. Unlike the NLS 3 split respondents, who split between the feedback and the prototype categories for more than half the experiment, these participants used the prototype category for most of the classifications, until the last set of trial blocks.

The remaining 25 participants were prototype respondents and, following the trend of the other NLS 9 participants, showed a more dramatic usage of the prototype category. With these participants, the improvement on the exception items near the end of learning was slight, and performance on these items was still well below chance (14%). The feedback displayed after each pattern did not seem to be enough to allow these participants to learn the exception patterns. These participants were not able to overcome the linear separability constraint.

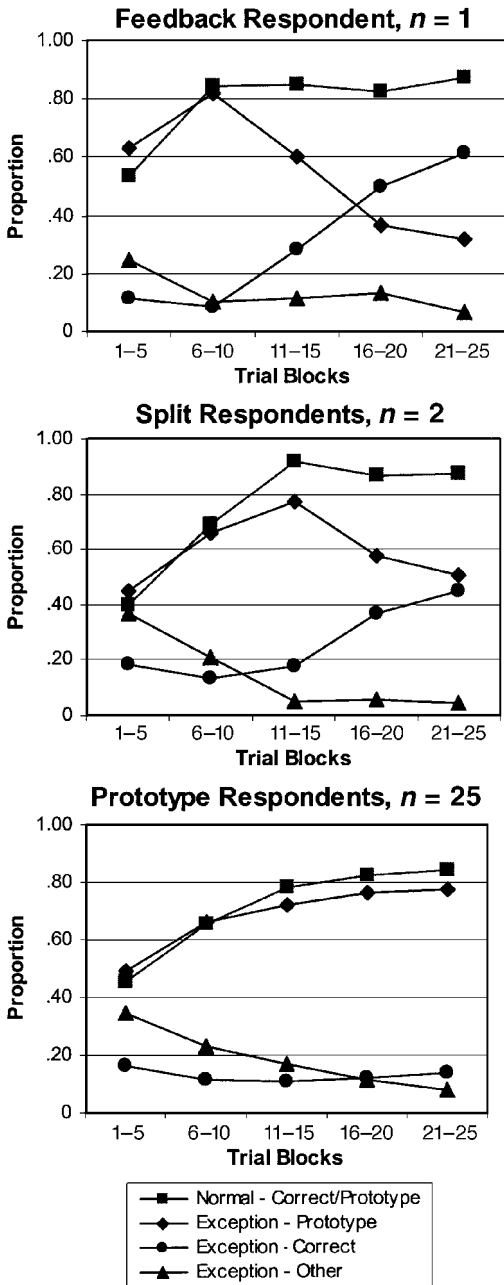


Figure 6. The performance profiles for participant subgroups in the NLS 9 condition. Random respondents ( $n = 3$ ) were not included.

Unlike the NLS 3 condition, in which the majority of participants showed performance patterns inconsistent with an active linear separability constraint, the NLS 9 condition yielded performance patterns consistent with an active linear separability constraint for 25 out of 32 participants and an active linear separability constraint in the first part of the experiment for 3 others. The next step

in the analysis is to establish whether an exemplar model can account for these data.

**Exemplar Model Predictions**

In this section, the data from the different participant groups discussed above is plotted against the predictions of an exemplar model. The model used to generate the predictions is a version of Nosofsky's (1986) generalized context model, which was modified by Maddox and Ashby (1993):

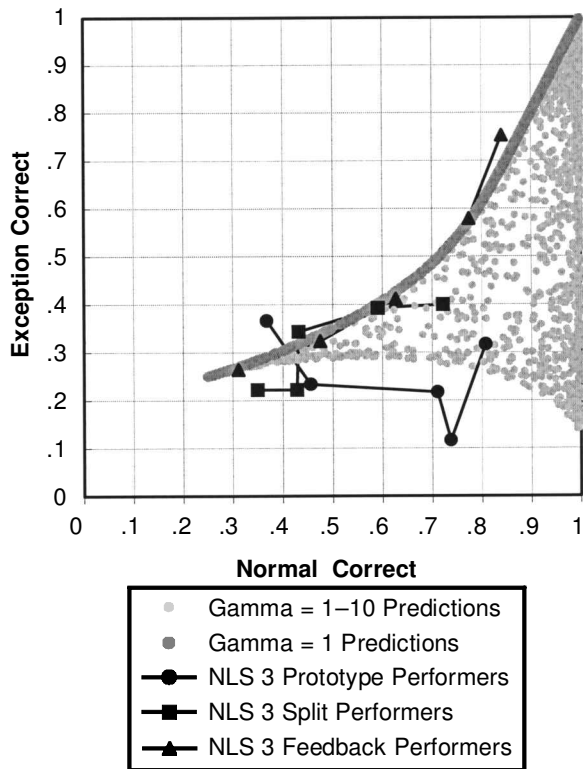
$$P(R_1 | S_i) = \frac{(\sum_{i \in C_1} \eta_{ij})^y}{(\sum_{i \in C_1} \eta_{ij})^y + (\sum_{i \in C_2} \eta_{ij})^y + (\sum_{i \in C_3} \eta_{ij})^y + (\sum_{i \in C_4} \eta_{ij})^y}$$

where similarity between pattern  $i$  and pattern  $j$  is an exponential decay function of the MDS distance  $d_{ij}$  between them:

$$\eta_{ij} = e^{-cd_{ij}}$$

The two free parameters were the response-scaling parameter,  $\gamma$ , which scales between a probability-matching decision rule when  $\gamma = 1$  and more deterministic responding when  $\gamma > 1$ , and a sensitivity parameter,  $c$ , which reflects discriminability in the psychological space. The predictions for each condition were generated twice, once using the original version of the model, (i.e., holding  $\gamma$  constant at 1) and once using the augmented version (i.e., allowing  $\gamma$  to take values from 1 to 10). The model predictions for the NLS 3 condition were generated using the distances derived in the MDS of the stimuli and 5,000 random configurations of the free parameters. For each version of the model, the sensitivity parameter was allowed to vary from 0 to 22. The upper bound of 22 was the smallest integer that allowed perfect performance with  $\gamma$  held at 1. A similar analysis was performed for the NLS 9 condition, except that the sensitivity parameter was allowed to vary from 0 to 27. The upper bound of 27 was the smallest integer that allowed perfect performance with  $\gamma$  held at 1. The predictions of the probability-matching exemplar model for both conditions will be discussed first and then compared with the predictions of the version that allows more deterministic responding.

Figure 7 shows the predictions of the exemplar model for the NLS 3 condition, along with the observed data from the three participant subgroups. These data are plotted in five block sets, as in Figures 5 and 6, such that the first datapoint for a series (the one farthest to the left) is the average percent correct of trials 1-5, the next point is the average for trials 6-10, and so on. The prediction space of the original version of the exemplar model, which has one free parameter in this case, is the curved dark line that goes from chance to perfect performance. The prediction space of the version augmented with  $\gamma$  is mapped out by the smaller gray dots.



**Figure 7.** The predictions of the original probability-matching exemplar model ( $\gamma = 1$ ) and the more deterministic version that allows variation in the response-scaling parameter ( $\gamma = 1-10$ ) and the observed data for the NLS 3 categories. The prediction space was generated using 5,000 random configurations of the similarity and response-scaling parameters. The respondent groups' data is plotted in blocks of five trials, beginning with trials 1-5 at the leftmost point of the series.

For the NLS 3 condition, the feedback respondents' (17 of 32 participants) data were well described by the probability-matching version of the exemplar model ( $\gamma = 1$ ). The data of the split respondents (7 of 32 participants) and the prototype respondents (3 of 32 participants) fell outside the prediction line of the probability-matching exemplar model. These are the participants who show a more active linear separability constraint, so the model's failure is expected.

For the NLS 9 condition, none of the datapoints fall on the predictions of the probability-matching exemplar model (see Figure 8).<sup>3</sup> Like the NLS 3 condition, and as predicted, the prototype respondents' (25 of 31 participants) data are far removed from the predictions of the model. Even though performance on the exception patterns did improve for the split respondents (2 of 31 participants) and a feedback respondent (1 of 31 participants), it improved only after performance on the normal patterns had improved to over 80% and performance on the exceptions had dropped below 20%, thus placing their learning path off the exemplar model's predictions. The data from this condition show a heavy linear separability

constraint, and so it is expected that the exemplar model should fail to predict the data from this condition. The best fit for the exemplar model is for the last two sets of trial blocks by the feedback respondent. This participant had some success in learning the exception patterns. The better fit may reflect a switch to using exemplar processes, a change that is necessary for the participant to eventually learn the exception patterns. If so, it is interesting that most of the participants are unable to make the switch.

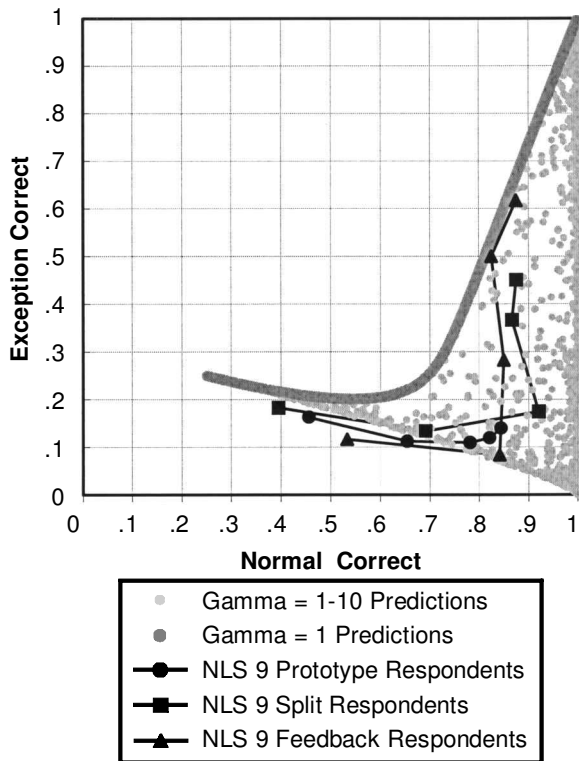
The  $\gamma$ -augmented exemplar model fit more of the data. For the NLS 3 condition, the prediction space spreads out to include the data from the last 10 trial blocks for the split respondents and from the last 5 trial blocks for the prototype respondents. For the NLS 9 condition, the more deterministic exemplar model does a much better job. By the third set of trial blocks, the data from all three groups of respondents fall inside the prediction space of the model.

Although the  $\gamma$ -augmented version of the exemplar model, with an additional free parameter, predicts more of the datapoints than does the simpler version, the prediction fits and failures do not make as much intuitive sense. The best fit for the 17 NLS 3 feedback respondents occurs at  $\gamma = 1$ . A larger  $\gamma$  means a worse fit, especially later in learning. With the other respondents in the NLS 3 conditions the opposite is true, a larger  $\gamma$  is necessary for good fits later in learning. Also, according to the interpretation of  $\gamma$  as probabilistic versus deterministic response rules, most of the participants in the NLS 3 condition are using probabilistic response rules, but in the NLS 9 condition, all the participants seem to be using more deterministic response rules. Why an increase in category size would change participants' decision rules in this way is unclear.

We have not found the  $\gamma$  model predictions to be particularly enlightening. Though more of the data are predicted by the model, the interaction between category size and the strength of the linear separability constraint on performance has no intuitive interpretation. Indeed, other recent work has pointed to similar issues regarding  $\gamma$  (Minda & Smith, 2001; Smith & Minda, 1998), although there is still considerable debate (Nosofsky & Johansen, 2000; Nosofsky & Palmeri, 1997).

**DISCUSSION**

There is strong evidence from the present experiment to support the hypotheses that a linear separability constraint on category learning exists and that its influence increases with category size. First, there was a performance advantage for the LS categories over the NLS categories. Second, there was a subset of participants, prototype respondents, who placed normal patterns and exceptions into the prototype category in roughly equal proportion, clearly showing a linear separability constraint. With the increase in category size, the proportion of respondents showing this pattern increased dramatically (9% in NLS 3, 81% in NLS 9). Formal modeling confirmed that an exemplar model alone could not predict many of these data.



**Figure 8.** The predictions of the original probability-matching exemplar model ( $\gamma = 1$ ) and the more deterministic version that allows variation in the response-scaling parameter ( $\gamma = 1-10$ ) and the observed data for the NLS 9 categories. The prediction space was generated using 5,000 random configurations of the similarity and response-scaling parameters. The respondent groups' data is plotted in blocks of five trials, beginning with trials 1-5 at the leftmost point of the series.

Why is the linear separability constraint not important in the Medin and Schwanenflugel (1981) categorization tasks, when it is found to exert a strong influence in the present study? According to prototype theory, there is a distinction between the memory representation of the category, which is the prototype, and the memory representation of any particular stimulus. A linear separability constraint applies only to classification decisions made using the prototype, not to decisions made on the basis of explicit memory for a specific pattern. The simplest and most intuitive explanation for the linear separability constraint's dependence on category size is that memory for specific patterns plays less of a role in performance for larger, more numerous categories.

In Medin and Schwanenflugel's (1981) experiments, it is possible to achieve perfect performance simply by memorizing the three or four members of Category A and responding B to everything else. In the present study, to reach perfect performance using a memorization strategy, a participant in the NLS 3 condition would need to memorize 9 different stimuli, and a participant in the NLS 9 condition would need to memorize 27 stimuli. Recent

work by Bourne, Healy, Parker, and Rickard (1998) on classification strategies suggests that a strategy's effectiveness and ease of use strongly influence when and whether participants will adopt the strategy. They showed that exemplar strategies dominated tasks where the rule strategy was difficult to implement, but they also showed that a rule strategy can dominate if the rule is simple to learn. They argued that increasing the number of training exemplars would increase the demand of an instance-based strategy and would make participants more likely to use a rule-based (or prototype-based) approach. The assertion that category size affects the use of exemplar strategies is well supported in the literature (Reed, 1978). Homa and colleagues showed the effect in multiple-category learning tasks with the random-dot polygons used in the present experiment (Homa, Dunbar, & Nohre, 1991; Homa, Sterling, & Trepel, 1981). Smith and Minda showed it in two-category learning tasks using letters and line-drawn bugs that varied on binary-valued dimensions (Minda & Smith, 2001; Smith & Minda, 1998; Smith et al., 1997). These studies also showed that prototype-based classification increases as exemplar-based performance decreases. The difference in the importance of linear separability across category sizes and numbers of categories fits perfectly with the evidence that exemplar processes are more important when learning small categories and prototype processes are more important when learning larger categories and suggests that the best explanation for these data will be a mixed model of classification that includes both the memorization of specific exemplars and the use of category prototypes (or rules).

Demonstrating that linear separability is important in larger categories leads one to be concerned about the proliferation of studies focusing exclusively on small categories (e.g., Lamberts, 1995; Medin & Schwanenflugel, 1981; Medin & Smith, 1981; Nosofsky, 1991). Obviously, understanding how people learn small categories is important. But just as obviously, data collected using tasks that elicit mostly exemplar processes are fundamentally limited as a basis for the development categorization theory. A broad range of tasks and stimuli is necessary to elicit the numerous strategies and processes participants can bring to bear in an experimental task, and a broad manipulation of important variables is necessary to generate comprehensive data (Homa, 1984). Mixed models of perceptual categorization acknowledge the existence of different participant abilities and so are obliged to address the problem of when these abilities are used. Fortunately, there is a growing number of mixed models (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998; Nosofsky & Palmeri, 1998; Nosofsky, Palmeri, & McKinley, 1994) and a growing concern for the importance of different processes at different stages of learning (Smith & Minda, 1998) and under different variable manipulations (Minda & Smith, 2001).

Mixed models, more generally, are being used in many domains. Babey, Queller, and Klein (1998) have proposed

a summary-plus-exception model of social memory. Lee, Chen, and Jiang (1998) developed an effective neural network classifier that uses distinct prototype nodes and exemplar nodes in order to learn categories that are nonlinearly separable. Using a model that incorporates category boundaries and prototypes, Huttenlocher and colleagues (Huttenlocher, Hedges, & Duncan, 1991) have explored category effects on spatial locations (Newcombe, Huttenlocher, Sandberg, Lie, & Johnson, 1999), estimates of time (Huttenlocher, Hedges, & Prohaska, 1988), and estimates of stimulus characteristics (Crawford, Huttenlocher, & Engebretson, 2000). Mixed approaches, using some combination of prototypes, decision boundaries, rules, or exemplars, are proving useful in accounting for data in many domains.

**Conclusion**

The issue of a linear separability constraint on category learning was first addressed experimentally by Medin and Schwanenflugel (1981). They found no evidence for a linear separability constraint when participants learned pairs of small categories. Prototype processes, and other processes that predict a linear separability constraint, therefore do not appear to drive categorization for all kinds of categories. The present study shows that there is a linear separability constraint active when participants learn greater numbers of large categories, showing that exemplar processes do not drive categorization for all kinds of categories. Taken together, these data sets provide strong motivation to explore multiple process accounts of categorization.

**REFERENCES**

ASHBY, F. G., ALFONSO-REESE, L. A., TURKEN, A. U., & WALDRON, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, **105**, 442-481.

BABEY, S. H., QUELLER, S., & KLEIN, S. B. (1998). The role of expectancy violating behaviors in the representation of trait knowledge: A summary-plus-exception model of social memory. *Social Cognition*, **16**, 287-339.

BOURNE, L. E., HEALY, A. F., PARKER, J. T., & RICKARD, T. C. (1998). The strategic basis of performance in binary classification tasks: Strategy choices and strategy transitions. *Journal of Memory & Language*, **41**, 223-252.

CRAWFORD, L. E., HUTTENLOCHER, J., & ENGBRETSON, P. H. (2000). Category effects on estimates of stimuli: Perception or reconstruction? *Psychological Science*, **11**, 280-284.

ERICKSON, M. A., & KRUSCHKE, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, **127**, 107-140.

GLUCK, M. A. (1991). Stimulus-generalization and representation in adaptive network models of category learning. *Psychological Science*, **2**, 50-55.

GLUCK, M. A., & BOWER, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, **117**, 227-247.

HOMA, D. (1978). Abstraction of ill-defined form. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 407-416.

HOMA, D. (1984). On the nature of categories. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18, pp. 49-94). San Diego: Academic Press.

HOMA, D., DUNBAR, S., & NOHRE, L. (1991). Instance frequency, cate-

gorization, and the modulating effect of experience. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 444-458.

HOMA, D., RHOADS, D., & CHAMBLISS, D. (1979). Evolution of conceptual structure. *Journal of Experimental Psychology: Human Learning & Memory*, **5**, 11-23.

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

HUTTENLOCHER, J., HEDGES, L. V., & DUNCAN, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, **98**, 352-376.

HUTTENLOCHER, J., HEDGES, L. V., & PROHASKA, V. (1988). Hierarchical organization in ordered domains: Estimating the dates of events. *Psychological Review*, **95**, 471-484.

KRUSCHKE, J. K. (1992). Alcové: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.

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

LAMBERTS, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General*, **124**, 161-180.

LEE, H. M., CHEN, K. H., & JIANG, I. F. (1998). A neural network classifier with disjunctive fuzzy information. *Neural Networks*, **11**, 1113-1125.

MADDOX, W. T., & ASHBY, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, **53**, 49-70.

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

MEDIN, D. L., & SCHWANENFLUGEL, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 355-368.

MEDIN, D. L., & SMITH, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 241-253.

MINDA, J. P., & SMITH, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 775-799.

NEWCOMBE, N., HUTTENLOCHER, J., SANDBERG, E., LIE, E. H., & JOHNSON, S. (1999). What do misestimations and asymmetries in spatial judgement indicate about spatial representation? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 986-996.

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

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

NOSOFSKY, R. M., & JOHANSEN, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, **7**, 375-402.

NOSOFSKY, R. M., & PALMERI, T. J. (1997). An exemplar-based random-walk model of speeded classification. *Psychological Review*, **104**, 266-300.

NOSOFSKY, R. M., & PALMERI, T. J. (1998). A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychonomic Bulletin & Review*, **5**, 345-369.

NOSOFSKY, R. M., PALMERI, T. J., & MCKINLEY, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, **101**, 53-79.

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

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

REED, S. K. (1978). Category vs. item learning: Implications for categorization models. *Memory & Cognition*, **6**, 612-621.

SHANKS, D. R. (1991). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 433-443.

- SHEPARD, R. N. (1962). The analyses of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, **27**, 125-140, 219-246.
- SMITH, J. D., & MINDA, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1411-1436.
- SMITH, J. D., MURRAY, M. J., & MINDA, J. P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 659-680.
- WATTENMAKER, W. D., DEWEY, G. I., MURPHY, T. D., & MEDIN, D. L. (1986). Linear separability and concept-learning: Context, relational properties, and concept naturalness. *Cognitive Psychology*, **18**, 158-194.

#### NOTES

1. Category structure is unavoidably confounded with category variance in this and all other linear separability research. NLS categories tend to have a number of patterns (normals) that have high within-category similarity to each other and one or more patterns (exceptions) that have low within-category similarity, whereas all the patterns in the LS categories tend to have roughly the same within-category similarity.

2. This ratio is different than other structural ratios in the linear separability literature, although it serves the same purpose. The structural

ratio calculated for binary-valued stimuli is based on the number of shared features, which yields larger numbers for similar stimuli and smaller ones for dissimilar stimuli. Therefore, ratios less than 1 represent categories that are unstructured, with more common features between categories than within category. Because distance in psychological space and similarity are inversely related, a structural ratio based on MDS distances approaches zero as within-category similarity increases and between-categories similarity decreases, and a ratio of over 1.0 indicates that the categories are unstructured.

3. It is possible that the exemplar model is very sensitive to small fluctuations in the MDS distances used to generate the predictions. This is especially relevant because the distances were obtained using less than the full set of stimuli. In order to verify that these failures to predict will not disappear if the MDS distances are changed slightly, the predictions were run again after the distance between low distortions of the same prototype was reduced from 0.28 to 0.1 and the distance between distortions of different prototypes was increased from 1.5 to 2.0. The exemplar model did not fit any more of the data.

(Manuscript received April 27, 2001;  
revision accepted for publication July 26, 2001.)