Journal of Experimental Psychology: General
1992, Vol. 121, No. 3, 278–304

0096-3445/92/$3.00

Similarity-Scaling Studies of Dot-Pattern Classification and Recognition

Hyun Jung Shin
Pusan National University
Pusan, Korea

Robert M. Nosofsky
Indiana University Bloomington

Classification performance in the dot-pattern, prototype-distortion paradigm (e.g., Posner & Keele, 1968) was modeled within a multidimensional scaling (MDS) framework. MDS solutions were derived for sets of dot patterns that were generated from prototypes. These MDS solutions were then used in conjunction with exemplar, prototype, and combined models to predict classification and recognition performance. Across 3 experiments, an MDS-based exemplar model accounted for the effects of several fundamental learning variables, including level of distortion of the patterns, category size, delay of transfer phase, and item frequency. Most important, the model quantitatively predicted classification probabilities for individual dot patterns in the sets, not simply general trends of performance. There was little evidence for the existence of a prototype-abstraction process that operated above and beyond pure exemplar-based generalization.

A classic theme in the categorization literature has involved a contrast between prototype and exemplar models. According to prototype models, subjects store an abstract summary representation of a category in memory. In formal models, this summary representation, or prototype, is usually defined as the central tendency of the category distribution (e.g., Ashby & Gott, 1988; Medin & Schaffer, 1978; Nosofsky, 1987; Reed, 1972). Classification decisions are based on the similarity of items to the abstracted prototype. By contrast, according to exemplar models, subjects store the individual training exemplars of a category in memory and make classification decisions on the basis of the similarity of items to these exemplars.

Two major research approaches have been used for testing and contrasting the predictions of prototype and exemplar models. Perhaps the most widespread experimental paradigm, dating back to at least the classic work of Posner and Keele (1968, 1970), has involved the use of dot-pattern and random polygon stimuli. In innumerable studies, the general procedure is to (a) start with dot-pattern prototypes that define each category; (b) generate various distortions of these prototypes by using a statistical-distortion rule; (c) train subjects in a learning phase by presenting these old distortions; and then (d) test subjects in a transfer phase that includes the old distortions, the prototypes, and various new distortions of the prototypes. As discussed, for example, by Homa (1984), a major advantage of these dot-pattern experiments is that the stimulus patterns are essentially infinitely variable and have a complex dimensional structure, so the properties of the artificial categories that are created seem to mimic those of real-world, natural categories.

Despite these advantages, there are several limitations associated with the dot-pattern experiments. First, many of the phenomena that were originally believed to provide evidence for a prototype-abstraction process, such as prototype-enhancement effects, resistance of the prototype to forgetting, category size effects, and so forth, have since been demonstrated to also be qualitatively consistent with the predictions of exemplar models (e.g., Busemeyer, Dewey, & Medin, 1984; Hintzman, 1986; Hintzman & Ludlam, 1980; Nosofsky, 1988a). Second, there is no clear understanding of the psychological similarity relations that hold among the various dot patterns. Objective measures of similarity are usually defined on the basis of the average distance between positions of corresponding dots in pairs of patterns. However, it is questionable whether the positions of the individual dots form the sole or primary psychological basis on which the patterns are coded. Rather, various emergent dimensions based on configurations and relations among dot positions are probably of critical importance (e.g., Hock, Tromley, & Polmann, 1988).

A second major approach to contrasting prototype and exemplar models, exemplified by the work of Reed (1972), Medin and his associates (e.g., Medin, Altom, & Murphy, 1984; Medin & Schaffer, 1978; Medin & Smith, 1981), Estes (1986), and Nosofsky (1987, 1988b, 1991b), has involved the use of simple perceptual stimuli varying along a few salient dimensions. In these studies, the relevant psychological dimensions are generally obvious and are often verified using multidimensional scaling (MDS) techniques. Exemplar and prototype models are formalized in quantitative fashion, and sharp contrasts between the predictions of the models can often be achieved. Indeed, in numerous recent studies, Nosofsky (1986, 1987, 1988b, 1989, 1991b; Nosofsky, Clark, & Shin, 1989) showed that a particular exemplar model known as the *generalized context model* (GCM), when used in con-

This article is based on a doctoral dissertation submitted by Hyun Jung Shin to Indiana University. The work was supported by Grants BNS 87-19938 from the National Science Foundation and PHS R01 MH48494-01 from the National Institute of Mental Health to Indiana University.

Correspondence concerning this article should be addressed to Robert M. Nosofsky, Department of Psychology, Indiana University Bloomington, Bloomington, Indiana 47405. Electronic mail may be sent to nosofsky@ucs.indiana.edu.

junction with MDS solutions for the exemplars, consistently yielded impressive quantitative fits to classification and recognition data that are superior to those of prototype models.

Nevertheless, these paradigms involving simple perceptual stimuli also have their limitations. Most salient among these limitations is that many researchers would view the category structures that are tested as unnatural and not very representative of the ill-defined categories found in the real world.

The central goal of our research was to bridge the gap between the dot-pattern studies and the second class of studies that have quantitatively contrasted prototype and exemplar models. The plan was to (a) generate categories of dot patterns following classic procedures adopted, for example, by Posner and Keele (1968, 1970), Homa (e.g., Homa, Cross, Cornell, Goldman, & Schwartz, 1973; Homa & Cultice, 1984; Homa, Sterling, & Trepel, 1981; Homa & Vosburgh, 1976), and others; (b) conduct preliminary similarity-judgment studies and use the resulting data to obtain MDS solutions for the patterns; and (c) use these derived scaling solutions, in conjunction with the competing models, to generate quantitative predictions of classification and recognition performance. Although some previous work on MDS of dot patterns has been performed (e.g., Homa & Cultice, 1984; Homa, Rhoads, Chambliss, 1979; Hyman & Frost, 1975), our project is the first to use such scaling solutions in conjunction with formal mathematical models to predict classification and recognition in quantitative detail.

The present scaling-based research has the potential for some significant advances over previous efforts to model classification in the dot-pattern paradigms. One well-known modeling technique, used by both exemplar and distributed-memory theorists (e.g., Hintzman, 1986; Knapp & Anderson, 1984; Metcalfe-Eich, 1982; Nosofsky, 1988a), is to randomly generate multi-element stimulus vectors and use these vectors in computer simulations of the dot-pattern experiments. The random vectors that are constructed are intended to mimic the dot-pattern stimuli, but little is known about the extent to which the random vectors capture the underlying psychological structure of the patterns. Rather than using objective measurements of positions of dots, our scaling-based research uses similarity-judgment data to position the patterns in multidimensional psychological spaces (cf. Shepard, 1962a, 1962b, 1987).

Other efforts to model dot-pattern classification have involved the application of models in which numerous free parameters were estimated to represent similarities among different types of patterns (e.g., Busemeyer et al., 1984; Homa, Dunbar, & Nohre, 1991; Homa et al., 1981). The numerous free similarity parameters were needed because the experiments provided no independent information about similarity relations among the patterns. Even with the numerous free parameters, these previous modeling efforts were limited to predicting fairly gross-level aspects of the data, such as average performance levels for the prototypes and old and new distortions. By contrast, our scaling-based research provides independent information about similarity relations among the patterns and thereby vastly reduces the number of free parameters that need to be estimated for predicting classification. Furthermore, the scaling allows measurement of similarities among all individual objects in the stimulus set. Thus, rather than simply predicting average levels of performance for the prototypes, old distortions, and new distortions, we have the capability of making quantitative predictions of classification for individual tokens of these main item types.

The exemplar model tested in this research is Nosofsky's (1986) GCM. According to the GCM, classification and recognition decisions are based on the summed similarity of an item to the exemplars of the alternative categories. Nosofsky (in press) has reviewed numerous studies in which the GCM, when used in conjunction with MDS solutions for the exemplars, has achieved impressive quantitative predictions of classification and recognition in highly simplified perceptual domains. Because the dot-pattern stimuli are undoubtedly psychologically complex, we do not expect the MDS solutions to perfectly account for their similarity structure. Therefore, we do not expect the quantitative predictions of the GCM to be as accurate as in previous work. Nevertheless, given the demonstrated successes of the scaling-based GCM in simple perceptual domains, and the vast amount of categorization research using the dot-pattern paradigm, our research direction seemed like an interesting and important one to pursue.

Experiment 1

The primary purpose of Experiment 1 was to quantitatively test the GCM and a prototype model on their ability to predict classification and old–new recognition performance for ill-defined categories composed of random dot patterns. One group of subjects engaged in a standard classification-learning experiment. Three categories of random dot patterns were constructed from each of three prototypes. Each category consisted of six old distortions. Following the training phase, subjects were tested in a transfer phase in which the old distortions, the prototypes, and various new distortions of the prototypes were presented. Subjects classified each pattern into one of the three categories and also judged whether each pattern was old or new.

A second group of subjects judged the degree of similarity between all pairs of exemplars, whereas a third group engaged in the classification-learning condition followed by the similarity-judgment condition. The purpose of collecting similarity-judgment data was to derive an MDS solution for the dot patterns. This MDS solution could then be used in conjunction with the GCM and the prototype model for predicting classification and recognition. In addition, a combined model was formalized and used in conjunction with the MDS solution to assess the relative contribution of exemplar and prototype information to classification performance.

The reason for collecting separate similarity-judgment data from two different groups was to investigate the influence of category-learning experience on the evolution of category structures. Homa et al. (1979) demonstrated that as category-learning experience increases, there are increases in the ratio of average judged within-category similarity to between-category similarity. Nosofsky (1984, 1986) suggested that changes in interexemplar similarities as a function of category learning may be the result of a selective-attention process, in which the psychological dimensions that compose the exemplars are differentially weighted. In Experiment 1 we explored whether changes in similarity judgments among exemplars as a function of category learning could be modeled in these terms.

Method

Subjects

The subjects were 70 undergraduates from Indiana University who were hired for participation. There were 30 subjects in the classification-only condition (Group C), 20 subjects in the similarity-judgment–only condition (Group S), and 20 subjects in the classification/similarity-judgment condition (Group C/S). All subjects were tested individually.

Stimuli and Apparatus

Three categories were used. Each category consisted of 10 dot patterns: a prototype, six training exemplars (the old distortions), and three new distortions of the prototype. The dot patterns were generated following Posner, Goldsmith, and Welton's (1967) procedure. Each pattern consisted of nine dots. Three prototypes were randomly generated, and six training exemplars were then constructed from each prototype by using Posner et al.'s (1967) statistical-distortion procedure. The distortion level of the training exemplars was 6 bits/dot. Each prototype was then redefined by calculating the mean positions of each of the nine dots from the six training exemplars of the same category. The reason for this procedure was to eliminate any random bias that might have occurred in the process of generating the training exemplars. Thus, the prototype was the true (objective) central tendency of each category distribution. Three new distortions of each redefined prototype were then generated. One was the same level of distortion as the training exemplars (6.0 bits/dot), one was a low-level distortion (3.0 bits/dot), and one was a high-level distortion (7.7 bits/dot). The physical specifications for each of the 30 patterns are reported in Shin (1990). We used IBM PCs to display the stimuli and control the experiment.

Procedure

Classification condition (Group C). A standard learning-transfer paradigm was used. In the learning phase, subjects learned to classify the 18 old distortions into three categories labeled *A*, *B*, and *C*. On each trial, a randomly selected old distortion was presented on the computer monitor and feedback was provided after subjects responded. Following 270 learning trials, a transfer phase was conducted in which subjects made classification and recognition judgments for the original training exemplars (the old distortions), the prototypes, and the new distortions. Subjects first judged whether a given item was old (presented during the learning phase) or new (not presented during the learning phase) and then classified it. Classification feedback was provided only for old exemplars. There were three blocks of transfer trials. Each exemplar was presented once in each block, and order of presentation was randomized within blocks. There was a total of 90 transfer trials.

Similarity-judgment condition (Group S). The same 30 transfer items used in the classification experiment (3 prototypes, 18 old distortions, and 9 new distortions) were paired with each other and presented side by side on the monitor. Subjects rated their similarities by using a 9-point scale (1 = *most dissimilar*, 9 = *most similar*). Subjects were urged to use the full range of ratings in making their judgments. To familiarize the subjects with the items, the dot patterns were presented one by one before the main judgment session began. There was a total of 435 trials. The presentation sequence of the 435 pairs and positions of the items within each pair were randomized.

Classification/similarity-judgment condition (Group C/S). The combined procedure of Groups C and S was used. That is, after completing the classification learning and transfer phases, subjects rated the similarities among the items. There was a break of about 3 min before the similarity-rating session began.

Results

Classification and Recognition

The data of Groups C and C/S were combined in all of the analyses. Average proportion correct was approximately .48 during the first 10 trials of learning and increased to approximately .90 by the final 10 trials of learning. Thus, subjects learned to classify the old distortions very accurately by the end of the learning phase.

The probability with which each individual pattern was classified in each of the categories during the transfer phase is reported in Table 1. Averaged across categories, the old distortions were classified most accurately (84.0% correct), followed by the prototypes (68.4% correct) and the new distortions (all three distortion levels less than 60.0% correct). All differences among these three main item types were statistically significant, $t(49) = 6.58$, $p < .01$, for the old distortions versus the prototypes, and $t(49) = 5.75$, $p < .01$, for the prototypes versus the new distortions.[1] However, as can be seen in Table 1, performance on individual tokens of the three main item types was highly variable across the three categories. For example, in Category 2 there was no difference among the old distortions, the prototype, and the low- and medium-level new distortions. Also, performance on the Category 3 prototype was extremely poor relative to the Category 1 and 2 prototypes. The ability of the competing models to predict classification probabilities for the individual items and to account for this variability in performance is tested in the Classification Theoretical Analyses section.

The probability with which each item was judged as old during the transfer phase is reported in Table 2. The recognition data showed a pattern of results similar to the pattern for the classification data. Averaged across categories, the old distortions were judged as *old* (80.3%) more than the prototypes (57.1%) and the new distortions (all three distortion levels less than 40.0%). All differences among these three main item types were statistically significant, $t(49) = 8.26$, p

[1] There was no effect of the manipulation of level of distortion for the new transfer items. Undoubtedly, the reason is that there was only one example of each distortion level for each of the categories. A larger sample of new distortions is probably needed to observe an effect of this variable. The reason that only a few new distortions were generated for each category is that if too many patterns were used, the number of pairs of objects that would be required to conduct the similarity-judgment task would become unwieldy. The main goal of this research was to use the MDS-based exemplar and prototype models to quantitatively predict performance for individual patterns, not to replicate qualitative trends that are already well documented in the literature.

Table 1
Category Response Probabilities Observed and Predicted by the GCM and the Prototype Model in Experiment 1

	Observed			GCM			Prototype		
Item	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3
				Category 1					
P_1	.767	.113	.120	.735	.076	.189	.855	.034	.112
O_{11}	.867	.047	.087	.886	.053	.061	.585	.198	.217
O_{12}	.827	.040	.133	.870	.030	.100	.439	.171	.390
O_{13}	.940	.027	.033	.900	.022	.078	.871	.024	.105
O_{14}	.827	.073	.100	.790	.059	.151	.688	.059	.253
O_{15}	.933	.020	.047	.892	.025	.084	.850	.027	.123
O_{16}	.973	.000	.027	.919	.023	.058	.874	.037	.089
N_{1l}	.340	.367	.293	.463	.288	.249	.767	.126	.107
N_{1m}	.293	.300	.407	.339	.153	.508	.313	.279	.408
N_{1h}	.647	.180	.173	.376	.270	.354	.362	.194	.444
				Category 2					
P_2	.033	.833	.133	.028	.842	.130	.015	.935	.050
O_{21}	.053	.800	.147	.035	.858	.107	.050	.850	.100
O_{22}	.053	.787	.160	.028	.867	.105	.044	.831	.124
O_{23}	.007	.833	.160	.027	.849	.124	.036	.863	.101
O_{24}	.013	.893	.093	.019	.893	.088	.024	.883	.093
O_{25}	.013	.853	.133	.029	.896	.075	.058	.830	.112
O_{26}	.027	.833	.140	.025	.886	.090	.044	.809	.147
N_{2l}	.013	.833	.153	.032	.828	.140	.019	.926	.055
N_{2m}	.027	.847	.127	.039	.853	.108	.055	.852	.093
N_{2h}	.093	.580	.327	.084	.636	.280	.109	.537	.354
				Category 3					
P_3	.060	.487	.453	.142	.391	.467	.157	.273	.569
O_{31}	.080	.067	.853	.074	.108	.818	.138	.287	.575
O_{32}	.060	.087	.853	.064	.097	.839	.167	.174	.658
O_{33}	.093	.027	.880	.101	.071	.827	.356	.159	.485
O_{34}	.120	.033	.847	.153	.064	.783	.130	.056	.814
O_{35}	.147	.107	.747	.143	.089	.768	.147	.079	.774
O_{36}	.047	.387	.567	.060	.329	.610	.094	.392	.515
N_{3l}	.027	.546	.427	.134	.418	.448	.164	.317	.518
N_{3m}	.200	.233	.567	.234	.213	.553	.212	.139	.649
N_{3h}	.200	.413	.387	.163	.459	.378	.154	.303	.543

Note. In each category, Stimulus P_i is the prototype of Category i, O_{ij} is Old Exemplar j of Category i, N_{il} is a new low-distortion, N_{im} is a new medium-distortion, and N_{ih} is a new high-distortion item of Category i. GCM = generalized context model. C_1, C_2, and C_3 indicate subject responses of Category 1, 2, or 3.

$< .01$, for the old distortions versus the prototypes, and $t(49) = 9.86$, $p < .01$, for the prototypes versus the new distortions. As was the case for classification, however, recognition probabilities for individual tokens of the main item types varied considerably (see Table 2).

MDS Analysis

The averaged similarity matrices for Groups S and C/S were used as input to the INDSCAL model (Carroll & Wish, 1974) to derive an MDS solution for the dot patterns. The output of the INDSCAL model is an MDS solution that is common to both groups of subjects, together with sets of dimension weights that are specific to each group. The weights reflect the relative importance of each dimension in determining similarities among the items. Because there were no obvious "elbows" in the plot of stress versus number of dimensions (see Kruskal & Wish, 1978), we decided to use the four-, five-, and six-dimensional solutions in conjunction with the GCM and the prototype model to predict the classification and recognition data. (Six dimensions was the maximum dimensionality that the statistical package would support.) It is not surprising that slightly better fits to the data were achieved by using the six-dimensional solution, although the overall pattern of results was the same regardless of the dimensionality. In our subsequent theoretical analyses, we report the results of the model-based analyses that used the six-dimensional solution. The coordinates for the six-dimensional solution are reported in the Appendix, together with the dimension weights for each group of subjects. The Pearson product-moment correlations between the INDSCAL-derived distances and the similarity ratings were −.898 and −.928 in Groups S and C/S, respectively.

Although not necessary for the ensuing theoretical analysis, it is of interest to consider possible interpretations of the underlying dimensions of the MDS solution. Table 3 lists objective measures defined on the dot patterns that were found to correlate significantly with the patterns' coordinates

Table 2
Observed and Predicted Old Recognition Probabilities in Experiment 1

Item	Observed	GCM	Prototype
	Category 1		
P_1	.573	.541	.756
O_{11}	.847	.782	.629
O_{12}	.767	.782	.754
O_{13}	.860	.853	.750
O_{14}	.840	.796	.653
O_{15}	.893	.856	.740
O_{16}	.807	.802	.626
N_{1l}	.120	.224	.597
N_{1m}	.140	.115	.344
N_{1h}	.200	.221	.442
	Category 2		
P_2	.795	.719	.899
O_{21}	.647	.821	.776
O_{22}	.727	.834	.778
O_{23}	.813	.813	.808
O_{24}	.867	.829	.794
O_{25}	.807	.789	.646
O_{26}	.753	.795	.514
N_{2l}	.707	.751	.885
N_{2m}	.653	.626	.571
N_{2h}	.267	.275	.507
	Category 3		
P_3	.347	.327	.545
O_{31}	.893	.783	.388
O_{32}	.913	.783	.559
O_{33}	.673	.786	.418
O_{34}	.887	.790	.787
O_{35}	.680	.797	.711
O_{36}	.773	.808	.661
N_{3l}	.353	.359	.453
N_{3m}	.080	.042	.509
N_{3h}	.193	.180	.373

Note. In each category, Stimulus P_i is the prototype of Category i, O_{ij} is old exemplar j of Category i, N_{il} is a new low-distortion, N_{im} is a new medium-distortion, and N_{ih} is a new high-distortion item of Category i. GCM = generalized context model.

on each of the six dimensions. Of course, our proposed interpretations should be treated with caution. It is likely that numerous other objective measures could be discovered that would yield higher correlations with the derived dimensions.

Classification Theoretical Analyses

According to the GCM, the evidence favoring Category J, given presentation of Stimulus i, is found by summing the similarity of i to all exemplars of Category J and then multiplying by the response bias for Category J. This evidence is then divided by the sum of evidences for all categories to predict the conditional probability with which Stimulus i is classified in Category J, as follows:

$$P(R_J \mid S_i) = b_J \sum_{j \in C_J} s_{ij} \Big/ \left[\sum_K b_K \sum_{k \in C_K} s_{ik} \right], \tag{1}$$

where b_J $(0 \leq b_J \leq 1, \sum b_J = 1)$ is the Category J response bias and s_{ij} is the similarity between exemplars i and j.

The interexemplar similarities are computed from the MDS solution for the patterns. First, the psychological distance between each pair of exemplars is computed by using a weighted Euclidean metric, as follows:

$$d_{ij} = [\sum w_m (x_{im} - x_{jm})^2]^{1/2}, \tag{2}$$

where x_{im} is the psychological value (MDS coordinate) of exemplar i on dimension m, and w_m $(0 \leq w_m \leq 1, \sum w_m = 1)$ is the weight given to dimension m. The weights are free parameters estimated from the data and are interpreted as reflecting the attention given to each dimension in making classification decisions (see Nosofsky, 1984, 1986, for extensive discussions). This distance is transformed to a similarity measure using an exponential decay function, as follows:

$$s_{ij} = \exp(-cd_{ij}), \tag{3}$$

where c is a sensitivity parameter reflecting overall discriminability in the psychological space. There is a vast amount of previous research supporting these assumptions of a Euclidean metric for computing distances among integral-dimension stimuli, which we presume compose the present dot patterns, and an exponential decay function for relating similarity to distance in psychological space (e.g., Garner, 1974; Shepard, 1958, 1987).

This full version of the GCM has eight free parameters: one overall sensitivity parameter c in Equation 3; five freely varying attention weights (w_m) in Equation 2 (the six attention weights are constrained to sum to one); and two freely varying bias parameters in Equation 1 (the three bias parameters are constrained to sum to one).

The GCM was fitted to the complete matrix of classification data (Table 1) by using a maximum-likelihood criterion.[2] The maximum-likelihood parameters and summary fits for the full model are reported in Table 4. A scatterplot of the observed classification probabilities against the predicted probabilities is shown in Figure 1 (panel A). The predicted probabilities are also listed with the observed probabilities in Table 1.

The GCM predicted the classification data very well, accounting for 96.7% of the variance. The model accounted nicely for the main qualitative trends in the classification matrix. Averaged across individual items, the GCM predicted 84.2% correct for the old exemplars, 68.1% correct for the prototypes, and 54.2% correct for the new distortions (compared with observed values of 84.0, 68.4, and 54.7%, respectively). The model also predicted quite well the patterns of variability for individual tokens of the main item types. For example, it predicted much worse performance on the Cate-

[2] A computer search was used to find the parameters that maximized the log-likelihood function $\ln L = \sum \ln N_i! - \sum\sum \ln f_{ij}! + \sum\sum f_{ij} \ln p_{ij}$, where N_i is the frequency with which Stimulus i was presented, f_{ij} is the frequency with which Stimulus i was classified in Category j, and p_{ij} is the predicted probability with which Stimulus i was classified in Category j. This likelihood function assumes that the responses for each stimulus are multinomially distributed and that the individual distributions are independent. Because multiple observations were collected from the same subjects, the independence assumption is probably incorrect. Thus, the results of statistical tests involving the ln L statistic should be interpreted with caution.

Table 3
Correlations Between Objective Measures and Coordinates of the Patterns on the Dimensions of the Multidimensional Scaling Solution

Dimension	Name	Objective measure	r
1	Diagonal unbalance	Number of dots above the main diagonal	.785
2	Compactness	Size of the smallest rectangle encompassing all dots of a pattern	.627
3	Nearness of dots to the rectangle	Average distance of dots to the nearest edge of the rectangle encompassing all dots of a pattern	.532
4	Diagonal width	Distance between two intercepts of two linear functions with slope of −1 that encompass all dots in the middle	.499
5	Highest position	Vertical coordinate of the highest positioned dot of a pattern	.781
6	Horizontal dispersion	Average distance of dots to the central vertical line	.486

gory 3 prototype than on the Category 1 and 2 prototypes, as was observed in the data. Likewise, it predicted the excellent performance observed for the low and medium distortions of Category 2 and the relatively poor performance for the low and medium distortions of Categories 1 and 3. As a final example, the model predicted the noticeably poor performance on Old Exemplar 6 of Category 3. The main discrepancies in the matrix of 90 classification probabilities were that the model mispredicted some of the classification probabilities for the new distortions of Category 1, particularly the high-level new distortion.

Table 4
Maximum-Likelihood Parameters and Summary Fits of the Classification Data for the GCM, Its Restricted Versions, and the Prototype Model in Experiment 1

Parameter	Model			
	Full GCM	Version 1[a]	Version 2[b]	Prototype
c	2.642	2.642	1.787	2.718
w_1	.337	.336	.720[c]	.375
w_2	.095	.094	.344[c]	.214
w_3	.064	.065	.251[c]	.131
w_4	.309	.310	.275[c]	.000
w_5	.116	.116	.263[c]	.161
w_6	.079	.079	.188[c]	.119
b_1	.335	1/3[c]	.319	.314
b_2	.335	1/3[c]	.357	.363
b_3	.330	1/3[c]	.324	.324
Fit				
−ln L	225.686	225.771	256.112	527.404
SSE	.303	.305	.433	1.576
% variance accounted for	96.746	96.724	95.356	83.093

Note. GCM = generalized context model; c = general sensitivity parameter; w_m = attention weight given to dimension m; b_j = bias for making Category response R_j; −ln *L* = negative value of log-likelihood; *SSE* = sum of squared deviations between observed and predicted classification probabilities.
[a] The bias parameters are held fixed at 1/3. [b] The weights are set at the INDSCAL weights. [c] These values were fixed a priori.

To evaluate the importance of the attention weights and response bias parameters in achieving these fits, several restricted versions of the GCM were tested. In Version 1, all the response bias parameters were set at $b_j = 1/3$ (i.e., a bias-free model). In Version 2, the attention weights were set at the values obtained when deriving the INDSCAL solution from the similarity-judgment data (Group S). These weights are the ones that would be expected if subjects devoted the same attention to the individual dimensions for classification as for making similarity judgments. Finally, in Version 3, the bias parameters were all set at 1/3 and the attention weights were held fixed at the INDSCAL weights. This third version is a single-parameter model, with only the sensitivity parameter free to vary.

As shown in Table 4, the bias-free model (Version 1) yielded predictions that were essentially identical to those of the full GCM and, indeed, did not fit significantly worse than the full model, $\chi^2(2, N = 4{,}500) = 0.17$, $p > .10$.[3] Thus, the bias parameters are not needed for fitting the data. We would expect a greater influence of differential biases in situations in which category frequencies or payoffs varied, and so forth. Version 2 of the model, with the INDSCAL similarity-judgment weights, fit significantly worse than the full GCM, $\chi^2(5, N = 4{,}500) = 60.85$, $p < .001$. Nevertheless, this restricted version was still able to account for 95.4% of the variance in the classification data, so the evidence for differential selective attention to the dimensions is not dramatic. We would expect a greater influence of the attention weights in situations in which the dimensions are highly separable (Garner, 1974;

[3] A restricted version of a full model arises when some of the parameters in the model are constrained on a priori grounds. To test if the restricted model fits significantly worse than the full model, one uses the method of likelihood-ratio testing (e.g., see Wickens, 1982). Let ln L_F and ln L_R denote the log-likelihoods for the full and restricted models, respectively. Assuming the restricted model is correct, the statistic 2 (ln L_F − ln L_R) develops a chi-square distribution as sample size gets large, with degrees of freedom equal to the number of constrained parameters. If the observed value of chi-square exceeds the critical value, then the restricted model fits significantly worse than the full one.

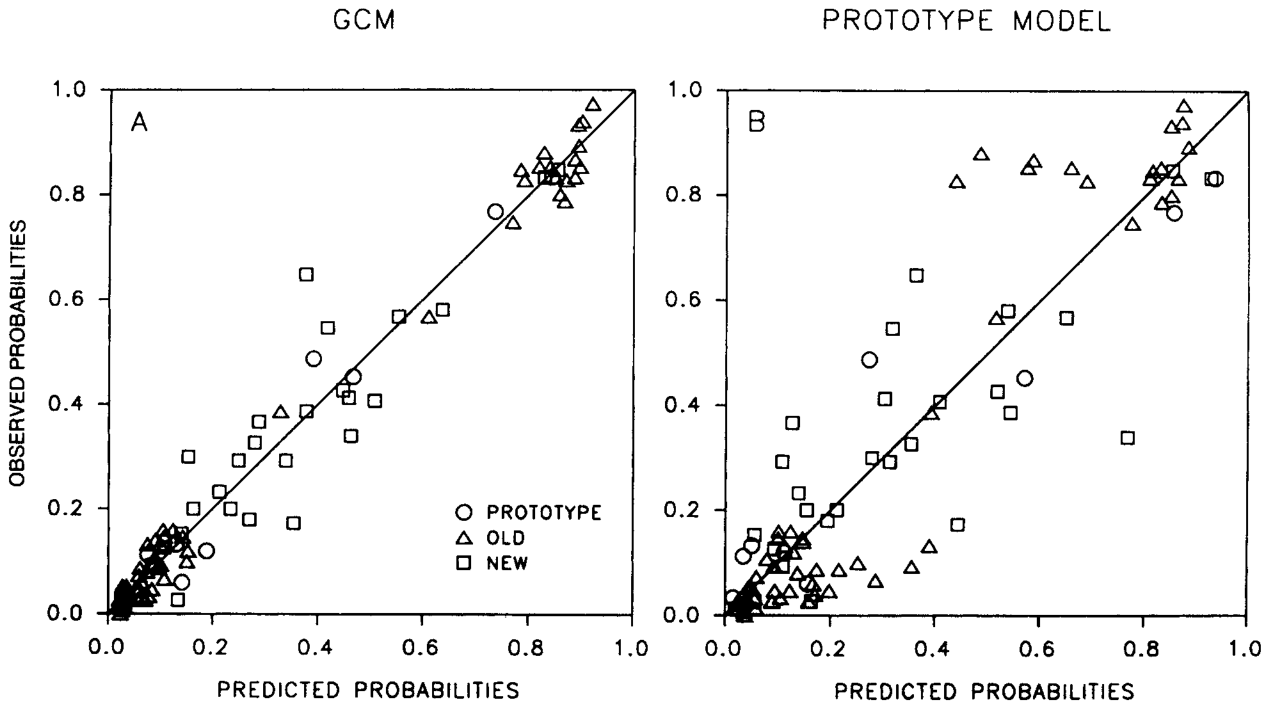

Figure 1. Scatterplot of observed against predicted classification probabilities for Experiment 1. (Panel A is the generalized context model [GCM]; panel B is the prototype model.)

Shepard, 1964) and only a subset of the dimensions is relevant for performing the classification (e.g., Nosofsky, 1989).

Finally, the single-parameter version of the GCM, with only sensitivity c free to vary, was able to account for 95.1% of the variance in the 90-cell classification matrix. The precision with which this single-parameter version of the GCM, when used in conjunction with the MDS solution, was able to predict the classification data in this dot-pattern domain seems truly remarkable.

Prototype Model

Following Reed (1972) and Nosofsky (1987), a prototype model was formalized within the MDS framework. The prototype model was the same as the GCM, except rather than computing the summed similarity of an item to the category exemplars, one computes the similarity between an item and the (psychological) category prototype. The psychological prototype of a category was defined as the central tendency or centroid of the distribution of category exemplars in the multidimensionally scaled space. Note that whereas the objective prototype is an actual stimulus pattern, the psychological prototype is a theoretical construct. Furthermore, the objective prototype and psychological prototype will not be located at precisely the same point in the multidimensional psychological space.

The formalization of prototype as the category central tendency is in keeping with much previous theoretical work aimed at testing models of classification (e.g., Ashby & Gott, 1988; Franks & Bransford, 1971; Hayes-Roth & Hayes-Roth, 1977; Medin & Schaffer, 1978; Nosofsky, 1987; Reed, 1972; Reitman & Bower, 1973; Rosch, Simpson, & Miller, 1976). Furthermore, this definition of prototype is very much in the spirit of the dot-pattern paradigm because, in the limit, as category size gets large, the objective prototype is indeed the (objective) central tendency of all the old and new distortions.

Formally, according to the prototype model, the probability that Stimulus i is classified in Category J is given by the following equation:

$$P(R_J \mid S_i) = b_J s_{iPJ} / \sum_K b_K s_{iPK}, \tag{4}$$

where s_{iPJ} denotes the similarity between Stimulus i and the psychological prototype of Category J. This similarity is computed as in the GCM. The distance between Stimulus i and Prototype J is given by the following equation

$$d_{iJ} = [\sum w_m (x_{im} - p_{Jm})^2]^{1/2}, \tag{5}$$

where p_{Jm} is the psychological value of Prototype J on dimension m. This distance is then transformed to a similarity measure by using an exponential decay function (Equation 3). The free parameters in the prototype model are the same as in the GCM, and they function analogously.

The prototype model was fitted to the classification data by using a maximum-likelihood criterion. The maximum-likelihood parameters and summary fits for the model are reported in Table 4. A scatterplot of the observed against predicted classification probabilities for the prototype model is presented in Figure 1 (panel B). The predicted probabilities are also reported with the observed probabilities in Table 1.

As is evident from inspection of Table 4 and Figure 1, the prototype model yielded a much worse fit to the classification data than did the GCM (e.g., $-\ln L = 527.4$ for the prototype model vs. $-\ln L = 225.7$ for the GCM). In general, the prototype model overpredicted correct classification for all the prototypes and underpredicted correct classifications for the old exemplars of Categories 1 and 3.

Combined Model

The finding that the GCM outperformed the prototype model does not imply that a prototype abstraction process did not take place. A salient possibility is that a prototype was abstracted and that classification decisions were based on both stored exemplars and the abstracted prototype. To test this idea, one can formulate mixture or combined models and then compare these models and the pure exemplar model on their ability to predict the classification data. One such mixture model was proposed by Busemeyer et al. (1984) and Homa et al. (1991). Basically, according to this mixture model, the probability that some item i, which was a member, for instance, of Category J, would be classified in Category J was given by the following equation:

$$P(R_J \mid S_i) = u_J + (1 - u_J)P_E(J \mid i), \tag{6}$$

where u_J is the probability that the Category J prototype is used to correctly classify members of its own category, and $P_E(J|i)$ is the probability that Item i is classified into Category J on the basis of the exemplar-generalization process. A conceptual problem with this mixture model, however, is that the same value u_J applies to all items, regardless of their similarity to the abstracted prototype. A more reasonable model would allow the probability of prototype-based classification to depend on an item's similarity to the prototype. Fortunately, our scaling approach allows one to formalize a variety of such models with a minimum of additional parameter estimation.

The most straightforward of the models that we tested was a combined model that views the abstracted prototype as simply another exemplar with a special status. Extending the GCM response rule, the probability of classifying Stimulus i into Category J is given by the following equation:

$$P(R_J \mid S_i) = b_J[y \cdot s_{iPJ} + \sum_{j \in C_J} s_{ij}]/[\sum_{K} b_K \cdot (y \cdot s_{iPK} + \sum_{k \in C_K} s_{ik})], \tag{7}$$

where the similarity parameters are computed as before, and y is a relative weighting term accorded to the prototypes. If the value of y is zero, then this model reduces to the pure GCM, whereas as y goes to infinity, classification is based purely on the prototypes. Intermediate values of y reflect combinations of exemplar-based and prototype-based classification. Note that the model captures the idea that the effectiveness of prototype-based classification should depend on an item's similarity to its category prototype.

Because the Equation 7 combined model generalizes the GCM, it must fit the classification data at least as well as the GCM. The critical question is whether the addition of the prototype-use parameter y leads to major improvements in fit over the pure exemplar model. In Experiment 1, the answer to this question is clear-cut: The maximum-likelihood estimate of y was zero, so the combined model led to no improvement over the pure exemplar model. A variety of alternative combined models and mixture models were also fitted to the classification data (see Shin, 1990), but none provided any evidence for a prototype-abstraction process that operated above and beyond pure exemplar-based generalization.

Recognition Theoretical Analyses

Generalized Context Model

Nosofsky (1988a, 1991b) demonstrated that old–new recognition data obtained in the context of classification learning situations can be modeled within the GCM framework. The critical assumption is that recognition judgments are based on the summed similarity of an item to all the exemplars of all the categories:

$$F_i = \sum\sum s_{ik}. \tag{8}$$

This summed similarity gives a measure of the overall familiarity for Item i (F_i), with higher familiarity values leading to higher recognition probabilities (cf. Gillund & Shiffrin, 1984; Hintzman, 1988).

Following Clark (1988), we used the following decision rule to predict the probability that Stimulus i is judged as old:

$$P(\text{old} \mid i) = F_i/[F_i + k], \tag{9}$$

where k is a free parameter to be estimated. The k parameter can be interpreted as a criterion for making *old* recognition judgments. Small ks reflect a lenient criterion and large ks a strict one.

The GCM with the Equation 9 decision rule was fitted to the old–new recognition data by using the same MDS solution as was used for fitting the classification data. Again, maximum likelihood was used as the measure of fit. The free parameters to be estimated were the sensitivity parameter c, attention weight parameters w_m, and criterion parameter k. The maximum-likelihood parameters and summary fits are reported in Table 5, and a scatterplot of the observed against predicted recognition probabilities is provided in Figure 2 (panel A). The predicted recognition probabilities are also reported along with the observed probabilities in Table 2.

The GCM predicts the recognition data fairly well, accounting for 94.6% of the variance. Averaged across individual items, the GCM predicts 80.6% *old* judgments for the old exemplars, 52.9% for the prototypes, and 31.0% for the new distortions (compared with observed values of 80.3, 57.2, and 30.1%, respectively). Concerning patterns of variability for individual tokens of the main item types, the model predicts fairly well the low recognition probability for Prototype 3, the moderate recognition probability for Prototype 1, and the high recognition probability for Prototype 2. It also predicts nicely the rather high recognition probabilities for the low and medium distortions of Category 2. Only one prediction deviated by more than 15% from the observed data (Old 1 of Category 2).

Nevertheless, an apparent shortcoming of the GCM revealed by Figure 2 is that it fails to account for the variability

Table 5
Maximum-Likelihood Parameters and Summary Fits of the Old–New Recognition Data in Experiment 1

Parameter	Model	
	GCM	Prototype
c	4.905	3.401
w_1	.006	.499
w_2	.084	.133
w_3	.102	.245
w_4	.392	.000
w_5	.218	.033
w_6	.198	.090
k	.280	.078
Fit		
−ln *L*	142.299	587.607
SSE	.285	2.921
% variance accounted for	94.623	44.935

Note. GCM = generalized context model; c = general sensitivity parameter; w_m = attention weight given to dimension m; k = criterion of judging *old*; −ln *L* = negative value of log-likelihood; *SSE* = sum of squared deviations between observed and predicted classification probabilities.

in recognition probabilities observed for just the old exemplars. Furthermore, the values of the sensitivity and attention weight parameters differed sharply for classification and recognition (compare the parameter estimates in Tables 4 and 5). Theoretical analyses reported by Shin (1990) in which the GCM was fitted to the classification and recognition data simultaneously indicated that the parameter estimates were significantly different across the two tasks. It is clear that more work is needed to understand why there is this lack of parameter invariance (but see Nosofsky, 1991b, for preliminary ideas along these lines).

Prototype Model

The prototype model for predicting old–new recognition was also formalized within the framework of the MDS approach. We assumed that overall familiarity (F_i in Equation 8) was given by the summed similarity of an item to all of the category prototypes. In all other respects, the prototype model was the same as the GCM. The maximum-likelihood parameters and summary fits for the prototype model are reported in Table 5, and a scatterplot of the observed against predicted recognition probabilities is provided in Figure 2 (panel B). The predicted probabilities are also reported with the observed probabilities in Table 2. The prototype model accounted for only 44.9% of the variance in the recognition data, far worse than the GCM (94.6%). In general, the model overpredicted *old* recognition probabilities for the prototypes and new distortions, and underpredicted *old* recognition probabilities for the old exemplars.

Analyses of the Similarity Data

A secondary purpose of this research was to study the role of category learning experience on subjects' similarity judg-

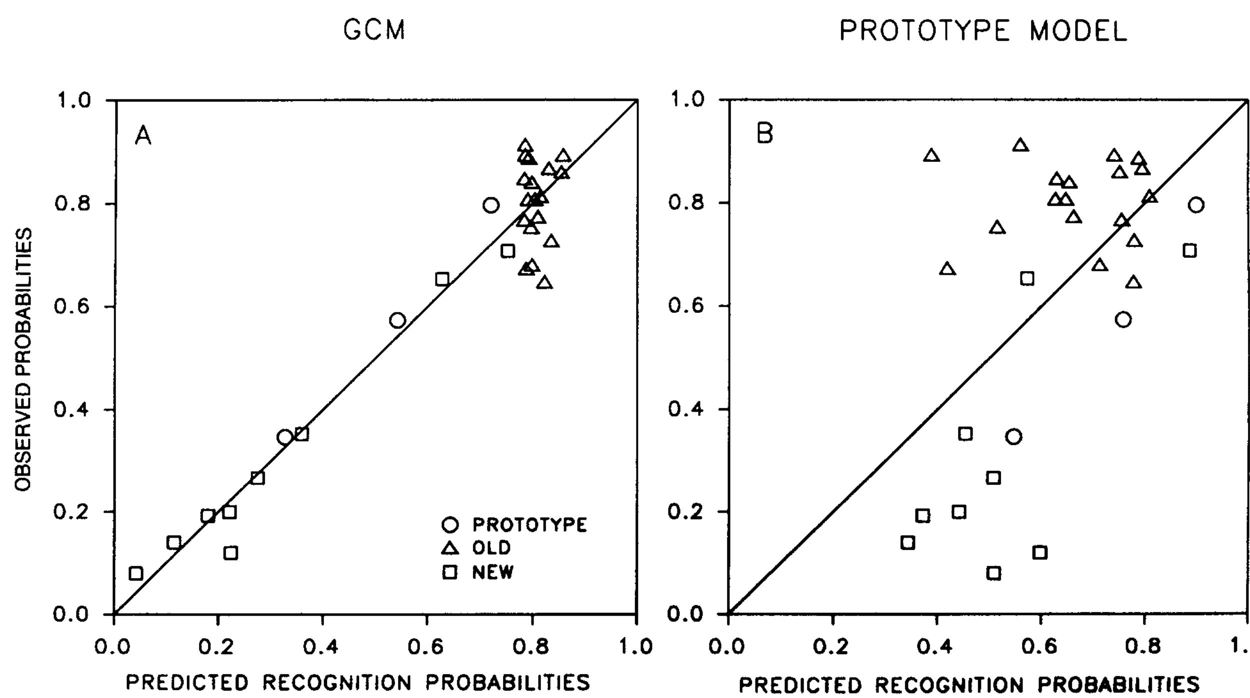

Figure 2. Scatterplot of observed against predicted recognition probabilities for Experiment 1. (Panel A is the generalized context model [GCM]; panel B is the prototype model.)

ments and to test whether changes in similarity judgments could be modeled in terms of an attention-weighting process (e.g., Nosofsky, 1986; Tversky, 1977). The average within- and between-category similarity judgments for Groups S and C/S are summarized in Table 6. An analysis of variance (ANOVA) was performed on these data, with learning experience as a between-subjects factor and within- versus between-category similarity as a within-subjects factor. The average within-category similarities (5.26) are larger than the average between-category similarities (4.17), $F(1, 38) = 300.6$, $MS_e = 0.080$, $p < .01$. An unexpected result is that learning experience increased both within- and between-category similarity judgments, $F(1, 38) = 5.02$, $MS_e = 1.047$, $p < .05$. (It was expected that whereas within-category similarities would increase, between-category similarities would decrease.) However, the interaction of the two main factors was statistically significant, $F(1, 38) = 4.93$, $MS_e = 0.080$, $p < .05$, reflecting that the increase in within-category similarities was larger than that of between-category similarities. Finally, the INDSCAL weights for Groups S and C/S were virtually identical (see the Appendix). Thus, our hypothesis that changes in similarity judgments as a function of category learning could be modeled in terms of changes in the attention weights was not supported. We do not have a good explanation of why between-category similarity judgments increased with learning experience in this paradigm. The same pattern of results just reported, however, was observed in all of our subsequent experiments.

Discussion

The focus of Experiment 1 was to quantitatively contrast the GCM with the prototype model in predicting classification and recognition performance in the dot-pattern category learning paradigm. Both models were conceptualized within the framework of an MDS approach. The most important demonstration was that the GCM, when used in conjunction with the MDS solution for the dot patterns, was able to achieve good quantitative predictions of both classification and recognition performance. Not only did the model accurately predict overall levels of performance for the main item types (i.e., the old exemplars, prototypes, and new distortions), it accurately predicted classification probabilities and patterns of variability for individual tokens of these main item types. By contrast, an MDS-based prototype model fared very poorly relative to the GCM. To evaluate the relative contribution of exemplar information and prototype information, a combined model was also formalized and fitted to the classification data. It is surprising that no evidence was found for the use of any prototype information.

Table 6
Average Within- and Between-Category Similarities Due to Category Learning Experience in Experiment 1

	Category learning experience		
Category	No (Group S)	Yes (Group C/S)	*M*
Within-category similarity	4.936	5.589	5.263
Between-category similarity	3.980	4.352	4.166
M	4.458	4.971	4.715

The design of Experiment 1 has limitations, however, and we would not conclude that prototype abstraction never occurs in the learning of ill-defined categories of random dot patterns. We used a relatively small number of training exemplars per category (six). Another limitation is that the classification and recognition transfer data were obtained immediately after the completion of the training phase. Homa and his colleagues (Homa, 1984; Homa et al., 1973, 1981; Homa & Vosburgh, 1976) have argued that with increased category size and time of delay, prototype information becomes dominant in the category representation. Therefore, in Experiment 2, we manipulated category size and time of delay of the transfer phase in an effort to produce conditions that might promote prototype abstraction.

Experiment 2

A robust finding in the dot-pattern classification learning paradigm is the category-size effect. As the number of old distortions defining a category increases in the learning phase, classification accuracy for the prototypes and new distortions increases in the transfer phase (e.g., Breen & Schvaneveldt, 1986; Homa & Chambliss, 1975; Homa et al., 1973; Homa & Cultice, 1984).

Since Posner and Keele (1970) demonstrated differential forgetting rates of the old exemplars and the prototypes, effects of delayed transfer tests have been obtained in many studies (e.g., Goldman & Homa, 1977; Homa et al., 1973, 1981; Homa & Vosburgh, 1976; Strange, Keeney, Kessel, & Jenkins, 1970). Whereas the old exemplars suffer substantially from forgetting over delay, the prototypes and new distortions often do not show significant forgetting.

On the basis of these previous findings, it seems reasonable that a prototype enhancement effect is likely to occur as category size increases and as the transfer test is delayed. Our aim in Experiment 2 was to produce experimental conditions that would yield prototype enhancement effects because such effects might prove extremely challenging to the quantitative predictions of a pure exemplar-based generalization model.

One previous experiment that showed a large prototype enhancement effect was reported by Homa and Cultice (1984). These investigators manipulated category size by using 3, 6, and 9 training exemplars for each of three categories. They also manipulated learning conditions by using training exemplars of the same distortion level or of mixed distortion levels. In the condition with mixed distortion levels, subjects were trained on low-, medium-, and high-level distortions simultaneously. In this mixed-distortion condition, Homa and Cultice (1984, Experiment 1, feedback condition) observed a prototype-enhancement effect for all three category sizes, wherein the prototypes were classified with higher accuracy than the average of the old exemplars. A particularly large prototype-enhancement effect was observed for the Size 9 category.

On the basis of these previous findings, we conducted an experiment similar to the one of Homa and Cultice (1984) by varying category size and by using exemplars of mixed distortion levels. We also manipulated time of transfer by testing subjects in both an immediate and a delayed condition. By introducing delay, we hoped to increase the magnitude of the prototype-enhancement effect.

In Homa and Cultice's (1984) mixed-distortions condition, there was a total of 80 transfer items. We reduced the number of transfer items to 30 (3 prototypes, 18 old exemplars, and 9 new distortions) because of a practical consideration: It was impractical to obtain scaling solutions for the exemplars if too many were used in the experiment. For example, to obtain scaling solutions for 30 exemplars requires similarity judgments for 435 pairs, but the number increases to 3,160 in the case of 80 exemplars.

Method

Subjects

A total of 90 undergraduates from Indiana University participated in Experiment 2 as part of a course requirement. Of the 90 subjects, 30 were assigned to the immediate transfer test condition (Group I), 30 to the delayed transfer test condition (Group D), and 30 to the similarity-judgment-only condition (Group S). The 30 subjects in the delayed condition also performed similarity judgments after the transfer phase. Two subjects in the delayed condition were replaced because their learning performances were almost at chance level. All subjects were tested individually.

Stimuli and Apparatus

Random dot patterns were generated by using a procedure similar to that used in Experiment 1. First, three prototypes were created. One was randomly assigned to a Size 3 category, the second to a Size 6 category, and the third to a Size 9 category. Three, six, and nine old distortions were then generated from the corresponding prototypes. In each category, an equal number of exemplars was generated at each of three distortion levels. For example, the Size 9 category had three low-level distortions (4.0 bits/dot), three medium-level distortions (6.0 bits/dot), and three high-level distortions (7.7 bits/dot) as training exemplars. Because the distortion levels within a category were not the same, we decided not to redefine the prototypes as in Experiment 1. (Had the prototype been recomputed, the old exemplars might not have retained their intended distortion levels.) Three new distortions were also generated from each prototype. One was a low-level distortion, the second was a medium-level distortion, and the third was a high-level distortion. We used IBM PCs to display the stimuli and control the experiment.

Procedure

Immediate condition. The general procedure was the same as for Group C in Experiment 1. The only difference was that during the learning phase the 18 old exemplars were presented by blocks. Each exemplar was presented once in each block, and order of presentation was randomized. Following Homa and Cultice (1984), there were eight blocks of learning trials. Subjects were told not to expect an equal number of items in each category. Immediately after the learning phase, the transfer test including all 30 items was conducted. Both classification and recognition data were collected, although we report only the classification data.

Delayed condition. The transfer test was performed after a delay of 1 week. Subjects were instructed to maintain consistency in assigning names to the items if they forgot which names (A, B, or C) corresponded to each category. In scoring the results, the assignment of names to categories that maximized performance was assumed. Subjects judged the similarities among the exemplars after the delayed transfer test.

Similarity-judgment-only condition. The procedure was the same as for Group S in Experiment 1.

Results

In both the immediate and delayed conditions, average proportion correct was approximately .42 in the first block of learning and increased to approximately .82 in the final block of learning. The final learning performance is somewhat worse than the level reported by Homa and Cultice (1984), perhaps because we used random dot patterns, whereas they used random polygons (dot patterns with connecting lines).

The probability with which each individual pattern was classified in each category during the transfer phase is reported in Table 7. A summary of some of the main trends is presented in Table 8. (Note, however, that as in Experiment 1, levels of performance varied considerably for individual tokens of the main item types.) In the immediate condition, averaged over categories, the old low distortions were classified the best (.891), but the prototypes were classified as well (.822) as both the old medium distortions (.819) and old high distortions (.796). The new distortions were classified worse than the other patterns (.641).

Overall performance was worse in the delayed condition for all the item types, and the pattern of performance was the same as in the immediate condition. In particular, we did not observe differential forgetting for the prototypes compared with the old exemplars. Averaged across categories, performance on the old exemplars declined by 12.6% and performance on the prototypes declined by 15.5%. A two-way ANOVA with conditions (immediate vs. delayed) and patterns (old vs. prototypes) as factors yielded a main effect of conditions, $F(1, 58) = 11.58$, $MS_E = .052$, $p < .01$, but no interaction $F(1, 58) = 0.55$, $MS_E = .011$, reflecting the nondifferential forgetting.

A category-size effect was observed for the new distortions and prototypes in both the immediate and delayed conditions. For example, in the immediate condition, classification accuracy for the new distortions increased from .518 in the Size 3 category to .793 in the Size 9 category. There was little if any category-size effect for the old exemplars (see Table 8).

The main result we wish to highlight is that the conditions of Experiment 2 produced at least a partial prototype-enhancement effect (albeit not as large as the one reported by Homa & Cultice, 1984). In particular, in the Size 9 category of the immediate condition, the prototype was classified significantly more accurately (.911) than the old high distortions (.789), $t(29) = 2.385$, $p < .05$, and at least as well as the old medium distortions (.830), $t(29) = 1.621$, $p \approx .10$. An advantage of the prototype over the old high distortions was also observed for the Size 9 category in the delayed condition. The question of whether the MDS-based exemplar model can predict these prototype-enhancement effects, as well as the

Table 7
Category Response Probabilities Observed and Predicted by the GCM and the Prototype Model in the Immediate and Delayed Conditions of Experiment 2

	Immediate condition									Delayed condition								
	Observed			GCM			Prototype			Observed			GCM			Prototype		
Item	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
Category 1																		
P_1	.822	.056	.122	.776	.087	.137	.779	.090	.131	.667	.200	.133	.679	.202	.119	.693	.190	.117
O_l	.833	.056	.111	.821	.081	.098	.762	.112	.126	.656	.211	.133	.736	.176	.088	.674	.217	.110
O_m	.833	.056	.111	.822	.053	.125	.730	.087	.183	.722	.222	.056	.735	.149	.117	.660	.186	.154
O_h	.733	.122	.144	.688	.168	.144	.513	.216	.271	.611	.189	.200	.668	.208	.124	.562	.248	.190
N_l	.700	.133	.167	.700	.115	.185	.722	.114	.165	.644	.256	.100	.597	.251	.153	.629	.230	.141
N_m	.411	.333	.256	.415	.284	.301	.488	.267	.246	.333	.456	.211	.389	.404	.207	.449	.374	.177
N_h	.444	.233	.322	.610	.150	.240	.607	.607	.146	.511	.356	.133	.544	.275	.181	.571	.247	.182
Category 2																		
P_2	.056	.733	.211	.049	.800	.152	.036	.863	.101	.133	.611	.256	.195	.617	.188	.139	.731	.130
O_{l1}	.022	.956	.022	.039	.853	.107	.051	.812	.137	.078	.800	.122	.146	.732	.122	.168	.674	.158
O_{l2}	.056	.778	.167	.035	.832	.133	.026	.879	.095	.089	.756	.156	.139	.689	.172	.109	.759	.132
O_{m1}	.056	.667	.278	.048	.769	.183	.093	.646	.261	.122	.556	.322	.148	.664	.188	.221	.551	.227
O_{m2}	.044	.922	.033	.030	.890	.081	.040	.829	.131	.056	.889	.056	.100	.804	.096	.127	.720	.153
O_{h1}	.044	.811	.144	.067	.796	.137	.117	.641	.242	.144	.711	.144	.213	.613	.174	.280	.506	.215
O_{h2}	.056	.867	.078	.026	.905	.069	.026	.906	.068	.144	.778	.078	.122	.785	.093	.112	.787	.100
N_l	.089	.778	.133	.074	.789	.137	.058	.840	.102	.222	.589	.189	.212	.632	.156	.181	.700	.119
N_m	.112	.467	.411	.088	.388	.524	.120	.372	.508	.433	.278	.289	.266	.332	.402	.249	.369	.382
N_h	.033	.589	.378	.043	.542	.415	.064	.649	.287	.222	.567	.211	.182	.523	.295	.180	.554	.266
Category 3																		
P_3	.033	.056	.911	.029	.082	.888	.024	.060	.917	.156	.122	.722	.084	.144	.772	.077	.130	.793
O_{l1}	.044	.033	.922	.028	.077	.895	.023	.056	.921	.122	.122	.756	.083	.136	.781	.076	.124	.800
O_{l2}	.000	.044	.956	.041	.076	.883	.036	.061	.903	.056	.144	.800	.106	.139	.754	.097	.129	.775
O_{l3}	.056	.044	.900	.036	.071	.893	.035	.057	.907	.133	.133	.733	.091	.144	.765	.096	.132	.773
O_{m1}	.022	.167	.811	.037	.095	.868	.032	.083	.885	.100	.167	.733	.093	.172	.735	.093	.169	.738
O_{m2}	.033	.244	.722	.045	.148	.807	.067	.178	.755	.133	.189	.678	.149	.190	.661	.168	.237	.596
O_{m3}	.000	.044	.956	.044	.067	.888	.056	.082	.862	.133	.122	.744	.111	.142	.747	.128	.165	.708
O_{h1}	.189	.000	.811	.112	.105	.783	.180	.150	.670	.300	.178	.522	.185	.243	.573	.241	.271	.487
O_{h2}	.000	.300	.700	.030	.341	.629	.059	.508	.434	.122	.344	.533	.144	.399	.456	.165	.437	.398
O_{h3}	.089	.056	.856	.060	.059	.880	.071	.061	.868	.167	.056	.778	.127	.135	.738	.131	.139	.731
N_l	.056	.178	.767	.036	.093	.871	.024	.053	.923	.111	.133	.756	.096	.149	.756	.077	.120	.803
N_m	.022	.122	.856	.034	.103	.863	.026	.076	.898	.022	.133	.844	.092	.153	.755	.082	.149	.769
N_h	.011	.233	.756	.049	.260	.691	.047	.185	.769	.211	.222	.567	.124	.287	.589	.123	.243	.634

Note. P_i denotes prototype of Category *i*; O_l, O_m, and O_h denote low-, medium-, and high-level old distortions, respectively; N_l, N_m, and N_h denote low-, medium-, and high-level new distortions, respectively. GCM = generalized content model. C_1, C_2, and C_3 refer to subject choices of Category 1, 2, or 3.

Table 8
Performance Summaries (Proportion Correct) in Experiment 2

Distortion type	Immediate condition Observed	Immediate condition Predicted	Delayed condition Observed	Delayed condition Predicted
New				
C1	.518	.575	.496	.510
C2	.611	.573	.478	.496
C3	.793	.808	.722	.700
Average new	.641	.652	.565	.569
Low old				
C1	.833	.821	.656	.736
C2	.867	.843	.778	.711
C3	.926	.890	.763	.767
Medium old				
C1	.833	.822	.722	.735
C2	.795	.830	.723	.734
C3	.830	.854	.718	.714
High old				
C1	.733	.688	.611	.668
C2	.839	.851	.745	.699
C3	.789	.764	.611	.589
Average old				
Low	.891	.863	.750	.743
Medium	.819	.841	.720	.724
High	.796	.780	.656	.639
C1	.800	.777	.663	.713
C2	.817	.841	.748	.715
C3	.848	.836	.697	.690
Prototype				
C1	.822	.776	.667	.679
C2	.733	.800	.611	.617
C3	.911	.888	.722	.772
Average prototype	.822	.821	.667	.689

Note. C1 = Size 3 category; C2 = Size 6 category; C3 = Size 9 category.

other main effects noted earlier, is addressed in the Theoretical Analyses section.

As was the case in Experiment 1, the similarity-judgment data from Groups S and D were used as input to the INDSCAL model. Again, we report the model-based analyses that used the six-dimensional solution, but the pattern of results of the ensuing theoretical analyses did not depend on the dimensionality of the solution. The MDS solution and dimension weights for each group are reported in the Appendix. The correlations between the INDSCAL-derived distances and the similarity ratings were −.937 and −.948 in Groups S and D, respectively.

Theoretical Analyses

Generalized Context Model

The GCM-predicted classification probabilities for the immediate and delayed conditions are shown along with the observed probabilities in Table 7, and the predictions of the main trends are presented with the observed data in Table 8. The maximum-likelihood parameters and summary fits are reported in Table 9, and scatterplots of observed against predicted probabilities are shown in Figure 3.

The GCM predicted the classification data fairly well, accounting for 97.7% and 95.3% of the variance in the immediate and delayed conditions, respectively. As shown in Table 8, it achieved excellent predictions of the main trends in both the immediate and delayed conditions, accounting accurately for overall performance levels on the low-, medium-, and high-level old distortions, the new distortions, and the prototypes. It also predicted the large category-size effect for the new distortions and the much smaller (or nonexistent) one for the old distortions. The model also accounted for the forgetting observed for all patterns due to the delay manipulation.

Patterns of variability for individual tokens of the main item types were also well predicted. The main result we wish to highlight is that the model did indeed predict the prototype enhancement effect observed for the Size 9 category (see Tables 7 and 8), wherein the prototype was classified more accurately than the old high distortions and at least as well as the old medium distortions.

Inspection of the maximum-likelihood parameters in Table 9 reveals that, according to the model, two main changes occurred when the transfer test was delayed. First, there was a decrease in the value of the sensitivity parameter c, suggesting that as the transfer test was delayed, discriminability among the items stored in memory decreased. Such changes in similarities among exemplars are one of the major ways that the context model explains learning and forgetting phenomena (e.g., Medin & Schaffer, 1978; Nosofsky, 1987; Nosofsky, Kruschke, & McKinley, 1992). Second, in the immediate condition, there was roughly equal response bias for the Size 3, Size 6, and Size 9 categories (perhaps slightly greater bias for Size 3 and Size 6 than Size 9), but in the delayed condition there was an ordering of biases that was inverse to category size. Note that according to the exemplar model, with nondifferential response bias, subjects would tend to more often classify items into the larger size categories. One reason is that, all other things being equal, summed similarity to exemplars is greater the larger the size of the category. The differential response bias observed in the delayed condition may have been the result of subjects trying to more nearly equalize their use of the alternative category labels.

Prototype Model

The predicted classification probabilities for the prototype model are presented with the observed probabilities in Table 7. The maximum-likelihood parameters and summary fits are reported in Table 9.

The prototype model accounted for more than 90% of the variance in both the immediate and delayed conditions but, compared with the GCM, the predictions of the prototype model were poor. For example, the sum of squared deviations between predicted and observed probabilities for the prototype model was approximately double that of the GCM in both the immediate and delayed conditions (see Table 9). The prototype model had a tendency to slightly overestimate performance on the prototypes and to underestimate performance on many of the old exemplars.

Table 9
Maximum-Likelihood Parameters and Summary Fits of the Classification Data for the GCM, the Prototype Model, and the Combined Model in Experiment 2

	Immediate condition			Delayed condition		
Parameter	GCM	Prototype	Combined	GCM	Prototype	Combined
c	2.181	2.207	2.181	1.617	1.566	1.753
w_1	.348	.351	.348	.445	.441	.391
w_2	.361	.355	.361	.186	.235	.176
w_3	.026	.000	.026	.000	.000	.000
w_4	.063	.145	.003	.152	.167	.189
w_5	.084	.061	.084	.000	.028	.000
w_6	.118	.088	.118	.217	.130	.243
b_1	.366	.213	.366	.478	.300	.126
b_2	.363	.373	.363	.335	.375	.567
b_3	.272	.414	.272	.187	.325	.307
y_1			.000			17.576
y_2			.000			.740
y_3			.000			1.110
Fit						
$-\ln L$	169.723	226.166	169.723	181.313	203.240	178.343
SSE	.233	.585	.233	.274	.437	.255
% variance accounted for	97.674	94.157	97.674	95.334	92.542	95.654

Note. GCM = generalized context model; c = general sensitivity parameter; w_m = attention weight given to dimension m; b_j = bias for making Category response R_j; $-\ln L$ = negative value of log-likelihood; *SSE* = sum of squared deviations between observed and predicted classification probabilities.

Combined Model

The combined model was elaborated to allow different degrees of prototype abstraction for the different size categories. In particular, three separate values of the prototype-contribution parameter (y in Equation 7) were allowed, one for each category size. The reason for elaborating the model in this manner is that prototype theorists have argued that a prototype-abstraction process is more likely to occur for larger size categories (e.g., Homa et al., 1981). The maximum-likelihood parameters and summary fits for the combined model are reported in Table 9. In the immediate condition, there was no evidence for the operation of a prototype-abstraction process, with the maximum-likelihood estimate

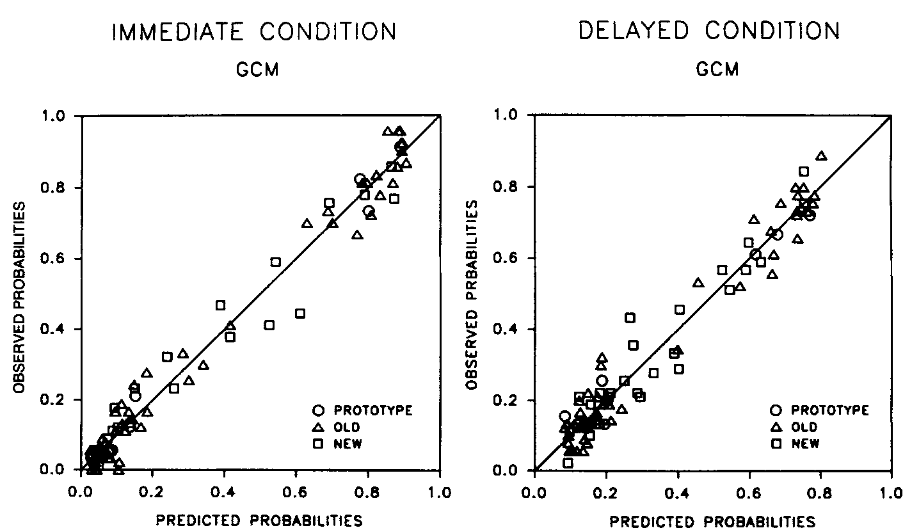

Figure 3. Scatterplot of observed classification probabilities against those predicted by the generalized context model (GCM) for the immediate and delayed conditions of Experiment 2.

of the mixing parameter y being zero for all three category sizes. The mixing parameter took on nonzero values for all three category sizes in the delayed condition. (Oddly, its magnitude was quite large for the Size 3 category, whereas prototype theorists would expect prototype abstraction to occur in large-size categories). However, the improvements in fit due to adding the prototype-use parameters were negligible. For example, by adding the three y parameters, the total sum of squared derivations (*SSE*) for the GCM was reduced from only .274 to .255. A likelihood-ratio test indicated that the improvement in fit due to adding these parameters was not statistically significant, $\chi^2(3, N = 2{,}700) = 5.94$, $p > .10$. In sum, these results provide little or no evidence for the use of prototype information in addition to exemplar-based generalization.

Discussion

In Experiment 2, we manipulated category size and time of delay of the transfer test and also used a learning design in which the patterns had mixed distortion levels. A major aim of the various manipulations was to produce conditions favorable to observing prototype enhancement effects. We were at least partially successful in achieving this aim, with the prototype being classified as well or better than the medium-level and high-level old distortions in the large-size category.

The main demonstration of interest is that the MDS-based exemplar model (the GCM) successfully predicted this prototype-enhancement effect. The exemplar model can predict prototype-enhancement effects because the prototype is at least moderately similar to most of the category exemplars. By contrast, any given exemplar may be highly similar only to itself. Thus, overall summed similarity to the exemplars of a category can be greater for the prototype than for individual exemplars themselves (cf. Hintzman, 1986; Nosofsky, 1988a).

In addition to accounting for the prototype-enhancement effect, the MDS-based exemplar model successfully accounted for the main effects of level of distortion, category size, and delay of the transfer test.

Most important, beyond predicting these main qualitative effects, the MDS-based exemplar model predicted fairly well the detailed quantitative structure of the classification data by accurately predicting the probability with which each of the individual patterns was classified into the alternative categories. Finally, as in Experiment 1, the prototype model yielded worse quantitative predictions of the classification data than did the GCM, and tests of the combined model again yielded little evidence for the operation of a prototype-abstraction process.

A surprising result is that we did not observe differential forgetting of the prototypes versus the old exemplars when delay was introduced. We suspect that observing this phenomenon may depend in subtle ways on the precise positioning of the prototypes and exemplars in the multidimensional psychological space, as well as on the overall levels of similarity that exist in the experiment. One way the GCM accounts for effects of delay is in terms of decreases in the overall level of the sensitivity parameter. The idea is that as delay increases, patterns may become less discriminable (more similar) in memory. Suppose that the prototype is centrally located with respect to the category exemplars in the multidimensional space and is distant from the exemplars of the contrast categories. Thus, when overall similarity increases because of delay, there could be dramatic increases in the within-category summed similarity for the prototype, yet only small increases in the between-category summed similarity. By contrast, for an old exemplar, which may be peripherally located in the category distribution, increases in within-category summed similarity may not be so dramatic. Indeed, if the old exemplar borders the distribution of a contrast category, its between-category summed similarity might increase dramatically with delay. This combination of factors would produce differential forgetting, in which there is less of a performance decrement for the prototypes than for the old exemplars.

In any case, differential forgetting was not observed in Experiment 2, and was not predicted by the exemplar model for the particular distribution of dot patterns that was generated. Future research will need to study more systematically the conditions under which the phenomenon does occur.

Experiment 3

In Experiment 3, we tested the MDS-based exemplar model on its ability to predict effects of other variables known to have influences on category learning in the prototype-distortion paradigm. Experiment 3 was a partial replication and extension of a recent study conducted by Homa et al. (1991), in which the variables of category size, frequency of individual training exemplars, and similarity relations of new distortions to specific training exemplars were factorially manipulated. We extended Homa et al.'s (1991) experiment by also collecting similarity judgments for the patterns and then using these similarity judgments to derive MDS solutions. At issue was whether the MDS-based exemplar model could predict in quantitative fashion the effects of these fundamental variables on classification performance.

In Experiment 3, subjects learned to classify objects into two categories. In one condition, category size was 3, whereas in a second condition, category size was 10. (Note that category size was manipulated as a between-subjects variable in Experiment 3, whereas in Experiment 2 it was a within-subjects variable.) The frequency of individual exemplars was also manipulated. In a baseline condition, all exemplars were presented with equal frequency during learning, whereas in a high-frequency condition, an individual exemplar from each category was presented five times as often as the other exemplars. Finally, we constructed new distortions that were either similar or dissimilar to the high-frequency old distortions.

According to the exemplar model, increases in presentation frequency lead to increases in the strength with which an exemplar is stored in memory. Because exemplar strength is assumed to combine multiplicatively with interitem similarity (see the later Theoretical Analyses section), the frequency-sensitive exemplar model predicts that classification accuracy should increase both for high-frequency exemplars and for category members that are similar to the high-frequency exemplars (Nosofsky, 1988b, 1991b). Little effect of presentation frequency is predicted, however, for items that are dissimilar to the high-frequency exemplars.

Method

Subjects

A total of 180 Indiana University undergraduates served as subjects. Some participated in Experiment 3 as part of an introductory psychology course requirement, whereas others were hired for their participation. (The same proportion of hired subjects participated in each condition.) Of the 180 subjects, 120 were assigned to four conditions of category size and frequency, with 30 subjects in each condition. Of the remaining 60 subjects, 30 provided similarity judgments for the items in the Size 3 condition and 30 provided similarity judgments for the items in the Size 10 condition. Five subjects were replaced because their performances in the learning phase were at chance level. All subjects were tested individually.

Stimuli and Apparatus

Following Homa et al. (1991), we used random polygons instead of dot patterns. A polygon was constructed by sequentially connecting the dots of a pattern with lines. The dot-generation procedure was the same as in Experiments 1 and 2. In addition to increasing comparability with the Homa et al. (1991) results, our main reason for using polygons was to increase the generality of the results of our studies.

Two categories of polygons were used in all of the classification learning conditions. First, two prototypes were created, and then 10 training exemplars (7.7 bits/dot) were generated from each prototype. All generated exemplars were used in the large category-size condition (Size 10). Three exemplars from each category were randomly selected as training exemplars for the small category-size condition (Size 3). One of the training exemplars in each category was randomly selected as a high-frequency exemplar (HF-old). The high-frequency old exemplars were the same in both category-size conditions.

The prototypes, as in Experiment 1, were redefined from the training exemplars in the Size 10 condition. They were also used as the redefined prototypes in the Size 3 condition. Two unrelated-new (URN) distortions were generated from each redefined prototype, and two related-new (RN) distortions were generated from each HF-old exemplar. The same new distortions were used in both size conditions. Note, therefore, that the patterns used in the Size 3 condition formed a subset of the patterns used in the Size 10 condition. The two URNs of each category were 7.7 bits/dot distortions of the redefined prototypes. One RN in each category was a 1 bit/dot distortion (RN_1) of the HF-old exemplar, whereas the second RN in each category was a 4.0 bits/dot distortion (RN_4) of the HF-old exemplar. Their distances from the prototype were roughly the same as for the training exemplars. IBM PCs were used for displaying the stimuli and controlling the experiment.

Procedure

Classification conditions. There were four conditions defined by category size (Size 3 and Size 10) and presentation frequency of the HF-old exemplars (equal frequency [EF] and unequal frequency [UF]). In Condition EF all exemplars were presented with the same frequency, whereas in Condition UF the HF-old exemplars were presented five times more frequently than the remaining exemplars. Each exemplar was presented once (or five times in the case of HF-old exemplars in Condition UF) in each block. There were eight blocks of learning trials. Order of presentation was randomized within each block. Again, there were three blocks of transfer trials following the learning phase, with each stimulus presented once in each block.

The Size 3 condition had six training exemplars (3 from each category) and 16 transfer items (3 old, 1 prototype, 2 related new, and 2 unrelated new from each category). The Size 10 condition had 20 training exemplars (10 from each category) and 30 transfer items (10 old, 1 prototype, 2 related new, and 2 unrelated new from each category). Subjects judged the similarity between all transfer items following the transfer test.

Similarity-judgment condition. The procedure was the same as in Experiments 1 and 2.

Results

In all conditions, learning performance in the first block was just slightly above chance (.50). By the final block, learning performance increased to slightly greater than .90 in the Size 3 conditions, and slightly greater than .85 in the Size 10 conditions. Not including the HF-old exemplars, performance was slightly better in the EF conditions than in the UF conditions but, including the HF-old exemplars, performance was better in the UF conditions.

The probability with which each item was correctly classified in its category in the transfer phase of each of the four main conditions (Size 3 EF, Size 3 UF, Size 10 EF, and Size 10 UF) is reported in Table 10.

The main results of the transfer phase are illustrated in Figure 4 (upper panels), which displays average performance levels for the old exemplars, HF-old exemplars, prototypes, related new distortions (RNs), and unrelated new distortions (URNs) as a function of category size and frequency conditions. Overall, in the Size 3 condition, the HF-old exemplars were classified better than the remaining old exemplars. Note, however, that in the equal frequency condition, the advantage for the HF-old exemplars is happenstance: The items that were randomly selected for the frequency manipulation accidentally turned out to be easier to classify than the remaining old exemplars. Indeed, even the new distortions that were similar to the HF-old exemplars (RN_1 and RN_4) were classified better than the old exemplars. Following the HF-old exemplars and RNs, in order of classification accuracy, were the old exemplars, the prototypes, and the unrelated new distortions. Perhaps the major difference in performance between the Size 3 and Size 10 conditions was that there was a large increase in classification accuracy for the prototypes and new distortions in the Size 10 conditions. Indeed, in the Size 10 conditions, the prototypes were classified as well as the old exemplars (with the exception of the HF-old exemplars in Condition UF).

In the Size 3 condition, there were clear increases in classification accuracy for the HF-old exemplars and the RNs as a result of the frequency manipulation (compare performance in Conditions UF and EF). Classification accuracy increased less for the remaining old exemplars and actually decreased slightly for the prototypes and unrelated new distortions. In the Size 10 condition, there were, once again, large increases in classification accuracy for the HF-old exemplars and RNs

Table 10
Correct Category Response Probabilities Observed and Predicted by the GCM, the Prototype Model, and Their Frequency-Sensitive Versions in Experiment 3

	Equal frequency condition			Unequal frequency condition		
Item	Observed	GCM	Prototype	Observed	GCM	Prototype
	Size 3 condition: Category 1					
P_1	.856	.759	.832	.767	.770	.871
HF-old	.878	.832	.712	1.000	.956	.941
O_2	.700	.795	.795	.767	.871	.601
O_3	.900	.831	.853	.911	.883	.730
RN_1	.789	.818	.744	.989	.937	.936
RN_4	.689	.773	.636	.867	.923	.902
URN_1	.589	.703	.661	.733	.839	.776
URN_2	.678	.712	.799	.656	.627	.718
	Size 3 condition: Category 2					
P_2	.589	.736	.805	.544	.658	.737
HF-old	.933	.867	.854	1.000	.941	.937
O_2	.556	.668	.481	.756	.804	.524
O_3	.911	.831	.722	.878	.865	.620
RN_1	.911	.852	.856	1.000	.927	.922
RN_4	.856	.817	.832	.911	.875	.926
URN_1	.578	.681	.728	.567	.682	.823
URN_2	.467	.492	.472	.333	.354	.282
	Size 10 condition: Category 1					
P_1	.922	.884	.953	.900	.882	.973
HF-old	.844	.865	.838	.956	.966	.959
O_2	.789	.865	.772	.900	.891	.695
O_3	.922	.922	.929	.933	.929	.921
O_4	.933	.913	.892	.867	.896	.920
O_5	.944	.848	.832	.900	.901	.887
O_6	.833	.795	.708	.911	.879	.738
O_7	.856	.913	.911	.800	.894	.870
O_8	.911	.893	.928	.944	.905	.915
O_9	.856	.894	.728	.911	.913	.773
O_{10}	.944	.875	.940	.933	.885	.955
RN_1	.822	.864	.885	.978	.953	.968
RN_4	.811	.832	.856	.900	.922	.948
URN_1	.711	.731	.726	.767	.806	.743
URN_2	.778	.804	.821	.789	.771	.854
	Size 10 condition: Category 2					
P_2	.811	.822	.955	.844	.772	.937
HF-old	.867	.862	.825	.978	.959	.891
O_2	.756	.842	.852	.833	.859	.861
O_3	.967	.951	.852	.967	.968	.913
O_4	.878	.876	.797	.956	.900	.800
O_5	.889	.888	.801	.933	.920	.870
O_6	.833	.875	.902	.822	.884	.924
O_7	.833	.867	.832	.900	.911	.860
O_8	.933	.891	.921	.867	.865	.873
O_9	.844	.856	.815	.778	.841	.798
O_{10}	.889	.840	.765	.778	.796	.676
RN_1	.900	.832	.821	.967	.914	.876
RN_4	.756	.759	.767	.844	.861	.832
URN_1	.867	.824	.871	.822	.824	.921
URN_2	.500	.556	.592	.578	.598	.567

Note. GCM = generalized context model; P = prototype; HF-old = high-frequency old exemplar; O = old exemplar; RN = related new; URN = unrelated new. O_1 in each category is the high-frequency old exemplar.

as a result of the frequency manipulation, but little if any effect for the remaining old exemplars, prototypes, and unrelated new distortions. These results are generally consistent with the exemplar model, which predicts increases in classification accuracy for high-frequency exemplars and items that are similar to these exemplars, but little change for items that are dissimilar to the high-frequency exemplars.[4]

Separate MDS solutions for items in the Size 3 and Size 10 conditions were derived from the similarity-judgment data by using the INDSCAL model. As in Experiments 1 and 2, we used the six-dimensional scaling solution for fitting the data in the Size 10 condition but decided to use a four-dimensional solution for fitting the data in the Size 3 condition. The reason for using the four-dimensional solution in the Size 3 condition was that in this condition there were only 16 items. With only 16 items, a six-dimensional scaling solution for the similarity data would probably be unstable. Moreover, with only 16 items, we wished to cut down on the number of free parameters used for fitting the exemplar and prototype models to the classification data. The MDS solutions for the patterns in the Size 3 and Size 10 conditions are reported in the Appendix. Averaged across Groups S and C/S, the correlations between the INDSCAL-derived distances and the similarity ratings were −.937 and −.943 in the Size 3 and Size 10 conditions, respectively.

Theoretical Analyses

Generalized Context Model

According to the frequency-sensitive version of the GCM (Nosofsky, 1988b, 1991b), the probability with which Stimulus i (Nosofsky, 1988b, 1991b) is classified in Category J is given by the following equation:

$$P(R_J \mid S_i) = b_J \sum_{j \in C_J} M_j s_{ij} / [\sum_{K} b_K \sum_{k \in C_K} M_k s_{ik}], \qquad (10)$$

where the b_J and S_{ij} parameters are defined as before, and M_j gives the memory strength for Exemplar j. In applications to Experiments 1 and 2, in which all exemplars were presented with equal frequency, all M_js were set at one. In the Experiment 3 applications, all M_js are set at one, except those corresponding to the high-frequency old exemplars in the UF conditions, which are allowed to be free parameters.

The frequency-sensitive GCM (Equations 10, 2, and 3) was fitted separately to the classification data in each of the four main conditions by using a maximum-likelihood criterion. The predicted probabilities are reported with the observed probabilities in Table 10, and the maximum-likelihood parameters and summary fits are reported in Table 11. The model accounted for 91.6, 96.4, 98.5, and 99.1% of the

[4] We were surprised to see such a large effect of the frequency manipulation in the Size 10 condition, however, because the exemplar model tends to predict that as category size increases, the role of individual exemplar frequency and of old–new similarity relations to specific old exemplars should be attenuated (Busemeyer, Dewey & Medin, 1984; Homa, Sterling, & Trepel, 1981; Homa, Dunbar, & Nohre, 1991). Any such attenuation effect is difficult to evaluate in this circumstance, however, because performance on the HF-old exemplars and RN_1s hits ceiling in the Size 3 condition.

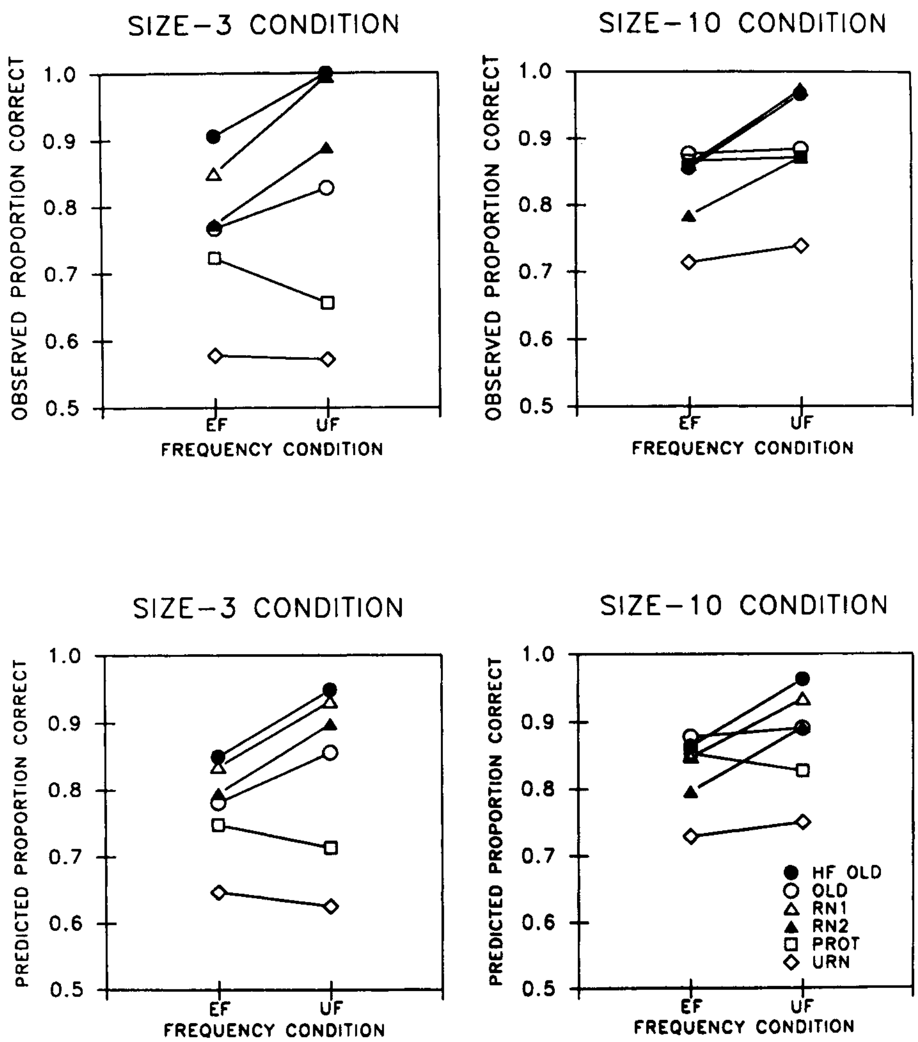

Figure 4. Proportion correct as a function of category size, frequency condition, and pattern type in Experiment 3. (Top panels show observed proportions. Bottom panels show proportions predicted by the generalized context model [GCM]. HF = high frequency; RN = related new; PROT = prototype; URN = unrelated new.)

variance in the Size 3 EF, Size 3 UF, Size 10 EF, and Size 10 UF conditions, respectively.

Figure 4 (bottom panels) illustrates that the model accounted for the major qualitative effects of the fundamental variables of category size, item frequency, and old–new similarity. First, the model predicted the category-size effect observed for the prototypes and unrelated new distortions. In general, as category size increased, there was an increased likelihood of the prototypes and new distortions contacting training exemplars to which they are similar. Second, the model predicted the increased classification accuracy for the HF-old exemplars and related new distortions that resulted from the frequency manipulation. Because increased item frequency leads to increased memory strength (Equation 10), high-frequency exemplars and their neighbors have large summed similarities. Third, the model predicted the effect of similarity of the new distortions to the HF-old patterns, in which the RN$_1$s are classified better than the RN$_4$s, which in turn are classified better than the URNs. In general, the more similar a transfer item is to a specific old exemplar, particularly one with a strong memory representation, the greater is its within-category summed similarity. The model even predicted the unexpected finding that in the Size 3 condition, there was a slight decrease in classification accuracy for the prototypes and unrelated new distortions when the HF-old exemplars were presented with high frequency. Apparently, some of these transfer items may have been more similar to an HF-old exemplar from the contrast category than from their own target category.

Beyond predicting these qualitative effects, the model makes fairly good quantitative predictions of classification probabilities for each of the individual items in the four main conditions, as is illustrated in scatterplots in Figure 5. Some quantitative shortcomings of the model, however, are that it underpredicted classification accuracy for the HF-old exemplars and RN$_1$s in the Size 3 UF condition (average observed

Table 11
Maximum-Likelihood Parameters and Summary Fits of the Classification Data for the GCM in Experiment 3

	Size 3 condition		Size 10 condition	
Parameter	EF	UF[a]	EF	UF
c	1.506	2.530	2.486	3.150
w_1	.484	.187	.348	.182
w_2	.023	.060	.184	.198
w_3	.107	.204	.239	.364
w_4	.386	.502	.104	.000
w_5			.212	.139
w_6			.000	.117
b_1	.503	.538	.486	.525
b_2	.497	.462	.514	.475
M		8.050		6.898
Fit				
$-\ln L$	63.437	62.241	84.232	75.017
SSE	.217	.140	.114	.076
% variance accounted for	91.639	96.368	98.522	99.144

Note. GCM = generalized context model; EF = equal frequency condition; UF = unequal frequency condition; c = general sensitivity parameter; w_m = attention weight given to dimension m; b_j = bias for making category response R_j; M = memory-strength parameter; $-\ln L$ = negative value of log-likelihood; *SSE* = sum of squared deviations between observed and predicted classification probabilities.
[a] The frequency-sensitive version of the GCM was used to fit the data for the UF condition.

= .997, and average predicted = .940) and slightly underpredicted accuracy for the prototypes in the Size 10 UF condition (average observed = .872, and average predicted = .827).

Inspection of the best-fitting parameters in Table 11 reveals that sensitivity was higher in the Size 10 condition than in the Size 3 condition. Although there was a total of eight blocks of training in both conditions, subjects had far more total trials of training in the Size 10 conditions than in the Size 3 conditions. Because perceptual/memorial differentiation is assumed to increase with learning experience (e.g., Gibson & Gibson, 1955; Nosofsky, 1987; Nosofsky et al., 1992; Shepard, 1957), the finding of increased sensitivity in the Size 10 condition seems reasonable.

In Table 12 we summarize the results of model-based analyses in which the sensitivity, attention weight, and bias parameters were held fixed across the EF and UF conditions, with only the memory-strength parameter allowed to vary. Although these constrained models fitted significantly worse than the full versions, the average percentage of variance accounted for decreased from only 94.0 to 92.2% in the Size 3 condition, and from only 98.8 to 98.4% in the Size 10 condition (see Tables 11 and 12). Thus, the MDS-based exemplar model can account parsimoniously for the effects of the exemplar-frequency manipulation with changes only in the value of the exemplar memory-strength parameter.

Prototype Model

To apply the prototype model to predict the effects of the frequency manipulation, the central tendency of each category is computed by using a weighted average over the dimensional coordinates of the training exemplars. The relative weight given to the high-frequency exemplars in computing the central tendency is allowed to be a free parameter in condition UF. Assuming a large relative weight, the central tendency would be shifted in the direction of the high-frequency exemplars in condition UF. In all other respects, the prototype model is the same as discussed previously.

The maximum-likelihood parameters and summary fits for the prototype model are reported in Table 13, and the predicted probabilities are shown with the observed probabilities in Table 10. Relative to the GCM, the prototype model performs very poorly in all conditions (despite the fact that it accounts for around 95% of the variance in the Size 10 conditions). For example, the total sum of squared deviations between predicted and observed probabilities for the prototype model is nearly four times as great as that of the GCM in the Size 10 conditions, and nearly three times as great as that of the GCM in the Size 3 conditions. The prototype model tended to overestimate performance on the prototypes and new distortions and to underestimate performance on some of the old exemplars.

Combined Model

One of Homa et al.'s (1991) central hypotheses was that when category size gets large, a prototype-abstraction process operates together with exemplar-based classification. To evaluate this hypothesis, the combined model was fitted separately to the transfer data in each of the four main conditions, and its fits were compared with that of the pure exemplar model. As explained previously, in the unequal-frequency conditions, the relative memory strength for the high-frequency exemplars was allowed to be a free parameter. (The same relative weighting was used to compute the weighted central tendency.) In all other respects, the combined model was the same as before.

The maximum-likelihood parameters and summary fits for the combined model are reported in Table 14. In the Size 3 condition, there was no indication of using prototype information, with the value of the prototype-use parameter (y) being zero in both conditions EF and UF. On the other hand, in the Size 10 condition, there was some suggestion for the use of prototype information, in accord with Homa et al.'s (1991) hypothesis. In the Size 10 EF condition, the value of the prototype-use parameter (y) was 14.234, and the improvement in fit over the pure exemplar model was statistically significant, χ^2 (1, N = 2,700) = 5.05, $p < .05$. However, in the Size 10 UF condition, the value of this parameter was only 2.262, and the improvement in fit was not statistically significant, $\chi^2(1, N = 2{,}700) = 1.602$, $p > .20$. Thus, the evidence for the use of prototype information is inconsistent. Even in the Size 10 EF condition, the improvement in fit due to the prototype-use parameter seems rather small (e.g., the

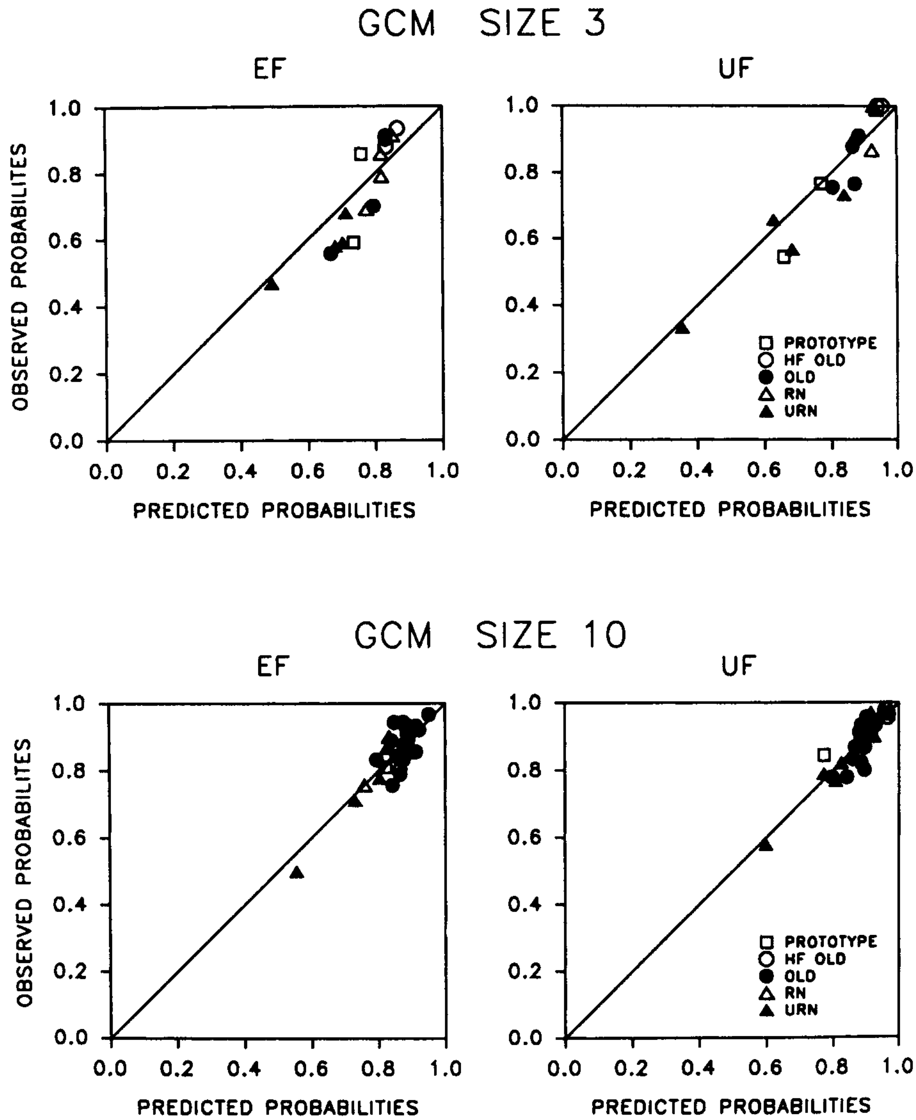

Figure 5. Scatterplots of observed classification probabilities against those predicted by the generalized context model (GCM) for each combination of category size and frequency conditions in Experiment 3. (EF = equal frequency condition; UF = unequal frequency condition; HF = high frequency; RN = related new; URN = unrelated new.)

SSE is reduced from only .114 to .099), so the evidence for the operation of a prototype-abstraction process is not compelling.

Discussion

In summary, Experiment 3 provided corroborating evidence for the well-known category-size effect for prototypes and new distortions. It also provided evidence of the importance of individual item frequencies in the learning of ill-defined categories and of the importance of similarity of new items to the high-frequency exemplars. Homa et al. (1991) suggested that with extensive training experience, individual item frequency may play only a limited role in shaping people's representations of ill-defined categories, especially for large-size categories. This hypothesis is an interesting and plausible one to pursue but, at least under the present conditions, with 10 exemplars in each of two categories and with eight blocks of training, individual item frequency clearly played a major role.

Most important to the theme of our investigation, the MDS-based exemplar model again provided excellent quantitative fits to the sets of classification transfer data and accounted for the major qualitative effects of the variables of category size, item frequency, and old–new similarity on classification performance. The MDS-based prototype model fared poorly relative to the exemplar model. Moreover, although our fits

Table 12
Maximum-Likelihood Parameters and Combined Fits of the Classification Data in the EF and UF conditions for the GCM in Experiment 3

Parameter	Size 3 condition	Size 10 condition
c	1.893	2.669
w_1	.331	.300
w_2	.049	.173
w_3	.090	.285
w_4	.530	.018
w_5		.147
w_6		.078
b_1	.517	.501
b_2	.483	.499
M[a]	3.507	3.794
EF fit		
−ln *L*	69.499	89.957
SSE	.267	.143
% variance accounted for	89.707	98.144
UF fit		
−ln *L*	73.894	83.994
SSE	.203	.112
% variance accounted for	94.728	98.738

Note. All parameters are common to both conditions. GCM = generalized context model; EF = equal frequency condition; UF = unequal frequency condition; c = general sensitivity parameter; w_m = attention weight given to dimension m; b_j = bias for making category response R_j; M = frequency parameter; −ln *L* = negative value of log-likelihood; *SSE* = sum of squared deviations between observed and predicted classification probabilities.[a] The memory-strength parameter was applied only to the data for the UF condition.

of the combined model provided some suggestion of the use of prototype information, the evidence was inconsistent and not compelling. We believe that more evidence is needed to demonstrate convincingly the operation of a prototype-abstraction process.

General Discussion

Summary

The aim of this research was to bridge two major research traditions that have compared and contrasted prototype and exemplar models of classification. The dot-pattern paradigm, initiated by the seminal work of Posner and Keele (1968, 1970), has been the source of innumerable studies and of a rich collection of experimental data. Exacting tests between exemplar and prototype models have not been forthcoming from this paradigm, however, because the models often make similar qualitative predictions, and the underlying psychological structure of the dot patterns is not specified. A second research tradition has compared formalized versions of exemplar and prototype models and tested these models quantitatively in highly simplified perceptual domains. These studies also have their limitations, however, because the naturalness and generalizability of the experimental materials is easily questioned.

By conducting similarity-judgment studies and deriving MDS solutions for the dot patterns, we hoped to bridge these two research traditions. Similarities of patterns to the category exemplars and prototypes could be computed from these derived scaling solutions and used in conjunction with the competing mathematical models to predict classification performance.

We feel that our efforts along these lines were surprisingly successful. Across three separate experiments with multiple conditions, an MDS-based exemplar model, namely, Nosofsky's (1986) GCM, consistently provided good quantitative predictions of classification. Not only did the model accurately predict overall levels of performance for types of dot patterns, such as the prototypes, old distortions, and new distortions, it accurately predicted classification probabilities for individual tokens of these main item types. The model also accurately predicted effects of fundamental learning variables on classification, such as effects of level of distortion of the patterns, category size, delay of transfer phase, individual item frequency, and old–new similarity relations.

Psychological Scaling

How important is the psychological scaling approach to achieving these quantitative predictions? Recall that in Experiment 1, a single-parameter version of the GCM, with only the sensitivity parameter allowed to vary, was able to account for 95.1% of the variance in the classification data when used in conjunction with the MDS solution for the dot patterns. For purposes of comparison, we again used this single-parameter GCM to predict the Experiment 1 classification data,

Table 13
Maximum-Likelihood Parameters and Summary Fits of the Classification Data for the Prototype Model in Experiment 3

	Size 3 condition		Size 10 condition	
Parameter	EF	UF[a]	EF	UF
c	2.503	1.721	3.617	3.268
w_1	.433	.399	.285	.275
w_2	.042	.279	.162	.335
w_3	.026	.000	.053	.000
w_4	.499	.322	.108	.000
w_5			.256	.286
w_6			.136	.104
b_1	.542	.511	.528	.517
b_2	.458	.489	.472	.483
M		9.959		6.341
Fit				
−ln *L*	82.706	83.372	119.578	137.971
SSE	.368	.640	.309	.404
% variance accounted for	85.810	83.372	95.983	95.442

Note. EF = equal frequency condition; UF = unequal frequency condition; c = general sensitivity parameter; w_m = attention weight given to dimension m; b_j = bias for making category response R_j; M = frequency parameter; −ln *L* = negative value of log-likelihood; *SSE* = sum of squared deviations between observed and predicted classification probabilities.
[a] The frequency-sensitive version of the prototype model was used to fit the data for the UF condition.

Table 14
Maximum-Likelihood Parameters and Summary Fits of the Classification Data for the Combined Model in Experiment 3

	Size 3 condition		Size 10 condition	
Parameter	EF	UF[a]	EF	UF
c	1.506	2.530	3.313	3.325
w_1	.484	.187	.243	.170
w_2	.023	.060	.120	.183
w_3	.107	.204	.322	.381
w_4	.386	.502	.036	.001
w_5			.130	.155
w_6			.150	.110
b_1	.503	.538	.516	.533
b_2	.497	.462	.484	.467
M		8.050		6.905
y	.000	.000	14.234	2.262
Fit				
−ln *L*	63.437	62.241	81.705	74.216
SSE	.217	.140	.099	.066
% variance accounted for	91.639	96.368	98.707	99.251

Note. EF = equal frequency condition; UF = unequal frequency condition; c = general sensitivity parameter; w_m = attention weight given to dimension m; b_j = bias for making category response R_j; M = frequency parameter; y = prototype-use parameter j; −ln *L* = negative value of log-likelihood; *SSE* = sum of squared deviations between observed and predicted classification probabilities. [a] The frequency-sensitive version of the combined model was used to fit the data for the UF condition.

except rather than using the MDS solution for the dot patterns, we used the objective coordinates of the dots. (Because the patterns were composed of nine dots, each with an *x* and a *y* coordinate, each dot pattern was described by an 18-element vector.) This objective-similarity GCM was able to account for only 81.5% of the variance in the classification data, with the sum of squared deviations between predicted and observed classification probabilities being nearly four times as great as that of the psychological similarity model. We conclude that the psychological scaling approach is critical in allowing the GCM to achieve its precise quantitative predictions of classification in the dot-pattern paradigm.

A fruitful avenue of future research would be to test alternative psychological scaling models to the ones used here. For example, instead of representing the dot patterns as points in a spatial solution, discrete clustering or network approaches may prove viable (e.g., Corter & Tversky, 1986; Hutchinson, 1989; Shepard & Arabie, 1979). In addition, in our research, each dot pattern was represented as a single point in the psychological space. More powerful similarity-scaling models have recently been proposed in which objects are represented as probabilistic distributions of points rather than as single points (e.g., Ashby & Lee, 1991; Ashby & Perrin, 1988; Ennis, Palen, & Mullen, 1988; MacKay, 1989; Zinnes & MacKay, 1983). Among other things, such probabilistic scaling models can account for asymmetries in similarities and other violations of the assumptions that underlie traditional MDS approaches (Tversky, 1977). Probabilistic scaling models might yield even better predictions of similarity and classification than were achieved in this study.

Prototype Abstraction

Another theme of this research was to compare the performance of an MDS-based prototype model with that of the GCM. In keeping with previous modeling efforts and with the spirit of the dot-pattern classification learning paradigm, we defined the prototype as the central tendency or centroid for all training exemplars of the category in the multidimensional psychological space. In all experiments, this central-tendency prototype model fared poorly relative to the GCM.

One might argue, however, that the actual prototype that subjects abstract during the learning phase is not the central tendency, but something else. Consider, therefore, the following preliminary exploration of this idea. We formalized an extremely general version of the prototype model and fitted it to the classification data obtained in Experiment 1. Rather than assuming that the prototype was the central tendency in the multidimensional psychological space, we allowed the prototype to be any single point in that space. Thus, we treated the values of the prototypes on each dimension as free parameters. Because there were three prototypes and six psychological dimensions, this procedure increased the number of free parameters in the prototype model by 18. Even this 26-parameter prototype model (18 prototype coordinates, 1 sensitivity parameter, 5 free dimension weights, and 2 free biases) yielded a worse fit to the classification data than did the 8-parameter GCM (−ln *L* = 225.7 for the GCM, −ln *L* = 240.4, for the generalized prototype model).

Beyond comparing the MDS-based exemplar model with the MDS-based prototype model, we also tested a model that assumed a combination of exemplar-based generalization and prototype abstraction. Because the combined model generalizes the exemplar model, it must fit at least as well as the exemplar model. The critical question in testing this model, therefore, was whether assuming a prototype-abstraction component added clearly to the ability of the exemplar model to predict the data. The general answer to this question appears to be "no." In only one condition across the three experiments was the improvement in fit, yielded by adding the prototype-use parameter, statistically significant and, even here, the improvement was far from dramatic. We believe more convincing evidence is needed before concluding that a prototype-abstraction process took place.

Our hope is that our research will stimulate tests of numerous alternative models on their ability to quantitatively predict performance in the dot-pattern classification paradigm. One model that is likely to perform well is Anderson's (1990, 1991) recently developed rational model of categorization, which is closely related to the GCM (see Nosofsky, 1991a, for discussion). In the rational model, individual exemplars are grouped into clusters during the learning process. The probability that an exemplar joins a cluster is determined jointly by the size of each cluster, the similarity of the exemplar to the cluster's central tendency, and the value of a coupling parameter, which is a free parameter in the model. Roughly, when the value of the coupling parameter is zero, each exemplar forms its own cluster, and the rational model is essentially the exemplar-based GCM. When the value of the coupling parameter is unity, the clusters that are formed correspond to the experimentally defined categories and the rational model becomes a pure prototype model. For intermediate values of the coupling parameter, however, the rational model behaves as a multiple-prototype model. Exemplars that are highly similar to one another are summarized by a prototype, but dissimilar exemplars in the same category form their own separate clusters that are summarized by new prototypes.

In a nutshell, instead of literally storing each exemplar as a unique memory trace, and instead of forming a single category prototype, the rational model provides a principled algorithm for how multiple prototypes may be formed to represent a category. By making use of psychological scaling solutions for the dot patterns, as we illustrated here for the GCM, we expect that the rational model could be applied directly to model performance in our experiments.

Limitations and Extensions

Although the scaling-based modeling of dot-pattern classification represents an important new research approach, our studies had some limitations. First, our model-based analyses involved the prediction of averaged subject data. In future work, it may be possible to collect extensive individual subject data and attempt to model performance at the level of individual subjects. Second, because the MDS solutions were derived from pairwise similarity-judgment data, the number of categories and training items used in the experiments was relatively small. (If too many items had been used, the number of pairwise similarity judgments needed to derive MDS solutions would have been enormous.) In Experiment 2, there was a condition in which one category had 9 training exemplars and the transfer phase was delayed for 1 week. In Experiment 3, there were conditions in which each of two categories had 10 training exemplars. Although these conditions would appear to be favorable to prototype abstraction, perhaps they are not sufficient. An aim of future investigation would be to increase category size and delay the transfer phase even more, and continue to challenge the MDS-based exemplar model. The experiments and analyses presented in this study can be viewed as representing a first step in that direction.

Finally, given our success in applying the MDS-based exemplar model in this dot-pattern domain, we wonder whether applications in perceptual domains with still more wealth and significance might be pursued, including how people learn to categorize faces, pieces of music, and works of art.

References

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429.

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

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

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

Breen, T. J., & Schvaneveldt, R. W. (1986). Classification of empirically derived prototypes as a function of category experience. *Memory & Cognition, 14,* 313–320.

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., & 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, pp. 57–105). New York: Freeman.

Clark, S. (1988). *A theory for classification and memory retrieval.* Unpublished doctoral dissertation, Indiana University.

Corter, J., & Tversky, A. (1986). Extended similarity trees. *Psychometrika, 51,* 429–451.

Ennis, D. M., Palen, J., & Mullen, K. (1988). A multidimensional stochastic theory of similarity. *Journal of Mathematical Psychology, 32,* 449–465.

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

Franks, J. J., & Bransford, J. D. (1971). Abstraction of visual patterns. *Journal of Experimental Psychology, 90,* 65–74.

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

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62,* 32–41.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Goldman, D., & Homa, D. (1977). Integrative and metric properties of abstracted information as a function of category discriminability, instance variability, and experience. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 375–385.

Hayes-Roth, B., & Hayes-Roth, F. (1977). Concept learning and the recognition and classification of exemplars. *Journal of Verbal Learning and Verbal Behavior, 16,* 321–338.

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

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

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.

Hock, H. S., Tromley, C., & Polmann, L. (1988). Perceptual units in the acquisition of visual categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 75–84.

Homa, D. (1984). On the nature of categories. *Psychology of Learning and Motivation, 18,* 49–94.

Homa, D., Chambliss, D. (1975). The relative contributions of common and distinctive information on the abstraction from ill-defined categories. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 351–359.

Homa, D., Cross, J., Cornell, D., Goldman, D., & Schwartz, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology, 101,* 116–122.

Homa, D., & Cultice, J. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 83–94.

Homa, D., Dunbar, S., & Nohre, L. (1991). Instance frequency, categorization, and the modulating effect of experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 444–458.

Homa, D., Rhoads, D., & Chambliss, D. (1979). Evolution of conceptual structure. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 11–23.

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

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 322–330.

Hutchinson, J. W. (1989). NETSCAL: A network scaling algorithm for nonsymmetric proximity data. *Psychometrika, 54,* 25–51.

Hyman, R., & Frost, N. H. (1975). Gradients and schema in pattern recognition. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 630–654). San Diego, CA: Academic Press.

Knapp, A. G., & Anderson, J. A. (1984). Theory of categorization based on distributed memory storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 616–637.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling* (Sage University paper series on Quantitative Applications in the Social Sciences, No. 07-001). Beverly Hills, CA: Sage.

MacKay, D. B. (1989). Probabilistic multidimensional scaling: An anisotropic model for distance judgments. *Journal of Mathematical Psychology, 33,* 187–205.

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, 10,* 333–352.

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.

Metcalfe-Eich, J. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10,* 104–114.

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

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 87–109.

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

Nosofsky, R. M. (1988b). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 54–65.

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

Nosofsky, R. M. (1991a). Relation between the rational model and the context model of categorization. *Psychological Science, 2,* 416–421.

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

Nosofsky, R. M. (in press). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition.* Hillsdale, NJ: Erlbaum.

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 282–304.

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 211–233.

Posner, M. I., Goldsmith, R., & Welton, K. E. (1967). Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology, 73,* 28–38.

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.

Reitman, J. S., & Bower, G. H. (1973). Storage and later recognition of exemplars of concepts. *Cognitive Psychology, 4,* 194–206.

Rosch, E. H., 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. (1958). Deduction of the generalization gradient from a trace model. *Psychological Review, 65,* 242–256.

Shepard, R. N. (1962a). Analysis of proximities: Multidimensional scaling with an unknown distance function: I. *Psychometrika, 27,* 125–140.

Shepard, R. N. (1962b). Analysis of proximities: Multidimensional scaling with an unknown distance function: II. *Psychometrika, 27,* 219–246.

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

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

Shepard, R. N., & Arabie, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review, 86,* 87–123.

Shin, H. J. (1990). *A similarity-scaling study of "dot patterns" classification and recognition.* Unpublished doctoral dissertation, Indiana University, Bloomington.

Strange, W., Keeney, T., Kessel, F. S., & Jenkins, J. J. (1970). Abstraction over time of prototypes from distortions of random dot patterns: A replication. *Journal of Experimental Psychology, 83,* 508–510.

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

Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology.* New York: Freeman.

Zinnes, J. L., & MacKay, D. B. (1983). Probabilistic multidimensional scaling: Complete and incomplete data. *Psychometrika, 48,* 27–48.

(*Appendix follows on next page*)

Appendix

Table A1
Six-Dimensional Scaling Solution for Experiment 1

Exemplar	Dimension 1	2	3	4	5	6
			Category 1			
P_1	−1.5597	0.0317	0.4041	−0.0806	0.7044	−0.6498
O_1	−0.3957	0.4043	−0.5009	1.3874	2.9568	1.3244
O_2	−0.4298	0.7670	0.1511	2.7576	−0.3260	−0.2771
O_3	−1.7295	0.3513	0.6034	−1.0291	0.2050	−0.1439
O_4	−1.0242	0.5849	0.8259	−0.7610	0.1401	2.1197
O_5	−1.7020	0.4002	0.7423	−0.7992	−0.0752	−0.1777
O_6	−1.7318	0.7908	0.6102	−1.3870	0.8574	−1.2062
N_l	−0.6580	1.3329	1.3447	0.5528	1.2421	−0.1631
N_m	−0.2911	−1.7077	1.2002	2.1050	−0.1559	−0.6702
N_h	−0.7722	0.7045	−0.8799	−0.2577	−0.9780	−2.2524
			Category 2			
P_2	1.3353	0.3275	0.0543	−0.6616	0.1806	−0.2880
O_1	1.0572	0.3240	0.7356	0.0128	0.5301	0.3156
O_2	1.1589	0.9646	0.5663	−0.2257	−0.3710	0.2569
O_3	1.2564	0.0153	−0.4414	−1.2021	0.9971	−0.0669
O_4	1.4310	−0.0661	0.6084	−0.1159	−0.4888	−0.3197
O_5	0.9951	0.8524	0.5920	0.0367	−0.5004	−2.0262
O_6	1.2649	0.0256	−1.4573	−0.3197	1.0346	−0.8378
N_l	1.2624	0.4207	0.3534	−0.7558	0.1718	−0.1813
N_m	1.0582	0.5241	1.4669	−0.2608	−0.1230	−0.5886
N_h	0.8404	−0.3782	0.8416	0.3203	−0.1264	1.7302
			Category 3			
P_3	0.1213	−2.0049	−0.5556	−0.1492	0.2473	0.3143
O_1	0.1726	1.4973	−2.2666	1.3864	−0.8222	−0.2460
O_2	−0.0099	1.5846	−0.8532	0.8982	−1.7078	1.9922
O_3	−0.7260	−2.1062	0.4022	0.9964	0.0206	−0.2523
O_4	−0.7126	−0.2446	−0.9565	−1.1702	−1.7786	0.8200
O_5	−0.5080	−0.5879	−1.7790	−1.4875	−0.1869	0.4917
O_6	0.6368	−0.6837	−0.2560	−0.8362	−0.0218	1.3730
N_l	0.2183	−2.0001	0.3115	−0.0454	−0.0266	0.5123
N_m	−0.6975	−1.0783	0.5422	0.6966	−2.5504	−0.4488
N_h	0.1393	−1.0457	−2.4102	0.3945	0.9511	−0.4545

Note. INDSCAL weights for Group C/S are .720, .344, .251, .275, .263, and .188. INDSCAL weights for Group S are .709, .362, .278, .249, .252, and .177. P_j = physical prototypes; O_j = old exemplars; N_l = low distortions; N_m = medium distortions; N_h = high distortions.

Table A2
Six-Dimensional Scaling Solution for Experiment 2

Exemplar	Dimension 1	2	3	4	5	6
			Category 1			
P_1	1.2067	−1.6245	−0.6660	0.2673	0.4606	−0.2204
O_l	1.3224	−1.4163	−1.3135	0.0713	0.3928	0.3859
O_m	0.8956	−1.8657	0.0891	−0.0168	0.6143	0.4090
O_h	0.7277	−0.4076	−0.3267	2.6634	−2.3710	0.2533
N_l	1.1242	−1.5320	−0.6839	−0.0063	0.2634	−0.6544
N_m	1.1231	−0.9766	0.2403	−0.5598	−0.3318	−1.5722
N_h	0.8087	−1.2878	−1.1225	0.1107	1.3103	1.8326
			Category 2			
P_2	0.7446	1.1592	0.9513	0.0775	−0.3944	−0.3311
O_{l1}	0.7952	0.9962	1.9097	−0.9617	−0.1548	1.9948
O_{l2}	0.6555	1.1649	0.8255	−0.6764	−0.3224	−0.7274
O_{m1}	0.5488	0.7225	0.1362	0.4104	0.5447	−1.7418
O_{m2}	0.9426	1.1294	−1.4550	−2.8740	−0.9043	0.7738
O_{h1}	0.6293	1.0874	0.5180	1.7720	−1.2869	−0.7537

Table A2 (*continued*)

Exemplar	Dimension 1	2	3	4	5	6
	Category 2 (*continued*)					
O_{h2}	1.1489	1.4846	0.1503	−1.3933	−0.4120	0.9162
N_l	1.0637	0.9462	−0.1822	0.0328	−0.9794	−0.8238
N_m	−0.1234	0.5741	1.0011	1.1934	2.2521	1.3463
N_h	0.2230	1.3547	0.3464	0.8936	2.1983	−0.3611
	Category 3					
P_3	−1.3762	−0.0586	−0.4475	−0.2407	−0.2455	−0.3910
O_{l1}	−1.4046	−0.0610	−0.1786	−0.1216	−0.2053	−0.1852
O_{l2}	−1.2485	−0.2151	−1.1538	0.2570	−1.0241	0.5001
O_{l3}	−1.3070	−0.4073	0.1492	−0.1017	−0.7556	−0.8327
O_{m1}	−1.0991	−0.1934	−1.6262	−0.7788	−0.3852	−0.7004
O_{m2}	−0.9270	0.4239	−0.1977	1.1662	0.2872	1.5573
O_{m3}	−1.2805	−0.5259	0.9888	−0.4293	−1.4318	1.5573
O_{h1}	−0.2649	−1.1087	2.1311	−1.1045	0.6115	−1.4226
O_{h2}	−0.2829	1.2821	−1.1526	1.0871	1.3395	−0.6966
O_{h3}	−1.1164	−1.0169	0.8775	−0.5990	−0.1211	0.4251
N_l	−1.2623	−0.0997	0.6224	−0.0140	−0.3670	−0.0646
N_m	−1.4201	0.0488	0.0137	−0.4580	0.3060	0.7006
N_h	−0.8474	0.4231	−1.4445	0.3334	1.1120	−1.1732

Note. INDSCAL weights for Group D are .757, .491, .173, .201, .159, and .167. INDSCAL weights for Group S are .762, .465, .216, .154, .171, and .153. P_j = physical prototypes; O_j = old exemplars; N_l = low distortions; N_m = medium distortions; N_h = high distortions.

Table A3
Four-Dimensional Scaling Solution for Size 3 Condition in Experiment 3

Exemplar	Dimension 1	2	3	4
	Category 1			
P_1	0.9694	0.0059	−0.6209	−0.1873
HF-old	0.9305	0.0776	−0.1429	−1.0717
O_2	0.6451	−1.6287	−0.8904	1.0324
O_3	1.1559	−0.3372	−0.7329	1.2986
RN_1	1.0060	0.1898	−0.2843	−0.9134
RN_4	0.6745	0.1773	−0.4199	−1.3818
URN_1	0.6511	−1.5177	0.4070	−0.9353
URN_2	0.8853	1.4833	−0.2765	1.0774
	Category 2			
P_2	−0.8916	0.4258	0.8141	−0.5962
HF-old	−1.5174	−0.2961	−0.5246	−0.2873
O_2	−0.0089	1.3389	0.4944	1.5733
O_3	−0.4424	−1.6693	2.7805	−0.4533
RN_1	−1.5231	−0.0079	−0.7041	−0.2034
RN_4	−1.5419	−0.2912	−1.4898	−0.4593
URN_1	−1.2977	0.1226	0.1353	1.9322
URN_2	0.3051	1.9219	1.4551	−0.4248

Note. INDSCAL weights for Group C/S are .697, .498, .311, and .226. INDSCAL weights for Group S are .703, .479, .328, and .218. P = physical prototypes; HF-old = high-frequency old exemplars; O = old exemplars; RN = related new distortions; URN = unrelated new distortions.

(*Appendix continues on next page*)

Table A4
Six-Dimensional Scaling Solution for Size 10 Condition in Experiment 3

Exemplar	Dimension 1	2	3	4	5	6
			Category 1			
P_1	1.0219	−0.7216	0.0976	0.2306	0.3107	0.8050
HF-old	1.0491	−0.2873	−0.9293	−1.3098	0.1752	1.0285
O_2	0.6616	−0.6742	−0.4599	1.6471	−2.0360	−1.2183
O_3	1.1630	−0.8203	0.1837	0.0547	−0.4740	−1.3086
O_4	0.6639	−1.3724	0.1793	−0.9942	−0.7799	0.5026
O_5	0.3289	−0.7510	0.9431	2.6868	0.8130	1.7058
O_6	−0.6615	−1.7869	−0.3285	0.3104	0.9705	1.8274
O_7	0.6187	−1.3887	−0.0961	0.5387	−0.3447	−1.6048
O_8	1.1293	−0.3747	0.4453	1.0226	−0.2598	−1.0630
O_9	1.3625	1.0750	−0.2169	0.1501	2.1266	0.2698
O_{10}	0.9936	−0.6340	1.6302	1.3545	−0.4047	−0.0140
RN_1	1.0806	−0.4745	−0.8090	−1.0060	0.3988	0.8344
RN_4	0.8147	−0.6511	−0.5738	−1.0349	0.7000	1.2420
URN_1	0.9205	−0.1507	−1.3600	−0.2456	−1.9087	0.0217
URN_2	0.8728	−0.5964	0.1759	−0.7507	2.0278	−1.1888
			Category 2			
P_2	−0.8352	1.1134	0.3223	−0.6718	−0.2788	0.1390
HF-old	−1.7199	−0.7317	−0.1396	−0.5504	−0.3510	−0.5780
O_2	−0.2496	1.1003	1.1637	−0.6821	−0.5739	−0.9194
O_3	−0.7028	1.0984	−3.1494	1.0944	0.2627	−0.3817
O_4	−0.0711	1.5427	1.3221	0.8833	−1.2754	0.5006
O_5	−1.4959	0.0652	−0.7188	0.8491	1.2659	−1.1767
O_6	−1.1607	0.5984	−0.1829	0.1958	0.6962	−0.5115
O_7	−0.3900	0.9843	2.1714	−0.1751	0.0056	0.1282
O_8	−0.4273	1.6937	−0.9903	0.1280	0.2572	−0.7118
O_9	−0.4664	1.0280	−0.4638	−1.0557	−1.9051	1.4363
O_{10}	0.2417	1.8244	−0.5165	0.2943	−0.3485	1.6476
RN_1	−1.7273	−0.7477	−0.0072	−0.7474	−0.1268	0.0096
RN_4	−1.7719	−1.1147	0.6233	−0.7751	−0.6200	−0.1825
URN_1	−1.5775	0.1587	0.3962	0.6772	0.6412	0.3347
URN_2	0.3344	0.9955	1.2879	−2.1187	1.0360	−1.5740

Note. INDSCAL weights for Group C/S: .700, .485, .304, .197, .209, and .172. INDSCAL weights for Group S: .695, .437, .316, .192, .198, and .182. P = physical prototypes; HF-old = high-frequency old exemplars; O = old exemplars; RN = related new distortions; URN = unrelated new distortions.

Received May 17, 1991
Revision received September 13, 1991
Accepted September 30, 1991 ■