

Comparisons between exemplar similarity and mixed prototype models using a linearly separable category structure

ROGER D. STANTON, ROBERT M. NOSOFSKY, and SAFA R. ZAKI
Indiana University, Bloomington, Indiana

Nosofsky and Zaki (2002) found that an exemplar similarity model provided better accounts of individual subject classification and generalization performance than did a mixed prototype model proposed by Smith and Minda (1998; Minda & Smith, 2001). However, these previous tests used a nonlinearly separable category structure. In the present work, the authors extend the previous findings by demonstrating a superiority for the exemplar generalization model over the mixed prototype model in a case involving a linearly separable structure. Because this structure has numerous features that Minda and Smith argued should be conducive to prototype-based processing, the results pose a significant challenge to the mixed prototype view.

A classic issue in the field of category learning concerns the contrast between exemplar and prototype models. According to exemplar models (Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1986), people represent categories by storing individual exemplars in memory and classify objects on the basis of how similar they are to the stored exemplars. By contrast, according to prototype models (Homa, 1984; Posner & Keele, 1968; Reed, 1972), people form a summary representation of a category, which is usually assumed to be a central tendency computed over the category-training instances. Classification decisions are based on the similarity of objects to the category prototypes.

There is a good deal of evidence that prototype models are insufficient as models of classification. Memories for individual exemplars appear to play at least some role in classification learning, especially in situations in which category size is small (Homa, Sterling, & Trepel, 1981; Nosofsky, 1992; Posner & Keele, 1968; Smith & Minda, 2000). However, in a recent series of articles, J. D. Smith and his colleagues (Minda & Smith, 2001; Smith & Minda, 1998, 2000; Smith, Murray, & Minda, 1997) have argued strongly for a *mixed prototype model* of classification. In this model, the prototype forms the core of the category representation, but observers develop all-or-none memories for specific exemplars as well. If an old item is presented, people may consult these all-or-none memories in order to classify it. However, generalization to new items is assumed to be based on similarity to the prototype.

In their work, Smith et al. (1997; Minda & Smith, 2001; Smith & Minda, 1998) claimed that the mixed prototype

model provided accounts of individual subject classification performance that were as good as or better than those of a well-known exemplar model, the generalized context model (GCM; McKinley & Nosofsky, 1995; Nosofsky, 1986). However, Nosofsky and Zaki (2002) pointed to limitations in the studies reported by Smith et al. that they argued invalidated these researchers' conclusions.

The first limitation was that in their model comparisons, Smith et al. (1997; Minda & Smith, 2001; Smith & Minda, 1998) confounded the issue of what is the nature of the category representation (prototypes vs. exemplars) with what type of response rule (probabilistic vs. deterministic) is used. The exemplar model they fitted was a constrained version of the GCM that assumed a probability-matching rule. By contrast, the prototype model they fitted made allowance for any level of probabilistic versus deterministic responding that was desired. Nosofsky and Zaki (2002) argued that to achieve more diagnostic tests of the type of category representation that is developed, it is essential that the competing models be equated in terms of their response rules.

A second limitation of Smith et al.'s (1997; Minda & Smith, 2001; Smith & Minda, 1998) model comparisons was that the paradigms they developed did not include tests of generalization to new transfer stimuli. Instead, these researchers tested the ability of the competing models to predict performance only on the old training exemplars themselves. However, the mixed prototype and exemplar models both predict that, with sufficient training, observers will learn to classify the training exemplars with near perfect accuracy. Therefore, at late stages of training, little information is provided by fitting the models only to the accuracy data obtained on the old exemplars. Instead, it is critical to test the models on their ability to predict generalization to new transfer stimuli as well.

Thus, Nosofsky and Zaki (2002) conducted an extension of the experimental paradigm developed by Smith

This work was supported by Grant R01 MH48494 from the National Institute of Mental Health. Correspondence concerning this article should be addressed to R. D. Stanton, Department of Psychology, Indiana University, Bloomington, IN 47405 (e-mail: rostanto@indiana.edu or nosofsky@indiana.edu).

et al. (1997; Smith & Minda, 1998). They used the same category-training structure as did Smith et al. (1997; Smith & Minda, 1998). However, following training, they conducted a transfer phase in which in addition to testing subjects on the old exemplars, they tested generalization performance on new items. Furthermore, the mixed prototype and exemplar models were equated in terms of their response rules. In these tests, the exemplar model provided dramatically better accounts of the observed classification performance of individual subjects than did the mixed prototype model.

The category structure that Nosofsky and Zaki (2002) used was highly diagnostic for contrasting the predictions from the mixed prototype and the exemplar models. A limitation of Nosofsky and Zaki’s study, however, is that the category structure was nonlinearly separable. As is well known, linear separability is a necessary requirement for a prototype strategy to yield perfect classification performance (Reed, 1972). The linear separability constraint can be overcome by the mixed prototype model by means of forming all-or-none traces of the individual training exemplars. Nevertheless, one can question the generality of Nosofsky and Zaki’s modeling results by arguing that they may hold only in cases involving nonlinearly separable category structures.

Therefore, the central purpose of the present research was to follow up on this earlier work by comparing the mixed prototype and the exemplar models on their ability to account for individual subject classification performance involving a linearly separable category structure. In particular, the category structure that we chose was one that Minda and Smith (2001) claimed gave strong advantages to the mixed prototype model. The categories are “well structured,” in the sense that there is high within-category similarity and low between-category similarity. In addition, the categories have reasonably large sizes and are composed of high-dimension stimuli. To the extent that the exemplar generalization model continues to provide a better account of individual subject classification performance in this domain, it would provide a strong challenge to the mixed prototype view of categorization.

THE FORMAL MODELS

Because the models have been described extensively in previous work, we will only briefly summarize them here. In the paradigms under investigation, the stimuli are composed of M binary-valued, separable dimensions, and observers learn to classify the stimuli into one of two categories. We describe the models as they are applied to such a paradigm.

Exemplar Generalization Model

According to the GCM, the probability that item i is classified into Category A is given by

$$P(A|i) = \frac{[\sum s(i,a)]^\gamma}{[\sum s(i,a)]^\gamma + [\sum s(i,b)]^\gamma}, \tag{1}$$

where $\sum s(i,a)$ and $\sum s(i,b)$ denote the summed similarities of item i to the exemplars of Categories A and B , respectively, and γ is a response-scaling parameter. The γ parameter (Ashby & Maddox, 1993; McKinley & Nosofsky, 1995) governs the extent to which responding is probabilistic versus deterministic. When $\gamma = 1$, observers respond probabilistically by “matching” to the relative summed similarities of each category, whereas as γ grows greater than 1, observers respond more deterministically with the category that yields the larger summed similarity.

In all of their tests of the GCM, Smith et al. (1997; Minda & Smith, 2001; Smith & Minda, 1998) held γ fixed at 1—that is, they considered only the version of the GCM that assumes a probability-matching response rule. It is well known, however, that individual observers often respond more deterministically than is predicted by this probability-matching version of the GCM (Ashby & Gott, 1988; Maddox & Ashby, 1993; McKinley & Nosofsky, 1995; Nosofsky, 1991). In the present research, we allow γ to vary as a free parameter to make allowance for differing degrees of probabilistic versus deterministic responding in classification.

According to the GCM, in situations involving binary-valued, separable dimensions, the distance between exemplars i and j is computed by using a weighted city-block metric,

$$d_{ij} = \sum_{m=1}^M w_m \cdot |x_{im} - x_{jm}|, \tag{2}$$

where x_{im} and x_{jm} denote the values of exemplars i and j on dimension m , respectively, and w_m ($0 \leq w_m \leq 1$, $\sum w_m = 1$) is the *attention weight* given to dimension m . The similarity between exemplars i and j is an exponential decay function of their distance (Shepard, 1987),

$$s_{ij} = \exp(-c \cdot d_{ij}), \tag{3}$$

where c is an overall sensitivity parameter. As has been explained in previous articles, the combination of the city-block distance metric formalized in Equation 2 and the exponential similarity function formalized in Equation 3 yields an interdimensional multiplicative rule for computing similarity (see Medin & Schaffer, 1978, and Nosofsky, 1984, for more extensive discussions). As a result, the context model is often referred to as a *multiplicative-similarity exemplar model*.

The similarities computed from Equations 2 and 3 are substituted into Equation 1 to yield the classification probabilities that are predicted by the model. The free parameters in the model are the response-scaling parameter γ , the sensitivity parameter c , and $M - 1$ freely varying attention weights (w_m).

Mixed Prototype Model

According to the mixed prototype model (Smith et al., 1997; Minda & Smith, 2001; Smith & Minda, 1998), the probability that an old training instance from Category A is correctly classified into Category A is given by

$$P(A|i) = e + (1 - e) \cdot \frac{g}{2} + (1 - e) \cdot (1 - g) \cdot \text{proto}(A|i), \tag{4}$$

where e ($0 \leq e \leq 1$) is the probability that the observer has formed a memory trace for the individual exemplar, g ($0 \leq g \leq 1$) is a guessing probability, and $\text{proto}(A|i)$ denotes the probability that item i is classified into Category A by a prototype process. If item i is a new item that was not presented during training, then $e = 0$, and classification is based solely on a combination of the guessing and the prototype processes.

Minda and Smith (2001) considered two different prototype processes, one based on a multiplicative-similarity calculation, such as occurs in the GCM, and a second based on an additive-similarity calculation. In both models, a prototype vector is defined for each of Categories A and B . The prototype for Category A is composed of the dimension values that occur most frequently among the members of Category A , and likewise for the prototype of Category B . We describe each submodel in turn.

Multiplicative-similarity prototype process. The multiplicative-similarity prototype model (Estes, 1986; Nosofsky, 1987, 1992) is directly analogous to the GCM. The distance between item i and Prototype A is given by

$$d_{iA} = \sum_{m=1}^M w_m \cdot |x_{im} - P_{Am}|, \tag{5}$$

where P_{Am} denotes the value of Prototype A on dimension m , and the w_m are the attention weights. The similarity between item i and Prototype A is then given by

$$s_{iA} = \exp(-c \cdot d_{iA}), \tag{6}$$

where c is the sensitivity parameter. Analogous equations are used to compute the similarity of item i to Prototype B . The probability with which item i is classified into Category A by the prototype process is given by

$$\text{proto}(A|i) = \frac{s_{iA}^\gamma}{s_{iA}^\gamma + s_{iB}^\gamma}, \tag{7}$$

where γ is the response-scaling parameter. However, as was explained by Nosofsky and Zaki (2002), within the framework of the multiplicative-similarity prototype model, the γ parameter cannot be estimated separately from the sensitivity parameter c .¹ Therefore, without loss of generality in this model, γ is set at 1. The free parameters in the multiplicative-similarity version of the mixed prototype model are the exemplar memory parameter e , the guessing parameter g , the overall sensitivity parameter c , and $M-1$ freely varying attention weights (w_m).

Additive prototype process. According to the additive prototype model advanced by Smith et al. (1997; Smith & Minda, 1998, 2000), when item i is presented, the evidence in favor of a Category A response is given by

$$E_{i,A} = \sum \delta_m \cdot w_m, \tag{8}$$

where w_m is the attention weight given to dimension m ; and δ_m is an indicator variable set equal to 1 if item i matches the prototype of Category A on dimension m , and set equal to zero if it mismatches the prototype of

Category A on this dimension. The probability that an observer classifies item i into Category A is given by

$$\text{proto}(A|i) = E_{i,A}. \tag{9}$$

The free parameters in the additive-similarity version of the mixed prototype model are the exemplar memory parameter e , the guessing parameter g , and $M-1$ freely varying attention weights (w_m).

We include a presentation of the additive prototype model in this article for purposes of completeness, because this version of the mixed prototype model is the one considered by Minda and Smith (2001, pp. 789–795). However, we find that this model never fits better, and often fits substantially worse, than does the multiplicative-similarity version of the mixed prototype model.

EXPERIMENT

The category training structure that we used for comparing the exemplar and the mixed prototype models in the present research is shown in Table 1. This linearly separable training structure is the same as one used previously by Minda and Smith (2001, Experiment 1, Category Set 2). There are two categories, A and B , composed of seven exemplars each. The exemplars vary along six binary-valued dimensions. Logical-value 1 on each dimension tends to indicate Category A membership, whereas logical-value 2 tends to indicate Category B membership. However, no single dimension is perfectly diagnostic. The training exemplars of Category A include the prototype ($A1$: 111111), two low distortions that differ from the prototype along a single dimension ($A2$ and $A3$), and four medium distortions that differ from the prototype along two dimensions ($A4$ – $A7$). A parallel structure exists for Category B .

Minda and Smith (2001) pointed to several aspects of this category structure that they suggested would be conducive to prototype-based processing. First, the cate-

Table 1
Category Structure Tested by Minda and Smith
(2001, Experiment 1, Set 2) and in the Present Experiment

Category A						
1.	1	1	1	1	1	1
2.	1	2	1	1	1	1
3.	2	1	1	1	1	1
4.	1	1	1	2	1	2
5.	2	1	1	1	1	2
6.	1	1	2	1	2	1
7.	1	2	2	1	1	1
Category B						
8.	2	2	2	2	2	2
9.	2	2	2	2	1	2
10.	2	2	1	2	2	2
11.	2	1	2	2	2	1
12.	1	2	2	2	2	1
13.	2	1	2	1	2	2
14.	1	2	1	2	2	2

gories have a larger size than has sometimes been used in previous tests between exemplar and prototype models; second, the stimuli vary along a larger number of dimensions; and third, the categories are well structured, in the sense that there is high within-category similarity and a low between-category similarity. Indeed, Minda and Smith observed that the multiplicative prototype model (MPM), even without extension in terms of the all-or-none exemplar memorization and guessing parameters, always yielded superior quantitative fits to the classification data than did the GCM in this paradigm. As we noted previously, however, they tested only a constrained version of the GCM in which the γ response-scaling parameter was held fixed at 1.

In the present experiment, we extended the paradigm used by Minda and Smith (2001) by conducting a transfer phase at the completion of classification learning. In the transfer phase, we tested the subjects with all 64 possible stimuli that can be constructed from the six binary-valued dimensions (including the original training stimuli). Unlike Minda and Smith, we included model comparisons in which the γ response-scaling parameter in the GCM was allowed to vary freely. Although the present category structure did not allow us to develop strong, a priori, qualitative contrasts between the predictions from the models, the complete set of transfer data nevertheless provided a rich source of constraints that was valuable for model testing. In particular, we identified for individual subjects various examples in which the best-fitting versions of each model made differing patterns of predictions.

Method

Subjects. Forty-six Indiana University students, both undergraduates and graduates, participated in the experiment. All the subjects were paid \$5. To ensure that the subjects would be motivated, they were also offered a \$5 reward for good performance. None of the subjects knew the purpose of the experiment.

Stimuli. The stimuli were the line drawings of buglike creatures used by Smith and Minda (1998; Minda & Smith, 2001). The bugs were constructed along the following six binary-valued dimensions: body size (short or long), eye type (closed green eye or open red eye), head shape (round or oval), antenna type (straight back with orange dot or curved forward with purple dot), feet type (blue semi-circle or gray triangle), and leg length (short or long). There were 14 stimuli in the training set (see Table 1). The transfer set consisted of all 64 stimuli that could be constructed from the six binary-valued dimensions, including the original training stimuli.

The assignment of physical dimensions to the logical structure of the stimuli was randomized. In addition, following Smith and Minda (1998), we used four different random assignments of the physical features to the logical values 1 and 2 shown in the structure in Table 1. For example, logical-value 1 on the body dimension could be either *short* or *long* for different subjects.

Procedure. In the training phase, the subjects were instructed to learn to classify the bugs into one of two categories. The dimensions were described prior to the start of the training phase. The subjects were instructed that no single dimension could be used to perfectly classify the bugs and that they would need to attend to multiple dimensions to achieve adequate performance. The subjects were also instructed that they would need to achieve perfect performance by the end of the training phase to have a chance to earn the \$5 bonus. The training phase consisted of 30 blocks of the 14 training patterns. The order of presentation of the training patterns

within each block was randomized. On each trial, one of the training patterns would appear on the computer screen, the subject would classify it, and feedback was then provided.

A transfer phase consisting of four blocks of the 64 transfer stimuli immediately followed the training phase. The subjects were instructed that both old and new bugs would be presented and that their task was to classify each bug into the correct category. The order of presentation of the transfer patterns was randomized within each block. No feedback was provided during the transfer phase. To ensure that the subjects would be motivated, they were instructed that they would receive a \$5 bonus if they classified each individual old training item with above-chance accuracy. The bonus was based only on the old items, because the new items did not have objective right and wrong answers.

Results

Our central aim in this research was to compare the GCM and the mixed prototype models on their ability to account for the complete set of transfer data (i.e., performance on both the old and the new items). The most incisive comparisons between the models with respect to these data are achieved when subjects learn to classify the training instances with high accuracy.² Therefore, following previous work (e.g., Medin, Altom, & Murphy, 1984; Nosofsky & Zaki, 2002), we grouped the subjects into two separate data sets on the basis of their performance. We defined *learners* to be those subjects who exceeded 50% classification accuracy for each individual old training instance presented during the transfer phase. Twenty-three of the 46 subjects achieved this criterion. The remaining 23 subjects were defined to be *nonlearners*. To provide a clearer picture of the results, in this article, we will often discuss the quantitative fits of the models to the learners' and the nonlearners' data separately. We emphasize, however, that none of our main conclusions would be changed if the model fits were evaluated on both groups of subjects simultaneously.

All of the models were fitted to the data of the individual subjects by using a maximum-likelihood criterion (Wickens, 1982). Specifically, we conducted a computer search for the parameter values of each model that maximized the log-likelihood statistic

$$\ln L = \sum \ln N_i! - \sum \sum \ln f_{ij}! + \sum \sum f_{ij} \ln p_{ij},$$

where N_i denotes the frequency with which stimulus i was presented, f_{ij} denotes the frequency with which the subject classified stimulus i into category j , and p_{ij} denotes the predicted probability with which the subject classified stimulus i into category j . (This likelihood function assumes that the responses for each subject are binomially distributed into the categories and that the distributions for each stimulus are independent.) As an auxiliary measure, we will also report the percentage of variance in the Category A response probabilities that is accounted for by each model.

Model Fits to the Old Items Only

Before reporting the results of our main analyses involving the complete set of transfer data, we start with some preliminary analyses. In particular, to achieve closer

comparisons with the results reported by Minda and Smith (2001), we conducted analyses in which the models were fitted only to the old-items data. In addition to fitting the full version of the GCM and the multiplicative mixed prototype model (MMPM), we also fitted the special cases of these models that were the focus of Minda and Smith's report. Specifically, we fitted the probability-matching version of the GCM in which the γ response-scaling parameter was held fixed at one, and we fitted the pure version of the MPM in which the exemplar memory and guessing parameters in the MMPM were held fixed at zero.

The summary fits of the models to the learners' and the nonlearners' old-items data are reported in Tables 2 and 3, respectively. (Note that a smaller value of the $-\ln L$ statistic reflects a better fit of a model.) Because numerous observers, especially the learners, classified the old items with uniformly high accuracy, the absolute quantitative fits of the models to these data are often extremely good. (It is easy to locate parameter values for which the models predict high accuracies on all of the training exemplars.) Nevertheless, as was found by Minda and Smith (2001), the MPM does provide a better overall quantitative fit to the individual subjects' data than does the probability-matching version of the GCM [with $\gamma = 1$; $t(45) = -3.54, p = .001$]. This same pattern of results holds for both the learners and the nonlearners, although it is stronger for the nonlearners. Thus, when

we restrict the line of inquiry to the one followed by Minda and Smith, we obtain a set of results entirely parallel to that observed by these researchers.

Critically, however, when the γ response-scaling parameter is allowed to vary, the fits of the GCM improve considerably. Indeed, the full version of the GCM provides a significantly better overall quantitative fit to the old-items data than does the MPM [$t(45) = 2.22, p < .05$]. Thus, when the exemplar model is granted the same flexibility as is the MPM with respect to response-scaling processes, the results favor the exemplar model.

Nevertheless, by augmenting the MPM with the exemplar memory and guessing parameters, the MMPM's fit is improved as well. The bottom line is that the GCM and the MMPM are indistinguishable with respect to the old-items data [$t(45) = 0.94, p = .352$]. Therefore, a more diagnostic form of data is needed to discriminate between the models. With this goal in mind, we will now compare the models on their ability to account simultaneously for classification transfer performance on both the old *and* the new items.

Modeling Results for the Learners

The summary fits of the full version of the GCM, the MMPM, and the additive mixed prototype model (AMPM) to the complete set of transfer data from the learners are reported in Table 4. The full GCM provided a better fit than did the MMPM to the data of 13 of the 23 learners

Table 2
Summary Fits of the Full Generalized Context Model (GCM), the Multiplicative Mixed Prototype Model (MMPM), the Probability-Matching GCM (PM-GCM), and the Multiplicative Prototype Model (MPM) to the Learners' Old-Items Data

Subject	GCM		MMPM		PM-GCM		MPM	
	$-\ln L$	% Var	$-\ln L$	% Var	$-\ln L$	% Var	$-\ln L$	% Var
1	0.000	100.00	0.000	100.00	0.000	100.00	0.000	100.00
2	0.000	100.00	0.000	100.00	0.000	100.00	0.000	100.00
3	0.000	100.00	0.000	100.00	0.000	100.00	0.000	100.00
4	0.863	100.00	0.863	100.00	1.342	99.63	0.863	100.00
5	0.863	100.00	1.628	99.06	1.628	99.06	1.628	99.06
6	0.000	100.00	0.000	100.00	0.000	100.00	0.000	100.00
7	7.470	95.10	5.201	98.05	7.470	95.10	9.168	95.71
8	2.545	98.92	2.491	98.99	3.256	97.99	2.491	98.99
9	5.568	96.92	3.817	98.47	5.635	96.48	5.799	96.30
10	4.113	97.93	5.162	97.26	4.113	97.93	5.162	97.26
11	0.000	100.00	0.000	100.00	0.000	100.00	0.000	100.00
12	0.000	100.00	0.000	100.00	0.000	100.00	0.000	100.00
13	0.000	100.00	0.000	100.00	0.000	100.00	0.000	100.00
14	0.863	100.00	0.863	100.00	1.628	99.06	0.863	100.00
15	1.726	100.00	3.651	97.79	1.726	100.00	3.652	97.79
16	1.726	100.00	1.726	100.00	2.923	98.83	1.728	100.00
17	0.863	100.00	0.863	100.00	1.628	99.06	0.863	100.00
18	0.000	100.00	0.000	100.00	0.000	100.00	0.000	100.00
19	0.000	100.00	0.000	100.00	0.000	100.00	0.000	100.00
20	2.898	99.72	2.796	99.82	3.978	98.26	2.796	99.82
21	1.726	100.00	1.726	100.00	2.683	99.21	1.727	100.00
22	0.863	100.00	0.863	100.00	1.342	99.63	0.863	100.00
23	1.726	100.00	1.726	100.00	2.923	98.83	1.728	100.00
Mean	1.470	99.50	1.451	99.54	1.838	99.09	1.710	99.35

Note.— $-\ln L$, negative log-likelihood fit; % Var, percentage of variance in the Category A response probabilities that is accounted for by each model.

Table 3
Summary Fits of the Full Generalized Context Model (GCM), the Multiplicative Mixed Prototype Model (MMPM), the Probability-Matching GCM (PM-GCM), and the Multiplicative Prototype Model (MPM) to the Nonlearners' Old-Items Data

Subject	GCM		MMPM		PM-GCM		MPM	
	$-\ln L$	% Var	$-\ln L$	% Var	$-\ln L$	% Var	$-\ln L$	% Var
1	0.863	100.00	0.863	100.00	3.090	95.34	0.863	100.00
2	20.039	39.09	19.989	39.41	23.090	21.26	19.989	39.41
3	8.324	92.04	7.728	93.24	9.081	89.78	8.318	92.06
4	0.863	100.00	3.906	91.41	3.906	91.41	3.908	91.40
5	4.283	98.15	5.319	97.99	4.693	96.73	4.421	97.78
6	11.147	84.22	11.523	82.93	11.160	84.36	11.579	82.59
7	2.708	98.85	2.609	98.98	7.204	89.99	2.610	98.98
8	0.863	100.00	0.863	100.00	6.690	92.60	0.863	100.00
9	2.709	100.00	2.707	100.00	10.554	79.32	2.708	100.00
10	5.356	96.41	6.957	93.68	5.356	96.40	10.508	89.59
11	0.876	100.00	9.452	74.75	10.373	68.89	9.451	74.75
12	2.609	98.98	3.037	98.64	5.727	91.98	3.038	98.64
13	15.418	70.16	12.410	82.41	15.418	70.16	15.641	70.32
14	9.756	89.63	8.980	90.64	10.987	87.41	9.767	89.60
15	0.863	100.00	0.863	100.00	3.906	91.41	0.863	100.00
16	7.540	94.98	7.522	94.99	9.833	91.19	7.522	94.99
17	0.863	100.00	0.863	100.00	3.906	91.41	0.863	100.00
18	21.673	5.29	21.673	5.29	21.673	5.29	21.673	5.29
19	6.883	92.19	6.710	92.62	9.383	80.59	6.710	92.67
20	0.000	100.00	0.000	100.00	4.523	90.85	0.000	100.00
21	4.339	98.81	4.335	98.82	7.321	92.12	4.343	98.80
22	0.863	100.00	0.863	100.00	3.090	95.34	0.863	100.00
23	0.000	100.00	0.000	100.00	4.523	90.85	0.000	100.00
Mean	5.602	89.51	6.051	88.51	8.499	81.94	6.370	87.69

Note— $-\ln L$, negative log-likelihood fit; % Var, percentage of variance in the Category A response probabilities that is accounted for by each model.

(with one tie). For many of the learners, the advantage in fit for the GCM was substantial. In no case was a substantial advantage in fit provided by the MMPM. When the results from the individual subjects were used as the units of analysis, the fit of the GCM (mean $-\ln L = 31.5$) was significantly better than that of the MMPM [mean $-\ln L = 43.9$; $t(22) = 2.75$, $p < .05$]. The advantage of the GCM is even more pronounced when compared with the AMPM, with the GCM yielding better fits for 22 of the 23 learners. A parallel set of results is observed for the percent-variance measure of fit. Averaged across subjects, the GCM accounted for 89.1% of the variance in the observed Category A response probabilities, the MMPM accounted for 80.4% of the variance, and the AMPM accounted for 68.2% of the variance.

Table 4 also reports fits of the probability-matching version of the GCM with $\gamma = 1$. As was expected, this constrained version of the model provides relatively poor fits to these data because it fails to account for the deterministic levels of responding exhibited at the individual subject level. To gain another sense of the importance of the γ response-scaling parameter, we considered the best-fitting values of γ that were obtained when the full version of the GCM was fitted to the data. This analysis revealed that 22 of the 23 learners had γ estimates greater than one.

Modeling Results for the Nonlearners

The log-likelihood fits for the 23 participants in the nonlearners group are reported in Table 5. The GCM

provided better quantitative fits than did the MMPM for 15 of the 23 nonlearners, although there were only a few cases in which the fit advantage was substantial. There were no cases in which the mixed model produced a distinct advantage in quantitative fit. When the results from the individual subjects were used as the units of analysis, the fits of the GCM (mean $-\ln L = 45.6$) and the MMPM (mean $-\ln L = 52.0$) were not significantly different [$t(22) = 1.78$, $p > .05$], although the direction of model fits still favored the GCM. If the data from the learners and nonlearners are combined, the fit advantage for the GCM is statistically significant [$t(45) = 3.26$, $p < .01$]. Finally, as was the case for the learners' group, the AMPM and the probability-matching version of the GCM provide relatively poor fits to the nonlearners' data.

In general, the reason that it is more difficult to distinguish between the predictions from the MMPM and those from the GCM in the nonlearners group is that many of these subjects used single-dimension "rules" as a basis for classification. The multiplicative-similarity prototype and the exemplar models account equally well for this type of behavior by placing all attention weight on the single dimension and by setting the response scaling (or sensitivity) parameter at a sufficiently high level (see Nosofsky & Zaki, 2002, for a more detailed discussion).

Qualitative Comparisons

To provide some sense of the reason for the better fits provided by the full GCM, as compared with the mixed prototype models, we will consider the results for some

Table 4
Summary Fits of the Full Generalized Context Model (GCM), the Multiplicative Mixed Prototype Model (MMPM), the Additive Mixed Prototype Model (AMPM), and the Probability-Matching GCM (PM-GCM) to the Learners' Complete Transfer Data

Subject	GCM		MMPM		AMPM		PM-GCM	
	$-\ln L$	% Var	$-\ln L$	% Var	$-\ln L$	% Var	$-\ln L$	% Var
1	69.376	60.57	62.850	69.08	93.830	50.57	76.648	58.70
2	24.048	97.62	20.608	97.61	69.507	66.75	54.398	77.16
3	13.457	98.87	40.505	78.98	71.973	60.10	13.644	98.82
4	16.608	98.08	15.408	98.12	64.345	73.04	27.883	91.01
5	30.255	87.41	53.418	73.53	55.239	72.30	31.421	88.07
6	9.169	98.54	33.723	83.45	47.368	79.06	30.026	86.99
7	41.537	90.21	39.389	91.62	42.058	90.61	68.040	71.53
8	12.745	99.02	56.480	72.92	56.480	72.92	37.635	85.52
9	38.535	92.05	57.860	74.21	67.808	67.27	38.592	91.56
10	52.672	80.46	52.395	80.70	62.934	72.27	53.529	79.93
11	31.108	93.78	65.998	65.54	69.881	63.50	31.520	92.84
12	59.208	69.46	55.449	69.86	89.868	51.35	82.836	55.87
13	0.026	100.00	0.026	100.00	83.653	61.06	36.389	81.03
14	21.582	97.15	30.769	91.07	44.140	81.98	30.073	90.03
15	14.591	98.53	28.698	90.76	53.739	73.50	35.005	84.76
16	38.396	90.06	36.935	92.24	36.601	91.94	45.860	87.09
17	13.062	98.58	13.128	98.58	65.593	72.63	13.328	98.51
18	7.850	99.58	96.261	43.28	96.917	44.61	30.027	86.02
19	4.779	99.40	4.731	99.40	59.958	74.32	7.306	99.24
20	27.901	95.73	26.006	93.79	68.332	70.59	33.684	92.44
21	44.783	88.40	59.839	74.93	60.544	74.61	46.987	86.51
22	54.403	75.45	56.924	73.33	67.944	66.73	54.739	75.31
23	98.494	40.67	102.132	37.11	102.299	36.13	98.494	40.67
Mean	31.504	89.11	43.893	80.44	66.566	68.17	42.785	82.59

Note— $-\ln L$, negative log-likelihood fit; % Var, percentage of variance in the Category *A* response probabilities that is accounted for by each model.

of the individual subjects in depth. Our general plan of analysis is as follows. First, we will use the best-fitting parameters from the model fits to determine which dimensions of the exemplars each individual subject attended to. Next, we will locate critical training stimuli with the following property: Although a particular training stimulus is a member of one category, its values along the attended dimensions are ambiguous with respect to category membership. That is, half its attended dimension values point to Category *A*, and the other half point to Category *B*. (We were unable to locate cases in which the majority of dimension values pointed to the opposite category.) Finally, we will consider performance on transfer stimuli that match the critical training stimulus along the attended dimensions. We will refer to such transfer stimuli as the *neighbors* of the critical training stimulus. Because the GCM assumes that generalization of new transfer stimuli is based on similarity to specific exemplars, it tends to predict that the neighbors are classified with high probability into the same category as the critical training stimulus. By contrast, because the mixed prototype model assumes that generalization of new transfer stimuli is based on similarity to the prototype, it tends to predict that the neighbors are classified with roughly equal probability into each category. The reason is that half the neighbors' attended dimensions point to Category *A*, and the other half point to Category *B*. Note that this contrast in predictions does not obtain for the critical training stimulus itself, because the mixed prototype model can adjust its all-or-

none exemplar memory parameter (e) to correctly classify this old stimulus. Thus, tests of generalization of the new transfer stimuli are critical to distinguishing between the predictions of the models.

Table 6 provides a listing of the observed Category *A* choice probabilities for each of the 64 transfer stimuli for Subjects 18, 9, 3, and 13 of the learners group. The greatest advantage for the GCM over the MMPM was observed for Subject 18. According to the GCM, this subject focused attention on Dimensions 3, 4, 5, and 6. Training Exemplar 6 from Category *A* provides an example of a critical training stimulus, with half its attended dimensions pointing to Category *A* and the other half pointing to Category *B*. The neighbors of Exemplar 6 (along its attended dimensions) are Transfer Stimuli 43, 53, and 59. In accord with the predictions from the GCM, and in contrast to the predictions from the MMPM, Subject 18 always classified these transfer stimuli into Category *A*. A similar example is provided by considering Training Exemplar 4 and its neighbors (Transfer Stimuli 40, 50, and 56). Again, Subject 18 almost always classified these transfer stimuli into Category *A*, despite the fact that at least half their attended dimension values point to the opposite category. The results for Subject 9, who also attended selectively to Dimensions 3–6, are similar to those just described for Subject 18, although not quite as extreme.

Another illustration of the reason for the GCM's fit advantage is provided by Subject 3. According to the GCM analyses, this subject focused attention on Dimen-

Table 5
Summary Fits of the Full Generalized Context Model (GCM), the Multiplicative Mixed Prototype Model (MMPM), the Additive Mixed Prototype Model (AMPM), and the Probability-Matching GCM (PM-GCM) to the Nonlearners' Complete Transfer Data

Subject	GCM		MMPM		AMPM		PM-GCM	
	-ln L	% Var	-ln L	% Var	-ln L	% Var	-ln L	% Var
1	6.612	99.22	14.868	96.26	69.544	69.69	25.315	88.48
2	77.316	53.74	77.190	53.96	76.761	54.07	87.406	38.02
3	49.934	85.34	50.126	85.37	50.547	84.46	52.670	84.12
4	39.389	87.54	42.962	86.24	66.657	69.87	47.154	82.57
5	61.075	70.35	59.662	71.68	66.259	67.04	69.406	63.43
6	57.754	77.77	57.066	78.93	58.534	76.86	66.646	71.74
7	37.709	92.72	38.084	91.88	39.630	90.25	44.238	89.35
8	42.846	88.71	42.873	88.84	42.732	88.59	49.148	86.32
9	26.516	93.35	36.301	87.53	43.501	83.86	71.745	69.18
10	28.441	92.19	109.995	42.26	109.995	42.26	29.699	92.19
11	46.209	77.60	47.632	77.45	58.346	72.63	75.551	63.72
12	60.001	71.79	59.983	72.64	60.943	70.94	63.919	68.19
13	86.426	42.15	99.968	29.94	99.164	30.43	88.628	42.15
14	32.965	94.63	32.783	94.66	33.154	94.59	48.835	90.48
15	26.250	92.12	35.968	85.20	51.139	77.88	33.962	86.81
16	35.374	94.76	35.198	94.80	35.413	94.74	45.056	91.21
17	27.495	95.85	28.315	95.24	63.313	73.06	42.089	85.94
18	83.329	-8.62	83.664	-6.91	83.802	-7.51	84.074	-8.65
19	81.440	49.70	81.334	49.83	88.232	46.05	86.624	46.45
20	39.459	89.47	55.166	72.88	70.186	62.86	44.133	83.66
21	56.215	77.27	57.855	75.91	71.100	65.78	65.251	72.77
22	34.214	91.14	37.453	88.70	42.578	85.06	41.688	87.08
23	11.727	98.39	10.416	98.63	71.394	68.96	44.379	83.68
Mean	45.595	78.57	51.950	74.43	63.171	67.93	56.679	72.13

Note—-ln L, negative log-likelihood fit; % Var, percentage of variance in the Category A response probabilities that is accounted for by each model.

sions 1, 4, 5, and 6. In this case, the critical training stimuli were Training Exemplars 4 and 5 from Category A and Training Exemplar 12 from Category B. (Again, for all three of these training exemplars, half their attended dimension values pointed to Category A, and half pointed to Category B.) Transfer Stimuli 36, 47, and 50 matched Training Exemplar 4 on the attended dimensions, whereas Transfer Stimuli 42, 58, and 61 matched Training Exemplar 5 on these dimensions. In accord with the exemplar similarity predictions from the GCM, Subject 3 always classified these transfer stimuli into Category A. The neighbors of Training Exemplar 12 (from Category B) were Transfer Stimuli 26, 48, and 51. Subject 3 always classified these neighbors of Training Exemplar 12 into Category B.³

As was noted previously, the additive version of the mixed prototype model performs even worse than does the multiplicative version. A good illustration of the reason for the additive model's poor fits is provided by the data of Subject 13 (see Table 6). According to the best-fitting parameter estimates of all of the models, Subject 13 split attention nearly equally among Dimensions 1–5 and ignored Dimension 6. Inspection of Table 6 reveals that in all cases in which a majority of the attended dimension values pointed to Category A, the subject classified the stimulus into Category A with a probability of one; whereas in all cases in which the majority of the dimension values pointed to Category B, the subject classified the stimulus into Category B with a probability of one. Both the GCM

and the MMPM account perfectly for this deterministic pattern of responding by setting the response-scaling (or sensitivity) parameter at a high value. Without an analogous response-scaling mechanism, however, the AMPM is constrained to predict a much different pattern of behavior. For example, consider a new transfer stimulus in which three of the five attended dimensions point to Category A. Assuming equal attention to the five dimensions, it is straightforward to verify from Equations 8 and 9 that the AMPM predicts that this stimulus will be classified into Category A with a maximum probability of .60 (which falls far short of 1.00). Thus, by including tests of generalization of new transfer stimuli, we find examples that clearly illustrate that the AMPM is in need of a response-scaling mechanism, just as occurs for the GCM and the MMPM. (See also Nosofsky and Zaki, 2002, Experiment 1, for a previous demonstration of this point.)

Power Analyses

In their previous research, Minda and Smith (2001) had claimed that results from the present category structure provided strong evidence in favor of prototype-abstraction processes and evidence against exemplar similarity models. Our purpose in replicating and extending their experiment in the present research was to dispute this earlier claim and to demonstrate that the exemplar similarity model accounts well for performance in this paradigm. Furthermore, our finding that the exemplar model yielded far superior fits to the data of

Table 6
Observed Category A Response Probabilities for Example
Subjects That Yield Large Differences in the Fits of the
Alternative Models

Stim	Dimension Coding	Subject			
		18	9	3	13
1	111111	1.00	1.00	1.00	1.00
2	121111	1.00	.75	1.00	1.00
3	211111	1.00	1.00	1.00	1.00
4	111212	1.00	.75	1.00	1.00
5	211112	1.00	1.00	1.00	1.00
6	112121	1.00	1.00	1.00	1.00
7	122111	1.00	1.00	1.00	1.00
8	222222	.00	.00	.00	.00
9	222212	.00	.00	.00	.00
10	221222	.00	.00	.00	.00
11	212221	.00	.00	.00	.00
12	122221	.00	.00	.00	.00
13	212122	.00	.25	.00	.00
14	121222	.00	.00	.00	.00
15	112111	1.00	1.00	1.00	1.00
16	111211	1.00	.25	1.00	1.00
17	111121	.00	1.00	1.00	1.00
18	111112	1.00	1.00	1.00	1.00
19	122222	.00	.00	.00	.00
20	212222	.00	.00	.00	.00
21	222122	.00	.50	.00	.00
22	222221	.00	.00	.00	.00
23	111122	.00	1.00	.00	1.00
24	112112	.00	.25	1.00	1.00
25	121112	1.00	1.00	1.00	1.00
26	111221	.00	.00	.00	1.00
27	121121	.00	1.00	1.00	1.00
28	211121	.00	1.00	.75	1.00
29	112211	.00	.00	1.00	1.00
30	121211	1.00	.75	1.00	1.00
31	211211	1.00	.50	.00	1.00
32	212111	1.00	1.00	1.00	1.00
33	221111	1.00	1.00	1.00	1.00
34	112222	.00	.00	.00	.00
35	122122	.00	.00	.25	.00
36	122212	.00	.00	1.00	.00
37	211222	.00	.00	.00	.00
38	212212	.00	.00	.00	.00
39	221122	.00	.75	.00	.00
40	221212	.75	1.00	.00	.00
41	221221	.00	.00	.00	.00
42	222112	.00	.00	1.00	.00
43	222121	1.00	1.00	1.00	.00
44	222211	.00	.00	.00	.00
45	111222	.00	.00	.00	1.00
46	112122	.00	.00	.00	1.00
47	112212	.00	.00	1.00	1.00
48	112221	.00	.00	.00	.00
49	121122	.00	1.00	.25	1.00
50	121212	1.00	.75	1.00	1.00
51	121221	.00	.00	.00	.00
52	122112	.00	.00	1.00	1.00
53	122121	1.00	1.00	1.00	.00
54	122211	.00	.00	1.00	.00
55	211122	.00	1.00	.25	1.00
56	211212	1.00	.25	.00	1.00
57	211221	.00	.00	.00	.00
58	212112	.00	.25	1.00	1.00
59	212121	1.00	1.00	1.00	.00
60	212211	.00	.00	.00	.00
61	221112	1.00	.75	1.00	1.00
62	221121	.00	1.00	1.00	.00
63	221211	1.00	.50	.00	.00
64	222111	1.00	1.00	1.00	.00

many of the individual subjects than did the prototype model indicates that the prototype model is insufficient as a general model of classification.

We should emphasize, however, that our results do not rule out the possibility that at least some of the individual subjects classified on the basis of prototypes. In particular, if a pure prototype-abstraction process governed performance, it would not be clearly revealed by the results from this paradigm. The reason is that, for the present category structure, the exemplar similarity model can fit well data that are generated by a prototype process. To see why, note that the present structure has the property of high within-category similarity and low between-category similarity. Thus, each category occupies a tightly packed region in multidimensional similarity space, with training exemplars that are clustered around their prototype. Intuitively, it is easy to see that, for such a structure, comparing how similar an item is to the category prototypes will often be roughly equivalent to comparing the item's summed similarity to the individual exemplars. In previous research, more diagnostic experiments have been tested that could reveal a prototype-abstraction process if it were operating (e.g., Ashby & Maddox, 1992; Medin & Schaffer, 1978). The results from these experiments, however, have not provided evidence in favor of prototype abstraction either.

On the other hand, we should also emphasize that if the process envisioned in Minda and Smith's (2001) *mixture* model were operating, the present design would be well suited to reveal it, at least for certain parameter settings in that model. In general, if observers develop strong all-or-none memories for the training exemplars, while generalizing to new transfer items on the basis of similarity to the prototype, the patterns of data that would be created would *not* be well fit by the exemplar similarity model. To document this point, we conducted a *power analysis* involving the models. Specifically, following Nosofsky and Zaki (2002), we simulated how hypothetical subjects would classify stimuli, assuming that their behavior followed the MMPM. We then tested the ability of the competing models to fit these simulated data. In the simulations of the MMPM, we set the exemplar memory parameter at $e = 1$, set the guessing parameter at $g = 0$, and assumed a broad-based prototype-abstraction process in which equal attention weight was given to each dimension. For purposes of generality, we conducted three separate series of simulations, in which the sensitivity parameter was set at $c = 1$, $c = 3$, and $c = 6$. In each series of simulations, we generated 40 separate data sets in which each simulated observer classified each transfer item four times. We then fitted the GCM and the MMPM to these simulated data by using a maximum-likelihood criterion. The results are summarized in Table 7 in terms of the mean log-likelihood fits of the models and in terms of the number of individual cases (out of 40 total) in which the GCM provided a better fit to the data than did the MMPM. In a nutshell, the table reveals that if the MMPM actually governed performance, the GCM would

Table 7
Fits of the MMPM and the GCM to 40 Data Sets Constructed
by Simulating the MMPM

<i>c</i>	Mean $-\ln L$		<i>t</i> Statistic	No. of Cases in Which GCM Gave Better Fits
	MMPM	GCM		
1	64.9	80.3	$t(39) = 24.4$	0/40
3	58.5	66.7	$t(39) = 17.5$	0/40
6	45.9	48.9	$t(39) = 12.8$	2/40

Note—MMPM, multiplicative mixed prototype model; GCM, generalized context model; $-\ln L$, negative log-likelihood; *c*, overall sensitivity parameter.

provide relatively poor fits to the resulting data. Thus, the consistently superior fits of the GCM in the present paradigm constitute strong evidence *against* the processes formalized in the mixed prototype model of Smith and Minda (2000; Minda & Smith, 2001).

GENERAL DISCUSSION

In summary, in previous tests, Smith et al. (1997; Minda & Smith, 2001; Smith & Minda, 1998) argued that a mixed prototype model provided better accounts of individual subject classification performance than did an exemplar similarity model. However, Nosofsky and Zaki (2002) argued that there were important limitations of these tests. First, the exemplar model that Smith et al. fitted was constrained to assume a probability-matching response rule, whereas the prototype model they fitted made allowance for differing degrees of probabilistic versus deterministic responding. Second, these researchers tested the competing models on their ability to predict classification performance only on old training instances. However, a fundamental contrast between the models concerns their predictions of generalization to new transfer stimuli. The mixed prototype model assumes that generalization is based on similarity to the prototype, whereas the exemplar similarity model assumes that generalization is based on similarity to specific old exemplars. Nosofsky and Zaki showed that when the competing models were equated in terms of their response rules and tests of generalization of new transfer stimuli were included, the exemplar similarity model provided dramatically better accounts of individual subject classification performance than did the mixed prototype model.

A limitation of Nosofsky and Zaki's (2002) earlier research, however, is that they compared the models on their predictions of classification involving a nonlinearly separable category structure. In the present research, we extended the earlier findings by showing that they also hold in a situation involving a linearly separable category structure. Although the qualitative contrasts between the predictions from the models were not as strong in the present design, the complete set of transfer data nevertheless provided sufficient constraints to allow the models to be distinguished. In particular, we found a number of examples at the individual subject level in

which observers classified new transfer stimuli according to their similarity to specific old training exemplars, rather than according to their similarity to the category prototype.

Minda and Smith (2001) suggested that the category structure used in the present design should be conducive to prototype-based processing. It features a reasonably large category size, high-dimensional stimuli, and a coherent structure involving high within-category similarity and low between-category similarity. Thus, the present advantage displayed by the exemplar similarity model in its accounts of individual subject classification performance poses a significant challenge to the mixed prototype view of categorization.

A natural embellishment of the mixed prototype model that would enable it to account for the present results would be to make allowance for generalization processes involving the stored exemplars. In other words, one would abandon the assumption that the memory traces for the old exemplars are all-or-none and that they support classification judgments of old items only. Instead, new items could be classified in accord with their similarity to the old stored exemplars.

For example, one possible embellishment would involve the assumption that instead of storing whole exemplars in memory, observers store subpatterns of the exemplars. New items that perfectly matched a stored subpattern would be classified into the subpattern's category. The process that gives rise to such subpatterns would need to be specified, however, before one could conduct rigorous tests of such a model. In addition, it seems clear that such a model would have difficulty accounting for more graded forms of exemplar-based generalization, such as occurs in continuous-dimension stimulus domains (see Nosofsky, 2000, for a more extended discussion).

Regardless of the type of exemplar generalization mechanism that might be added to the mixed prototype model, however, the key question then becomes whether there is a *need* to posit a prototype-abstraction process at all. According to the exemplar similarity model, the category representation consists *only* of stored exemplars, and a process involving similarity comparisons with these exemplars is the sole basis for classification performance. The question is whether this exemplar similarity model is sufficient or needs to be extended with a prototype-abstraction process as well.

REFERENCES

- ASHBY, F. G., & GOTT, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *14*, 33-53.
- ASHBY, F. G., & MADDUX, W. T. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception & Performance*, *18*, 50-71.
- ASHBY, F. G., & MADDUX, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*, 372-400.

- ESTES, W. K. (1986). Array models for category learning. *Cognitive Psychology*, **18**, 500-549.
- HINTZMAN, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, **93**, 411-428.
- HOMA, D. (1984). On the nature of categories. *Psychology of Learning & Motivation*, **18**, 49-94.
- HOMA, D., STERLING, S., & TREPPEL, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 418-439.
- MADDOX, W. T., & ASHBY, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, **53**, 49-70.
- MCKINLEY, S. C., & NOSOFSKY, R. M. (1995). Investigations of exemplar and decision-bound models in large-size, ill-defined category structures. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 128-148.
- 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, & Cognition*, **10**, 333-352.
- MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, **5**, 207-238.
- MINDA, J. P., & SMITH, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 775-799.
- NOSOFSKY, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, & 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, & Cognition*, **13**, 87-108.
- NOSOFSKY, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 3-27.
- NOSOFSKY, R. M. (1992). Exemplars, prototypes, and similarity rules. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes* (pp. 149-167). Hillsdale, NJ: Erlbaum.
- NOSOFSKY, R. M. (2000). Exemplar representation without generalization? Comment on Smith and Minda's (2000) "Thirty categorization results in search of a model." *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1735-1743.
- NOSOFSKY, R. M., & JOHANSEN, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, **7**, 375-402.
- NOSOFSKY, R. M., PALMERI, T. J., & MCKINLEY, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, **101**, 53-79.
- NOSOFSKY, R. M., & ZAKI, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 924-940.
- POSNER, M. I., & KEELE, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, **77**, 353-363.
- REED, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, **3**, 382-407.
- SHEPARD, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.
- SMITH, J. D., & MINDA, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1411-1436.
- SMITH, J. D., & MINDA, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 3-27.
- SMITH, J. D., MURRAY, M. J., & MINDA, J. P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 659-680.
- WICKENS, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: Freeman.

NOTES

1. To see why, note that $s_{iA}^{\gamma} = [\exp(-c \cdot d_{iA})]^{\gamma} = \exp(-c \cdot \gamma \cdot d_{iA}) = \exp(-c^* \cdot d_{iA})$, where $c^* = c \cdot \gamma$. By contrast, because the GCM involves a summing of similarities in its response rule (Equation 1), the sensitivity and response-scaling parameters can be estimated separately in that model.

2. Nosofsky and Zaki (2002) found that a common strategy in this type of paradigm is for subjects to rely on the information provided by a single dimension as a basis for classification (i.e., a single-dimension "rule" strategy). (For similar results and a related discussion, see Nosofsky & Johansen, 2000, and Nosofsky, Palmeri, & McKinley, 1994.) This strategy, however, does not allow subjects to classify correctly all training exemplars in the present paradigm. The GCM and mixed prototype models can account equally well for this type of behavior by placing all attention weight on this single dimension and by setting the response scaling (or sensitivity) parameter at a sufficiently high level.

3. The MMPM fitted well Subject 3's classification of some of these critical transfer stimuli by placing almost all of its attention weight on Dimensions 4 and 5. However, with this attention configuration, the model then mispredicted Subject 3's classification of numerous other transfer stimuli.

(Manuscript received February 14, 2002;
revision accepted for publication May 9, 2002.)