How many exemplars are used? Explorations with the Rex Leopold I model

MAARTEN DE SCHRYVER, KATLEEN VANDIST, AND YVES ROSSEEL
Ghent University, Ghent, Belgium

The goal of this research is to test the hypothesis that a category is not necessarily represented by all observed exemplars, but by a reduced subset of these exemplars. To test this hypothesis, we made use of a study reported by Nosofsky, Clark, and Shin (1989), and replicated their Experiment 1 in order to gather individual-participant data. Both a full exemplar model and a reduced exemplar model were fit to the data. In general, the fits of the reduced exemplar model were superior to those of the full exemplar model. The results suggest that only a subset of exemplars may be sufficient for category representation.

A main topic in the categorization literature concerns the question of how a category is represented. According to prototype models, a category is represented by a single prototype. This prototype is an abstract summary of the category and reflects the central tendency of all the exemplars experienced. By contrast, according to exemplar models, a category is represented by a (full) set of observed specific exemplars (Hayes-Roth & Hayes-Roth, 1977; Medin & Schaffer, 1978; Nosofsky, 1986).

Most empirical studies that have been reported in the categorization literature show that exemplar models have been extremely successful (see Nosofsky, 1992, for an overview). Nevertheless, despite its success, the exemplar approach has been criticized by many authors (e.g., Maddox, 1999; Minda & Smith, 2001; J. D. Smith & Minda, 2002). Several issues have been raised, but the one we will focus on in this article concerns the ecological validity of a full exemplar model: Is it really plausible that all individual exemplars are stored in memory (Juslin & Persson, 2002; Maddox & Ashby, 1993; Nosofsky, Palmeri, & McKinley, 1994; E. E. Smith & Medin, 1981)? For example, do we actually build a representation of the category “dog” by storing a memory trace for every dog we have encountered in our life?

The central goal of this research is to test the hypothesis that a category is not necessarily represented by all observed exemplars, but by a subset of these exemplars. This hypothesis, although implicitly accepted by some researchers, has not been formally and empirically tested.

To make clear what we mean by a subset representation, consider the following example. Suppose that in a categorization experiment, a participant observes 10 exemplars (5 Category A exemplars, 5 Category B exemplars) in the training phase. Perhaps the participant uses a subset of, say, 3 Category A exemplars and a subset of 2 Category B

exemplars. In other words, only 5 out of 10 exemplars are actually used. On the other hand, another participant may use all 5 Category A exemplars and a subset of, say, 4 Category B exemplars. For this participant, 9 exemplars are used. Indeed, an important assumption of our work is that different participants may use different subsets. As a consequence, to test our hypothesis (that only a subset is used), we should avoid aggregating our data over participants, and focus on individual data sets.

Our starting point will be Experiment 1 in Nosofsky, Clark, and Shin (1989). This study was chosen because it has contributed significantly to the categorization literature, and is often cited as a “success story” for exemplar-based models. We will refit the original data of Nosofsky et al. (1989) using both a full exemplar model (the generalized context model, or GCM) and a reduced exemplar model (the Rex Leopold I, or Rex LI, model),¹ whereby the full set of exemplars is replaced by a smaller subset of these exemplars. In addition, we will fit our models to data that we gathered in an experiment with a design similar to the design of the Nosofsky et al. (1989) study. At this point, it is important to stress that in this article, no effort is made to explain why some exemplars are retained in the subset and others are not retained. All we want to establish is that the assumption that not all exemplars are used is a plausible one.

In what follows, we will first discuss the various formal models (the GCM and the Rex LI model). Next, we will summarize the Nosofsky et al. (1989) study and refit the original data, using the two models. Finally, we will describe and discuss our own experiment.

Overview of the Models

The generalized context model. According to the GCM (Nosofsky, 1984, 1986), the probability that stimulus i is classified in category C_A is given by

$$P(C_A | S_i) = \frac{b_A \sum_{j \in C_A} \eta_{ij}}{b_A \sum_{j \in C_A} \eta_{ij} + (1 - b_A) \sum_{j \in C_B} \eta_{ij}}, \quad (1)$$

where b_A ($0 \leq b_A \leq 1$) is the category C_A response bias and η_{ij} denotes the similarity between exemplars i and j . The similarity measure is assumed to be related to the psychological distance d_{ij} by

$$\eta_{ij} = \exp(-d_{ij}^q), \quad (2)$$

where $q = 1$ yields an exponential function and $q = 2$ yields a Gaussian function. In a two-dimensional space, the psychological distance between stimuli i and j is given by

$$d_{ij} = c \left[w_1 |x_{i1} - x_{j1}|^r + (1 - w_1) |x_{i2} - x_{j2}|^r \right]^{1/r}, \quad (3)$$

where x_{im} is the psychological value of exemplar i on dimension m . The parameter w_1 ($0 \leq w_1 \leq 1$) is the attention weight for Dimension 1. The parameter c ($0 \leq c < \infty$) is a sensitivity parameter reflecting overall discriminability in the psychological space. The exponent r defines the distance metric ($r = 1$, city-block metric; $r = 2$, Euclidean metric).²

The Rex Leopold I model. A model that can be used for testing the assumption that participants retain only a subset of exemplars is the Rex LI model, a variant of the reduced exemplar models (Rosseel, 2002). The Rex LI is designed to be identical to the GCM, with the exception that the full set of exemplars can be replaced by a reduced set of exemplars. The remaining exemplars form a true subset of the full set.

Importantly, the Rex LI model does not allow us to determine a priori which subset should be retained, nor does it explain why some exemplars are retained and others are not. Consequently, all possible subsets need to be tested. For a category containing N_k exemplars, there are $R_k = 2^{N_k} - 1$ possible subsets (note that the full set is considered as a possible subset). If there are two categories C_A and C_B containing N_A and N_B exemplars, respectively, $R_A \times R_B$ different models need to be tested. The exemplars used by the best-fitting model are assumed to be the exemplars that form the category representations. We immediately acknowledge a disadvantage of this procedure: Rex LI is presumably too flexible, and may very well overfit the data. Rex LI can freely choose which exemplars it keeps and which exemplars are removed. The GCM, on the other hand, can work only with the full set of exemplars. To temper the flexibility of the Rex LI model, we followed a conservative strategy for fitting and evaluating the models, as described in the next section.

Fitting the models to the data. As in Nosofsky et al. (1989), the weighted city-block metric is applied because separable-dimension stimuli were used in this experiment. Similarity is calculated by using the exponential function. When fitting the GCM, the sensitivity parameter c , the attention weight parameter w_1 , and the bias parameter b_A were freely estimated. A computer search was used to find

the parameters for each participant that maximized the log-likelihood function

$$L_{\text{test}} = \sum_i \ln N_i! - \sum_i \sum_k \ln f_{ik}! + \sum_i \sum_k f_{ik} \ln p_{ik}, \quad (4)$$

where N_i is the frequency with which stimulus i was presented and f_{ik} and p_{ik} are, respectively, the observed frequency and predicted probability with which stimulus i is classified in category k (Nosofsky et al., 1989).

When fitting the Rex LI model, all possible subsets need to be tested. The subset that maximizes the log-likelihood function is selected as the best-fitting variant of Rex LI. However, an important restriction is that we do not reestimate the free parameters (c , w_1 , and b_A) for every subset. This procedure would result in an overflexible model that almost surely would overfit the data. Instead, in the present approach, the best-fitting parameter values of the GCM are simply substituted into the Rex LI model. Effectively, the only difference between the GCM model and the selected variant of the Rex LI model will be the number of exemplars. The values of the free parameters are kept the same.

To test the models for generalizability, we followed a cross-validation procedure (Pitt, Kim, & Myung, 2003). In our own experiment, two different test sets were used, and hence two different log-likelihoods were computed. The log-likelihood L_{test1} was calculated by fitting the models to the data (i.e., the responses of the participants) obtained in the first test phase. The computation of the second log-likelihood L_{test2} was based on data obtained in the second test phase. Importantly, for both models, free parameters were kept fixed and the same (sub)sets of exemplars were used when computing the second log-likelihood L_{test2} .

The reasoning behind this cross-validation procedure is that more flexible models may overfit the data in the first test phase (Pitt et al., 2003). If we use only a single test phase to evaluate the quality of the model (as is the common approach in the categorization literature), we may tend to prefer the more flexible model. However, by using a second test phase, we expect that the overfitting model will be penalized, resulting in a fairer comparison between models.

Experiment 1 by Nosofsky et al. (1989)

In the categorization condition of Nosofsky et al.'s (1989) Experiment 1, 197 participants were tested. The stimuli were circles, varying in size and angle of orientation of a radial line. Seven stimuli were given during the training phase (3 were Category A exemplars and 4 were Category B exemplars). Sixteen stimuli (7 old and 9 new) were presented during the subsequent test phase (see Figure 1 for the design). Each of these 16 stimuli was presented five times during this test phase.

The GCM and rule-based model were fit to the data, using a maximum likelihood criterion. For the results, see Table 2 in the original article, but in short, the exemplar model fits were excellent. In order to obtain a better insight into the aggregated data, Nosofsky et al. (1989) reported some interesting analyses on the individual level. An individual classification partition was created for each

Figure 1. The category structure: Dimension “size” is on the horizontal axis, and dimension “angle” is on the vertical axis. Black circles are Category A training exemplars; gray circles are Category B training exemplars; white circles are the stimuli for the first test phase. The dots are the stimuli for the second test phase.

participant (see Figure 5 in the original article). A partition was composed as follows: Each new test exemplar that was classified as Category A three times or more was regarded as being a Category A exemplar. If not, it was regarded as a Category B exemplar.

Nosofsky et al. (1989) reported analyses showing that six out of the seven main observed partitions would have occurred with high probability if decisions had been made on the basis of similarity with all training exemplars. The authors concluded that most individuals’ behavior can be described accurately by the exemplar model. The model, however, also showed some inadequacies: Certain observed partitions were inconsistent with the exemplar model’s predictions.

These deviating partitions make the study attractive to replicate. We believe that the variation in the observed partitions may be due to a differing category representation among participants. It is unclear why the category representations of some individuals are different from those hypothesized by the exemplar model. In this article, we speculate that individual participants may have used a different subset of exemplars in order to represent the

categories. For example, whereas some participants may have used all four exemplars that belong to Category B, others may have used only two or three out of these four exemplars.

Refitting the models to Nosofsky et al.’s (1989) original data. The GCM and the Rex LI were fit to the data reported in Figure 4 in Nosofsky et al. (1989). The log-likelihood L is calculated by fitting the models to the data obtained in the test phase. The best-fitting parameters and fits are reported in Table 1.³

Apparently, according to Rex LI, all seven exemplars were used to represent the categories. No smaller subset could be retained. Following the assumption of Nosofsky et al. (1989), whereby the authors explored whether the aggregated data might have originated out of a confluence of different rules, we assume that the aggregated data might have arisen out of a confluence of different category representations. Therefore, we replicated the experiment to gather individual data.

METHOD

Participants

Five students from Ghent University participated in exchange for a small payment.

Materials

Stimuli were circles with a radial line and were presented on a 17-in. monitor with an 800×600 resolution. The circles varied in size and angle of orientation of the radial line. For both the training phase and the first test phase, the angles were 25° , 50° , 130° , and 155° . The lengths of the radii (size) were 25, 31, 44, and 50 mm. For the second test phase, on an 18×18 grid, 18 angles varied from 25° to 155° . The 18 sizes for this phase varied from 25 to 50 mm (see Figure 1).

Procedure

Participants were tested individually in a dimly lit room. The experiment consisted of four phases. In each phase, stimuli were presented until the participants responded by pressing X (for Category A) or N (for Category B) on the keyboard.⁴

In a first training phase, the 3 Category A stimuli (Exemplars 6, 7, and 11 in Figure 1) and the 4 Category B stimuli (Exemplars 1, 4, 9, and 14 in Figure 1) were presented in four blocks of 70 trials. In each block, the 7 training stimuli were presented 10 times in a random order. Feedback was given after each trial. In the first test phase, the 16 stimuli (7 old and 9 new) were presented in one block of 80 trials. In contrast to the training phase, no feedback was given. The training phase and the first test phase correspond to the procedure of Experiment 1 for the two categorization conditions as described by Nosofsky et al. (1989), except that in our test phase no feedback was given for old exemplars.

A second training phase served only as a recapitulation of the first training phase. One block of 70 trials of training stimuli was presented. Corrective feedback was given after each trial. Finally, in a second test phase, the 324 test stimuli were randomly presented in four blocks of 81 trials. No feedback was given.

RESULTS

Exemplars of the last training block were correctly classified in 98.57% (Participant 1), 97.14% (Participant 2), 97.14% (Participant 3), 97.14% (Participant 4), and 88.57% (Participant 5) of the cases. The frequency with which each stimulus of the first test set was classified in Category A for each of the 5 participants is presented in

Table 1
Maximum Likelihood Parameters and Fits to the
Data Reported in Nosofsky, Clark, and Shin (1989)
for the GCM and the Rex LI

Model	Best Subset	L	Parameter Values		
			w_1	c	b_A
GCM	full	-127.95	.64	0.77	.45
Rex LI	{1, 4, 6, 7, 9, 11, 14}	-127.95	.64	0.77	.45

Note—GCM, generalized context model; Rex LI, Rex Leopold I model; best model, the set of exemplars (see Figure 1) used by Rex LI; L , maximum log-likelihood.

Table 2
The Response Pattern for Each of the 5 Participants
in the First Test Phase

Stimulus	Participant				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	0	0	4
3	0	0	0	0	0
4	1	0	0	0	0
5	3	1	5	0	0
6	5	5	5	5	4
7	5	5	4	5	5
8	5	5	5	5	5
9	1	0	0	2	0
10	5	5	2	5	1
11	5	5	4	4	5
12	5	5	5	5	5
13	0	0	0	0	0
14	0	0	0	0	1
15	0	0	0	0	0
16	1	0	0	0	0

Note—The numbers are the frequencies with which each stimulus was classified in Category A.

Table 2. Note that participants continued to classify the original training exemplars (fairly) accurately during the first test phase. As in the Nosofsky et al. (1989) study, several different response patterns can be observed among the participants. Recall that a second training phase followed the first test phase. Exemplars of the second training phase were correctly classified in 98.57% (Participant 1), 100% (Participant 2), 100% (Participant 3), 97.14% (Participant 4), and 100% (Participant 5) of the cases. Finally, the results of the second test phase are depicted in Figure 2, where the response patterns of each participant are shown in a scatterplot. Note that the test stimuli that correspond to the training stimuli (the large circles) are correctly classified by all participants. For the other test stimuli, some rather large differences can be observed in the responses of the 5 participants, resulting in a challenging test set for our models.

When fitting the models to the data, the psychological space was assumed to reflect the physical two-dimensional space (see Nosofsky et al., 1989, Figure 2, panel A). The best-fitting parameters and fits of the individual and aggregated data are reported in Table 3. Also, for each participant, Table 3 gives the “best model,” or the subset of exemplars retained by the Rex LI model. For example,

the set of exemplars used to represent the category structures for Participant 2 is {1, 4, 7, 11, 14}. This means that Exemplars 7 and 11 (see Figure 1) are stored for the Category A representation, and Exemplars 1, 4, and 14 are stored for the Category B representation. For the first test set, the fits of the Rex LI model were better than those of the GCM, for each of the 5 participants. Of course, this result is not surprising because the Rex LI model is a much more flexible modeling approach. The real question is how well the models perform in the second test set. As can be observed in Table 3, the fits of Rex LI were better for 4 of the 5 participants, whereas the fits of both models for Participant 5 were very similar.

DISCUSSION

The results demonstrate that although a reduced set was used to classify new stimuli, the Rex LI model performed better than the GCM not only in the first test set, but also in the second test set. In particular, the reduced exemplar model did a better job describing the response patterns in the second test set (L_{test2}) of Participants 1, 2, and 4. For these 3 participants, a subset of exemplars seems clearly much more appropriate than the full set. For Participants 3 and 5, almost equal fits are obtained for both models, suggesting that a reduced set of exemplars may be sufficient, but not necessarily superior, in these cases. Recall that the values of the free parameters are the same in both models. The only difference between the GCM and the selected variant of the Rex LI model is the number of exemplars that are used to represent the categories.⁵

It is interesting to observe that the subsets selected by the Rex LI model contain a different selection of exemplars for each participant (only for Participants 2 and 4 is the same subset retained). This suggests that individual participants may have used a different subset of exemplars in order to represent the categories. These results strongly suggest that the variation in the partitions as observed in the Nosofsky et al. (1989) study may be due to a differing category representation among participants.

Finally, the results suggest that the aggregated data might have arisen out of a confluence of different category representations. When the Rex LI model was fitted to the aggregated data, the model retained a subset of exemplars that was observed for only 1 participant (Participant 1).

Table 3
Maximum Likelihood Parameters and Fits for the
GCM and the Rex LI Model in Experiment 1

Part.	Parameter Values			$G L_{\text{test1}}$	$G L_{\text{test2}}$	$R L_{\text{test1}}$	$R L_{\text{test2}}$	Best Subset
	w_1	c	b_A					
1	.27	0.16	.66	−15.84	−113.23	−13.10	−99.69	{1, 4, 7, 9, 11, 14}
2	.36	0.18	.41	−20.12	−126.83	−14.48	−84.21	{1, 4, 7, 11, 14}
3	.26	0.21	.39	−7.60	−83.53	−6.72	−80.09	{1, 4, 6, 9, 11, 14}
4	.26	0.12	.46	−25.46	−148.13	−17.32	−106.23	{1, 4, 7, 11, 14}
5	.51	0.13	.36	−25.54	−133.62	−24.17	−134.36	{1, 6, 9, 11, 14}
AGG	.34	0.14	.45	−62.96	−365.11	−53.39	−395.12	{1, 4, 7, 9, 11, 14}

Note—G, generalized context model; R, Rex Leopold I model; best subset, the set of exemplars (see Figure 1) used by the Rex LI model; L_{test1} (L_{test2}) = maximum log-likelihood for the first (second) test phase data; AGG, aggregated data.

Figure 2. For each participant, a scatterplot shows the response pattern in the second test phase. A small dot indicates a Category A response; “+” indicates a Category B response. Black circles are Category A training exemplars; gray circles are Category B training exemplars.

Although a better fit was obtained for the first test set, the model failed to generalize.

Recovery Study

By using cross-validation, we already tested our model for generalizability. In our experiment, the Rex LI model yielded better results for the second test phase for 4 participants. However, one may still question these results, because it is not unlikely that, despite the cross-validation procedure, the more complex model has the advantage over less complex models. As a consequence, it could be possible that Rex LI always outperforms the GCM simply because it is more flexible. To assess the inherent flexibility of the Rex LI model, we conducted a recovery study. A model recovery analysis can be used to verify that the procedure as used in the former analysis will choose the GCM when it is the actual generating model (Navarro, Pitt, & Myung, 2004). The purpose of this recovery study is to verify our cross-validation procedure. If the GCM is the “correct” model, the Rex LI model should not fit the data better than the GCM, especially the data generated for the second test phase.

An important issue in a recovery study concerns the choice of the parameter values that are used to generate the data under the true model. For our purposes, we decided to conduct the recovery study with five different sets of parameter values corresponding to the ones that we obtained in our experiment (see Table 4). First, we generated the model predictions, $P(C_A|S_i)$, under the GCM for both test phases. Using these predictions, 500 data sets were generated. Each data set consists of the number of Category A responses and the number of Category B responses for each of the test stimuli in both test phases (16 stimuli for Test 1, and 324 stimuli for Test 2). For each stimulus, the total number of responses equals $N = 5$ for the first test set and $N = 1$ for the second test set. For each data set, the Category A responses were randomly sampled from a binomial distribution of size N , using $P(C_A|S_i)$ as the probability of success. By using this binomial distribution, noise—typical of psychological data—was added to the data. The fitting process was identical to the one described above. Finally, for both test sets, we calculated the recovery rates. The recovery rate is defined as the proportion of data sets for which the correct model (GCM) fits the data best. Notice that the GCM is nested within the

Rex LI model. As a consequence, if the full set is retained by the Rex LI model, equal fits are obtained. In this case, we consider the GCM to be the best model.

We expect the recovery rates for the first test phase to be lower, due to the inherent flexibility of the Rex LI model (note that we have 105 variants to choose from and we pick only the best-fitting one). However, this flexibility should be penalized when calculating the second log-likelihood and should result in a much higher recovery rate for the GCM.

Table 4 contains the recovery rates for both test sets for the different sets of parameter values. Looking at the (total) recovery rates, it can be observed that the GCM recovered its own data in 35.7% of all Test 1 data sets and in 83.4% (min = 72.8%, max = 95.6%) of all cross-validation test sets. These results suggest that—at least for the design used in this article—if the GCM is the correct model, we can be fairly confident that our cross-validation procedure will choose the GCM as the best model.

GENERAL DISCUSSION AND CONCLUSION

The essence of exemplar theory is that a category is represented merely by its collection of exemplars. An exemplar simply refers to a specific observed instance, and importantly, no abstracted information or any type of (verbal) rules are involved in the category representation. However, a widely spread misconception is that exemplar theory insists on using all observed instances that have been identified as belonging to a specific category as an exemplar of this category. Indeed, this all-exemplars-are-used view of exemplar theory has been voiced by some authors. For example, the proximity model of Reed (1972) postulates that a category or concept is represented by all previously seen exemplars. On the other hand, other authors have suggested that only the best, most typical, or most frequent exemplars are used (E. E. Smith & Medin, 1981), or that most or many (but not necessarily all) exemplars are used (see Komatsu, 1992, for more references). In theory, we believe that few exemplar theorists would insist on adhering to the all-exemplars-are-used view. In practice, however, the all-exemplars-are-used principle is widely spread.

As soon as exemplars-based models are fitted to empirical data sets, they tend to include all exemplars, even if the number of exemplars is excessively high (e.g., 4,000 exemplars in McKinley & Nosofsky, 1995). We believe that this practice is largely responsible for the misconception that exemplar theory insists on the all-exemplars-are-used principle. Unfortunately, the not-all-exemplars-are-needed view has hardly received any attention in the categorization literature and has, as far as we know, never been explicitly tested.

Several (nonexemplar) models have been proposed to avoid the “memory load” problems that are often associated with exemplar models. Some examples are the rational model (Anderson, 1991), the striatal pattern classifier (Ashby & Waldron, 1999), and the mixture models of categorization (Rosseel, 2002). But few (if any) formal exemplar models have been described that explicitly avoid the usage of all exemplars. To fill this gap, we developed

Table 4
Recovery Rates (RRs)

Set of Parameters	Parameter Values			GCM vs. Rex LI Model	
	w_1	c	b_A	RR_{Fit1}	RR_{Fit2}
1	.27	0.16	.66	25.4	79.4
2	.36	0.18	.41	27.6	87.6
3	.26	0.21	.39	52.2	72.8
4	.26	0.12	.46	20.8	81.8
5	.51	0.13	.36	52.4	95.6
Total				35.7	83.4

Note— RR_{Fit1} (RR_{Fit2}) = recovery rates or the proportion of data sets for which the GCM is selected as the best-fitting model in the first (second) test set.

the reduced exemplar (Rex) family of models. The Rex models were first introduced in Rosseel (2002), where they were described as a specific implementation of a mixture model of categorization. Since then, they have grown to a larger family of models with one common theme: A category is represented not by a full set, but by a reduced set of exemplars. The several variants differ from each other in how they choose the (reduced) set of exemplars. The most basic model, Rex LI, is the one that we have employed in this article.

It is argued sometimes that the all-exemplars-are-stored principle is only problematic for categories with a large number of exemplars. After all, for categories with few exemplars, the memory load is rather limited or even negligible. However, the results of our experiment suggest that even for categories with only four or five exemplars, participants may not store the full set of exemplars.

A limitation of Rex LI is that it only allows us to determine post hoc which subset was retained. It does not explain why some exemplars are retained and others are not. Hence, in future research, the challenge will be to predetermine this subset of exemplars.

AUTHOR NOTE

Correspondence concerning this article should be addressed to M. De Schryver, Department of Data Analysis, Ghent University, B-9000 Ghent, Belgium (e-mail: maarten.deschryver@ugent.be).

REFERENCES

- ANDERSON, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, **98**, 409-429.
- ASHBY, F. [G.], & MADDOX, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, **37**, 372-400.
- ASHBY, F. G., & WALDRON, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin & Review*, **6**, 363-378.
- HAYES-ROTH, B., & HAYES-ROTH, F. (1977). Concept learning and the recognition and classification of exemplars. *Journal of Verbal Learning & Verbal Behavior*, **16**, 321-338.
- JUSLIN, P., & PERSSON, M. (2002). PROBABILITIES from EXemplars (PROBEX): A "lazy" algorithm for probabilistic inference from generic knowledge. *Cognitive Science*, **26**, 563-607.
- KOMATSU, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin*, **112**, 500-526.
- MADDOX, W. T. (1999). On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization. *Perception & Psychophysics*, **61**, 354-375.
- 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, ill-defined category structures. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 128-148.
- MEDIN, D. L., & SCHAFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 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.
- NAVARRO, D. J., PITT, M. A., & MYUNG, I. J. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology*, **49**, 47-84.
- 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. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multi-dimensional models of perception and cognition* (pp. 363-393). 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, & Cognition*, **15**, 282-304.
- 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 prototypes revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 924-940.
- PITT, M. A., KIM, W., & MYUNG, I. J. (2003). Flexibility versus generalizability in model selection. *Psychonomic Bulletin & Review*, **10**, 29-44.
- REED, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, **3**, 382-407.
- ROSSEEL, Y. (2002). Mixture models of categorization. *Journal of Mathematical Psychology*, **46**, 178-210.
- SMITH, E. E., & MEDIN, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- SMITH, J. D., & MINDA, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 800-811.

NOTES

1. The Rex (reduced exemplar) Leopold I model is the first member of a family of reduced exemplar models. The different variants have been named after the subsequent monarchs of the Kingdom of Belgium. Leopold I was the first king of the Belgians.
2. An important extension of the GCM was proposed by Ashby and Maddox (1993). By including a response-scaling parameter (γ), the GCM- γ can account for more deterministic response patterns. Such a deterministic response pattern is often observed for individual participant data (Ashby & Maddox, 1993; Maddox & Ashby, 1993; McKinley & Nosofsky, 1995; Nosofsky & Zaki, 2002). However, for simplicity, we decided not to include this parameter. When adding this extra parameter to the response rule of both models (GCM and Rex LI), no meaningful differences in the model fits were found when fitting the models to our data.
3. The fits obtained for the GCM are close but not identical to those reported in Nosofsky et al. (1989): We used the coordinates and the response probabilities as reported in Nosofsky et al.'s (1989) Table A2, Appendix A, and Figure 4, respectively. Rounding off errors may have caused small but negligible differences.
4. X and N were the category labels used in the experiment, whereas in this article we label them A and B.
5. If we allow all three parameters (c , w_1 , and b_A) to vary freely, the fits of the Rex LI model in the first test set are even more spectacular. Unfortunately, this is due to overfitting. The fits in the second test phase are extremely poor (at least for 2 of the 5 participants). This illustrates the fact that a modeling approach where we allow for a combination of (1) choosing any possible subset as a representation of the categories and (2) estimating a separate set of parameter values for each subset is too flexible.

(Manuscript received July 12, 2007;
revision accepted for publication October 17, 2008.)