Journal of Experimental Psychology:
Learning, Memory, and Cognition
1995, Vol. 21, No. 3, 548-568

Copyright 1995 by the American Psychological Association, Inc.
0278-7393/95/$3.00

Recognition Memory for Exceptions to the Category Rule

Thomas J. Palmeri and Robert M. Nosofsky
Indiana University

Experiments were conducted to demonstrate the utility of a rule-plus-exception model for extending current exemplar-based views of categorization and recognition memory. According to the model, exemplars that are exceptions to category rules have a special status in memory relative to other old items. In each of 4 experiments, participants first learned to categorize items organized into 2 ill-defined categories and then made old–new recognition judgments. Although the categories afforded no perfect rules, a variety of imperfect rules could be formed combined with memorization of exceptions to those rules. In each experiment, superior recognition of exceptions to imperfect logical rules was found. In addition, participants demonstrated better memory for old exemplars than new ones. A mixed model, which assumed a combination of rule-plus-exception processing and residual exemplar storage, provided good quantitative accounts of the data.

A common assumption underlying many modern models of classification learning is that a great deal of information about the originally presented exemplars is retained in the category representation. According to exemplar models, for example, category representations consist of the storage of all previously presented exemplars, and classification decisions are made by summing the similarity of an object to the stored exemplars of the alternative categories (e.g., Estes, 1986a; Heit, 1992; Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1986). Exemplar models have had success in accounting for numerous fundamental categorization phenomena (e.g., Estes, 1994; Medin & Florian, 1992; Nosofsky, 1992). Despite this success, it is reasonable to question the plausibility of exemplar-storage processes and the vast memory resources they seem to require.

An alternative model of category learning, much different in spirit from exemplar models, was recently proposed by Nosofsky, Palmeri, and McKinley (1994). These investigators formalized a rule-plus-exception (RULEX) model of category learning that follows in the spirit of classic hypothesis-testing models (e.g., Levine, 1975; Trabasso & Bower, 1968). According to RULEX, participants learn categories by forming logical rules over single dimensions or conjunctions of dimensions and then supplement these rules with occasional exceptions. In contrast to exemplar models, the category representations in RULEX contain relatively little information—just a simple rule or two, plus a few exceptions.

RULEX is limited in its development thus far to problems involving deterministic category assignments and stimuli varying along binary-valued dimensions. Nevertheless, within this domain, Nosofsky, Palmeri, et al. (1994) demonstrated that RULEX was capable of accounting for many fundamental categorization phenomena. These phenomena include prototype and specific exemplar effects (Medin & Schaffer, 1978), selective attention effects (Medin & Smith, 1981; Nosofsky, 1984), sensitivity to correlated dimensions (Medin, Altom, Edelson, & Freko, 1982), differential difficulty of learning linearly versus nonlinearly separable categories (Medin & Schwanenflugel, 1981), and the relative difficulty of learning various rule-described categories (Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994; Shepard, Hovland, & Jenkins, 1961). In addition, beyond simply accounting for averaged classification data, RULEX predicted fairly well the wide variety of patterns of generalization displayed by individual participants (Nosofsky, Palmeri, et al., 1994; Palmeri & Nosofsky, 1993).

The central goal of the present research was to provide further tests of this rule-plus-exception model by using it to predict patterns of old–new recognition data observed at the completion of category learning. Old–new recognition data are often used as a source of converging evidence for the types of representations that are formed when people learn categories (e.g., Estes, 1986b; Hayes-Roth & Hayes-Roth, 1977; Medin & Schaffer, 1978; Metcalfe & Fisher, 1986; Omohundro, 1981; Reitman & Bower, 1973). Indeed, previous work has demonstrated that, following category learning, people often have memories for old exemplars and that exemplar models provide excellent accounts of patterns of old–new recognition performance (Estes, 1994; Medin & Schaffer, 1978; Nosofsky, 1988, 1991). According to such models, recognition judgments are based on a measure of overall "familiarity" for an item, computed by summing the similarity of that item to all exemplars stored in memory (Gillund & Shiffrin, 1984; Hintzman, 1988; Nosofsky, 1988).

If all that is stored in the category representations is a rule and a few exceptions, then why do exemplar models yield good fits to old–new recognition data? One approach to answering this question is to admit that, even if the dominant strategy for

Editor's Note. Douglas L. Medin served as a Guest Action Editor in the handling of this article.

Thomas J. Palmeri and Robert M. Nosofsky, Department of Psychology, Indiana University.

This work was supported by Grant PHS R01 MH48494-04 from the National Institute of Mental Health. We thank Douglas Hintzman and Thomas Ward for their useful critiques of previous versions of this article. We also thank Jason Bechler, Lora Bigelow, Debra Coomes, Brenda Graves, Brian Krecik, and Lacinda Waldroup for testing participants.

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

solving classification problems is to form rules and exceptions, some participants may have memories for at least some of the exemplars presented during training. Such memories could be a natural byproduct of the processing that takes place when rules and exceptions are formed.

The approach we took in our research was to ask whether the predictions of exemplar models could be improved by considering the learning processes that are assumed in the RULEX model. Specifically, a natural prediction is that the exceptions to the category rule should have a special status in memory. We used this idea to develop a combined model, which posits that recognition decisions are based on summing similarities to stored exemplars, but where the exceptions are accorded a special weight when computing this summed similarity.

It is important to realize that sensitive procedures are needed to reveal the special role of the exceptions in memory. According to RULEX, people learn categories by forming rules and exceptions, but the particular rules and exceptions that are used will vary idiosyncratically across individual participants. The reason is that the learning process in RULEX is stochastic, and multiple dimensions and exceptions are available for solving most problems. Thus, in many cases, averaged categorization and recognition data will not reveal a special role for the exceptions because the particular items serving as the exceptions will differ across individual participants.

We used a variety of techniques in this research in an attempt to reveal the special role of the exceptions in participants' category representations. In each of the first three experiments, we used ill-defined category structures consisting of stimuli varying along highly separable, binary-valued dimensions. In these structures, no perfect single-dimension rule or conjunctive rule existed for defining category membership; however, a variety of imperfect rules were available that could be supplemented by exceptions. In the first experiment, participants were supplied with explicit rule-plus-exception instructions to control the strategy that was used for categorizing the items. We hypothesized that the exceptions to the supplied rule would be the best recognized items. Armed with knowledge of how these explicit rule-plus-exception instructions influence the types of representations that are formed, in the second experiment we examined old–new recognition following free-strategy category learning. Participants were first clustered into two groups on the basis of the types of rule-based generalizations they produced. We then examined differential recognition of individual items within each subgroup. If rule-plus-exception processes dominate during free-strategy conditions, then patterns of results within each subgroup should be similar to those observed in Experiment 1. In the third experiment we sought additional evidence for differential old–new recognition of exceptions under free-strategy conditions. We designed a category structure affording an extremely limited number of rules so that the exceptions could be determined beforehand, thus bypassing the need to examine different subgroups of participants. Finally, in Experiment 4, we extended the domain of inquiry from discrete, binary-valued stimuli to stimuli varying along fuzzy, continuous dimensions, again testing for differential recognition of exceptions to category rules. In all cases, we made use of explicit formal models to corroborate our interpretations of the use of exemplars in old–new recognition judgments, but where exceptions to the category rule have a special status in memory.

Models of Categorization and Recognition

In the following section we describe the formal models that guided our research. The exemplar model is Medin and Schaffer's (1978) well-known context model, which has had enormous success accounting for a variety of categorization and recognition data. The rule-plus-exception model that is used to supplement the context model's predictions of old–new recognition is the RULEX model proposed by Nosofsky, Palmeri, and McKinley (1994). At the outset, we restrict attention to stimuli varying along binary-valued dimensions.

Categorization

Context model. Categorization decisions in the context model are based on similarity relations among exemplars. The probability that a given stimulus S_i is classified into Category A is found by summing the similarity of S_i to all members of Category A and then dividing by the summed similarity of S_i to members of both Category A and Category B,

$$P(A|S_i) = \frac{\sum_{j \epsilon A} s_{ij}}{\sum_{j \epsilon A} s_{ij} + \sum_{j \epsilon B} s_{ij}}. \tag{1}$$

As formulated for discrete binary-valued dimensions, the similarity between exemplar S_i and S_j is given by the multiplicative rule,

$$s_{ij} = \prod_{m=1}^{M} s_m^{\delta_m(i,j)}, \tag{2}$$

where the s_ms are free parameters indicating the similarity of mismatches along dimension m, and $\delta_m(i, j)$ is an indicator function equal to 0 if stimuli S_i and S_j match along dimension m and set equal to 1 if they mismatch along dimension m. The similarity parameters, s_m, represent a combination of dimensional salience and selective attention (see Nosofsky, 1984, 1986). In general, when the physical dimensions are randomly assigned to each abstract dimension for every participant, the dimensional saliences can be assumed to be equal, hence, the similarity parameters reflect selective attention.

RULEX. The basis for category learning in RULEX is the acquisition of simple single-dimension rules or conjunctive rules supplemented by the partial storage of exceptions to those rules. One of the defining characteristics of RULEX is that the behavior of an individual participant is highly idiosyncratic. Different participants form different rules and remember different partial exceptions to those rules. The notion that much of category learning is based on the extraction of simple rules with the occasional storage of exceptions to those rules is not entirely new (cf. Martin & Caramazza, 1980; Medin, 1986; Medin, Wattenmaker, & Michalski, 1987; Ward & Scott, 1987). However, RULEX is the first such model to have explicitly been formulated and tested on a wide variety of psychological data (see Nosofsky, Palmeri, & McKinley, 1994, for additional details).

The learning process in RULEX works basically as follows. First, RULEX searches for a perfect single-dimension rule. A dimension is sampled and a single-dimension rule is formed. In the general version of the model, each dimension is sampled according to its intrinsic salience, W_i, where the saliences are free parameters. However, when physical dimensions are randomly assigned to abstract dimensions, we assume equal saliences (the only exception is Experiment 1 in which participants were supplied with a rule, causing that dimension to acquire a higher salience). If a rule fails it is discarded and a new dimension is sampled. If no dimension yields a perfect single-dimension rule (as was true in all experiments presented in this article), then RULEX searches for imperfect single-dimension rules. A dimension is selected and a single-dimension rule is formed. This rule is retained for a minimum number of trials (set equal to the number of training items by default). From this point on, the imperfect rule is retained only if performance exceeds a lax criterion (set to 60% correct by default). After a given number of trials (set equal to twice the number of training items by default) the rule is evaluated. If performance exceeds a strict criterion, *scrit* (which is a free parameter), the rule is permanently stored, otherwise it is discarded and another dimension is selected. If all dimensions have been sampled, RULEX searches for conjunctive rules by using a similar process.

After a single-dimension rule or conjunctive rule has permanently been stored, RULEX begins the exception storage stage. If an item is encountered that contradicts the rule, then RULEX attempts to store that exception. Each dimension of the item is probabilistically sampled with probability *pstor* (which is a free parameter); the dimension(s) that were part of the failed rule are sampled with probability one. Any dimension not sampled is stored as a "wildcard" (*) that can match any value. Storage of the exception is also probabilistic, with P(successful storage) = $pstor^N$, where N is the number of sampled dimensions. For example, given the stimulus structure shown in Table 1, suppose a rule has been formed along dimension 1 such that value 1 on dimension 1 signals an A and value 2 signals a B. When A5 (2111) is encountered, an error occurs, and an attempt is made to store this item as an exception. Dimension 1 is sampled with probability one, and dimensions 2–4 are each sampled with probability *pstor.*

During the categorization decision process, an item is first matched with all exceptions that have been stored. For example, 2111 and 2112 match the exception 21**. If an item matches more than one exception, then a response is made probabilistically, depending on how many of those exceptions signal Category A or Category B. If an exception causes an error, it is removed from memory. If none of the exceptions match the item, then the rule is applied.

Table 1
Category Structure Used in Experiments 1 and 2

Category and item	Dimension values
Category A	
A1	1112
A2	1212
A3	1211
A4	1121
A5	2111
Category B	
B1	1122
B2	2112
B3	2221
B4	2222
Transfer item	
T1	1221
T2	1222
T3	1111
T4	2212
T5	2121
T6	2211
T7	2122

Note. Based on Medin and Schaffer (1978).

Recognition

Context model. According to the context model, whereas categorization decisions are based on the relative summed similarity of an item to exemplars of each category, recognition decisions are based on the absolute summed similarity of an item to exemplars of both categories (Gillund & Shiffrin, 1984; Hintzman, 1986, 1988; Nosofsky, 1988, 1991). The overall summed similarity, or familiarity, F_i, of each stimulus, S_i, is given by

$$F_i = \sum_{j \epsilon A} s_{ij} + \sum_{j \epsilon B} s_{ij}, \tag{3}$$

where the similarities are defined in Equation 2. Recognition judgments are assumed to be a monotonically increasing function of this summed similarity.

RULEX. In an extreme version of RULEX, recognition decisions are only based on the partial exceptions. Note that recognition judgments may vary among individual participants because different partial exceptions have been stored. The familiarity of a stimulus, S_i, is based on its summed similarity to each of the exceptions, X_j, and is given by

$$F_i = \sum_{j \epsilon Exc} s_{ij}. \tag{4}$$

The similarity between a stimulus, S_i, and an exception, X_j, is given by

$$s_{ij} = \prod_{m=1}^{M} \theta_m(i,j), \tag{5}$$

where

$$\theta_m(i,j) = \begin{cases} s_w & \text{if } X_j \text{ contains a wildcard on dimension } m \\ s_s & \text{if } S_i \text{ mismatches } X_j \text{ on dimension } m \\ 1 & \text{if } S_i \text{ matches } X_j \text{ on dimension } m, \end{cases} \tag{6}$$

where s_w and s_s are free parameters. s_w reflects the similarity of a value to a wildcard, whereas s_s reflects the similarity of mismatching values.[1]

Combined model. In the combined model, we assume that recognition decisions are primarily based on the exceptions, but that there is also residual memory for old items. We formulated a model in which the summed similarity was a combination of summed similarities to stored exceptions and residual summed similarities to all old exemplars. The familiarity of a stimulus S_i was defined by

$$F_i = \omega F_i^X + (1 - \omega)F_i^R, \quad (7)$$

where F_i^X is the summed similarity of S_i to the exceptions, as defined above for RULEX, and F_i^R is the residual summed similarity to all exemplars, as defined above for the context model. Because exceptions are presumably more strongly encoded, they are weighted by ω. A four-parameter model was formalized with parameters s_s, s_w, s, and ω, where s is the residual exemplar-similarity parameter in Equation 2. Setting ω equal to zero results in a pure version of the context model, in which $s_w = s$ for all dimensions m. Setting ω equal to one results in a pure version of RULEX.

Experiment 1

In the first experiment we supplied participants with explicit rule-plus-exception instructions to control as carefully as possible the type of strategy that was used during category learning. The goal was to investigate the pattern of recognition data that would be observed when rule-plus-exception strategies are used for categorization. We hypothesized that under such conditions the exceptions to the supplied rule would be the best recognized items. After verifying and modeling such a pattern of results, the next step would be to search for evidence of rule-plus-exception processes under free-strategy conditions.

Experiment 1 was a partial replication and extension of a study by Medin and Smith (1981). These researchers found that the context model could adequately account for categorization under conditions of explicit rule-plus-exception instructions and free-strategy instructions. Therefore, they argued, an account of categorization under different strategies could be made within the framework of a single process model (see also Medin, 1986). The effect of different strategies was merely to alter the amount of information stored about each attribute. However, Medin and Smith did not compare the predictions of the context model with those of a rule-plus-exception model. Thus, a subsidiary goal of this experiment was to compare the predictions of RULEX and the context model with regards to categorization judgments under conditions in which explicit rule-plus-exception instructions were provided.

Method

Participants. The participants were 58 undergraduate and graduate students from Indiana University who were paid $5.00 for their participation. All participants were individually tested.

Stimuli. The stimuli were computer-generated line drawings of rocketships varying along four binary-valued dimensions: shape of wing (triangular or rectangular), nose (staircase or half circle), porthole (circular or star), and tail (jagged or boxed) (modeled after stimuli used by Hoffman and Ziessler, 1983). The abstract category structure is shown in Table 1 (Medin & Schaffer, 1978). The assignment of physical dimension to abstract dimension was randomized for every participant, as was the assignment of physical values along a dimension to abstract values (1 or 2). In all experiments, stimulus presentation and response recording were controlled by IBM-compatible personal computers.

Procedure. Participants were instructed to follow a rule-plus-exception strategy at the start of the experiment. A sample stimulus was shown on the screen with one of the four dimensions highlighted. Instructions read as follows:

> We want you to use a particular strategy to learn to classify the rocketships. You might call it a "rule-plus-exception strategy." First, pay attention to the part of the rocketship which is highlighted on the display in front of you. During the experiment, we want you to pay attention to this dimension and learn which shape of this part is associated with rocketships from Planet A and which shape is associated with rocketships from Planet B. You will find a rocketship from each planet that is an exception to this rule. Memorize these rocketships. When you have mastered the task, you will be doing something like looking to see if the rocketship is one of the exceptions; if so, make the memorized response; if not, apply the rule.

For every participant, the "rule" dimension corresponded to abstract dimension 1 in Table 1. Hence, for every participant, A5 and B1 were the exceptions. Because the assignment of physical dimensions to abstract dimensions was randomized for every participant, the particular physical dimension defined as the rule was different for every participant. After the participants had read the instructions they were again explicitly told by the experimenter which dimension to form a rule along and that there would be a couple of exceptions to this rule.

There were 16 blocks of training trials. Each of the nine training stimuli, A1–A5 and B1–B4, was presented once per block. The presentation order was randomized for each subject. On each trial, the participant was presented with a rocketship and judged if it was from Planet A or Planet B. Responses were made by pressing one of two labeled keys on the computer keyboard. After a participant responded, corrective feedback was provided for 2 s. After an interval of 1 s, the next stimulus was shown. In addition, during the first two blocks of training, only the stimuli that followed the rule were displayed. This procedure was used to avoid the possibility of having one of the exceptions appear very early, causing participants to abandon the strategy they were explicitly told to use. This aspect of our procedure differs from Medin and Smith (1981).

Following training, participants completed a transfer phase in which they categorized and made old–new recognition judgments about all 16 stimuli. On each trial, the participant first judged if the rocketship was from Planet A or Planet B and then judged whether the rocketship was old or new. No corrective feedback was provided after either response. Each of the 16 stimuli was presented once per block for a total of three blocks. Participants were told to judge a rocketship as old only if it had been seen during training.

[1] In a complete version of the model, the actual features making up the rule would also be represented in memory. So, if a new feature value were presented along the dimension forming the rule, the participant should immediately recognize it as a new item. Such manipulations were not carried out in the experiments reported in this article.

Participants also completed five blocks of a speeded categorization task in which they categorized each of the 16 stimuli as quickly as possible without sacrificing accuracy. On each trial, a small crosshairs appeared on the screen for 500 ms to alert the participant to the next stimulus. Participants were instructed to keep one finger from each hand on the response keys at all times during this phase. Feedback was supplied only for old stimuli originally seen during the training phase.

Results

Categorization and recognition data. A fairly high criterion was set to ensure that only those participants who followed the rule-plus-exception strategy were included in the analyses. A reasonable assumption is that most of the participants who did poorly in the experiment were those who either did not follow the instructions or abandoned the explicit strategy that was given to them. Such participants might have adopted different rules or might have tried to memorize all of the items. Including them would have disrupted our goal of studying old–new recognition under highly controlled conditions deemed to promote the rule-plus-exception strategy. Participants making six or more errors in the last four blocks of training (36 trials) were excluded. A total of 35 participants (60%) satisfied the criterion for inclusion in the study.

The observed categorization data obtained during the transfer phase are shown in Table 2 as the probability of classifying each of the 16 rocketships as a member of Category A. Observe that T1–T3 each have value 1 on dimension 1 and T4–T7 each have value 2 on dimension 1. As shown in Table 2, there was a tendency to classify transfer items T1–T3 into Category A and T4–T7 into Category B. Thus, the pattern of categorization for the new transfer items is consistent with the use of the rule provided. Moreover, this rule-described tendency was weakest for T2, T5, and T6, which are each very similar to one of the exceptions, A5 or B1. For example, T2 (1222) can be classified according to the rule as a member of Category A or according to the similar stored exception, B1 (1122), as a member of Category B. Finally, as might be expected, more errors were made on the exceptions, A5 and B1, than were made on the other training items. In summary, the overall pattern of categorization data is consistent with the use of the rule-plus-exception strategy that participants were instructed to use, a point that we corroborate in the *Categorization theoretical analysis* section.

The observed recognition probabilities are shown in Figure 1. Although overall recognition performance was not very good, the training items, on average, were recognized as old with higher probability than the new transfer items, $t(34) = 4.40, p < .001$. Furthermore, if one excludes the exceptions from the analysis, the remaining old items still had higher recognition probabilities than the new items, $t(34) = 2.45, p < .01$. Of particular interest, the exceptions, A5 and B1, were recognized as old with higher probability than any other item (all relevant pairwise t tests were significant, $t(34) > 1.83$, $p < .05$). New items that were similar to the exceptions also tended to have relatively high recognition probabilities.

Table 2
Categorization Response Probabilities P(A) Observed and Predicted by RULEX and the Context Model in Experiment 1

Stimulus				
Category and item	Dimension values	Observed	RULEX	Context model
Category A				
A1	1112	.943	.972	.943
A2	1212	1.000	.983	.996
A3	1211	.971	.988	.999
A4	1121	.981	.972	.878
A5	2111	.924	.932	.872
Category B				
B1	1122	.133	.129	.176
B2	2112	.057	.032	.123
B3	2221	.023	.014	.004
B4	2222	.019	.012	.001
Transfer item				
T1	1221	.943	.934	.935
T2	1222	.686	.654	.553
T3	1111	.943	.983	.969
T4	2212	.029	.061	.065
T5	2121	.143	.255	.415
T6	2211	.324	.302	.461
T7	2122	.076	.019	.059

Note. RULEX = rule-plus-exception model.

The median response times for each item in the speeded categorization phase are shown in Figure 2. The most striking finding was that the exceptions, A5 and B1, were classified several hundred milliseconds slower than any other item (all relevant pairwise Wilcoxon tests were significant, $z > 4.77$, $p < .01$). We discuss the interpretations of these results following presentation of the formal modeling analyses.

Categorization theoretical analysis. Our first goal was to corroborate that participants followed our instructions and used a rule-plus-exception strategy by comparing the predictions of RULEX and the context model with respect to the categorization judgments. We first fitted a version of RULEX with three free parameters to the categorization data. The parameters were the criterion for accepting single-dimension rules, *scrit*, the exception storage probability, *pstor*, and a salience parameter for the rule dimension, W_1 (all other parameters were set at their default values; see Nosofsky, Palmeri, & McKinley, 1994). We modeled the explicit rule instructions by assuming a high salience along the dimension on which the rule was provided, W_1. Because RULEX is inherently stochastic, all of the fits to the categorization data reported in this article were averaged over Monte Carlo simulations of 5,000 individual runs. The best fitting parameters were found by conducting an extensive grid search using summed squared error (*SSE*) as the measure of fit. The predicted categorization response probabilities are shown in Table 2. RULEX provided an excellent fit to the average classification data, $SSE = 0.022$, accounting for 99% of the variance. The best fitting parameters were $scrit = 0.72, pstor = 0.67$, and $W_1 = 0.89$ (with $W_2 = W_3 = W_4 = 0.037$; the W_is sum to 1.0).

The excellent fit of RULEX contrasts with a relatively poor fit of the context model. A version of the context model with four free similarity parameters was fitted to the categorization

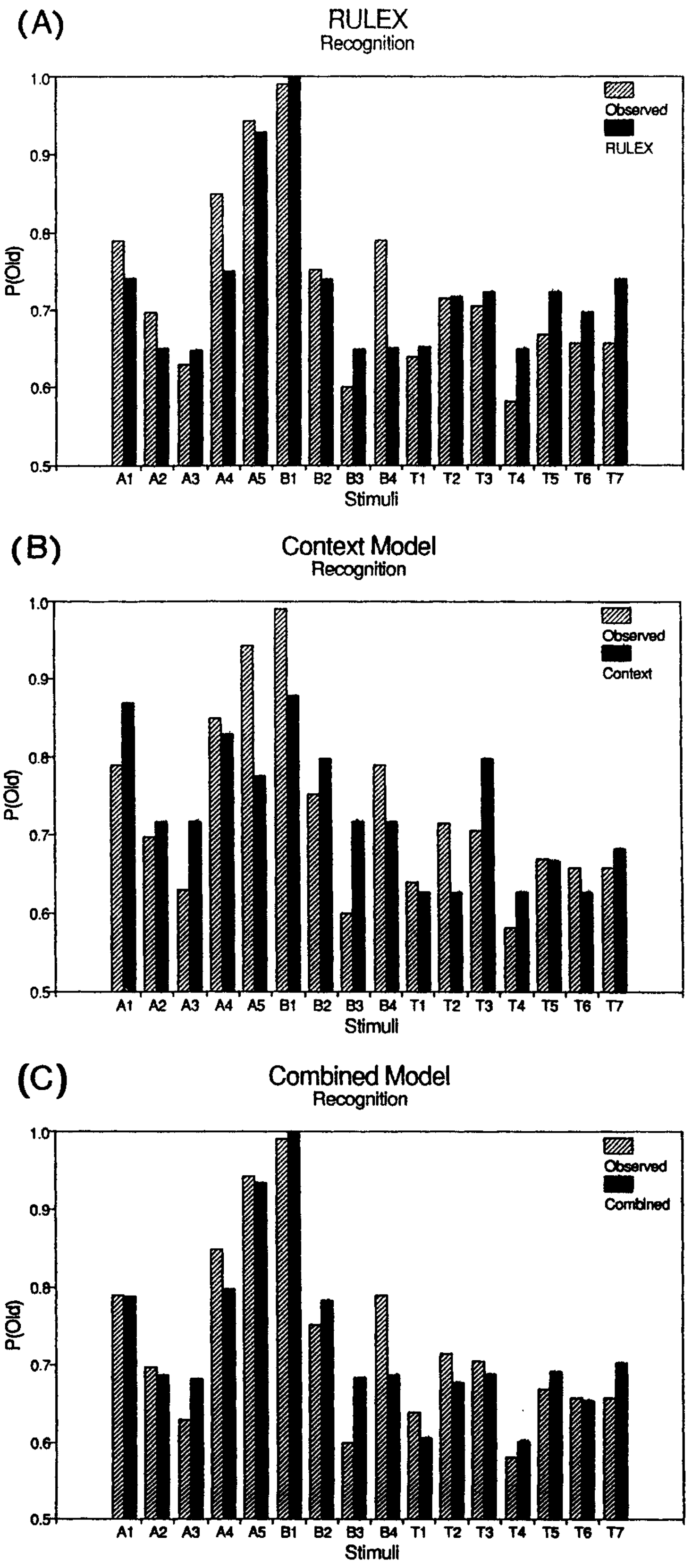

Figure 1. Old–new recognition probabilities observed and predicted in Experiment 1. (A): RULEX (rule-plus-exception model), (B): context model, and (C): combined model.

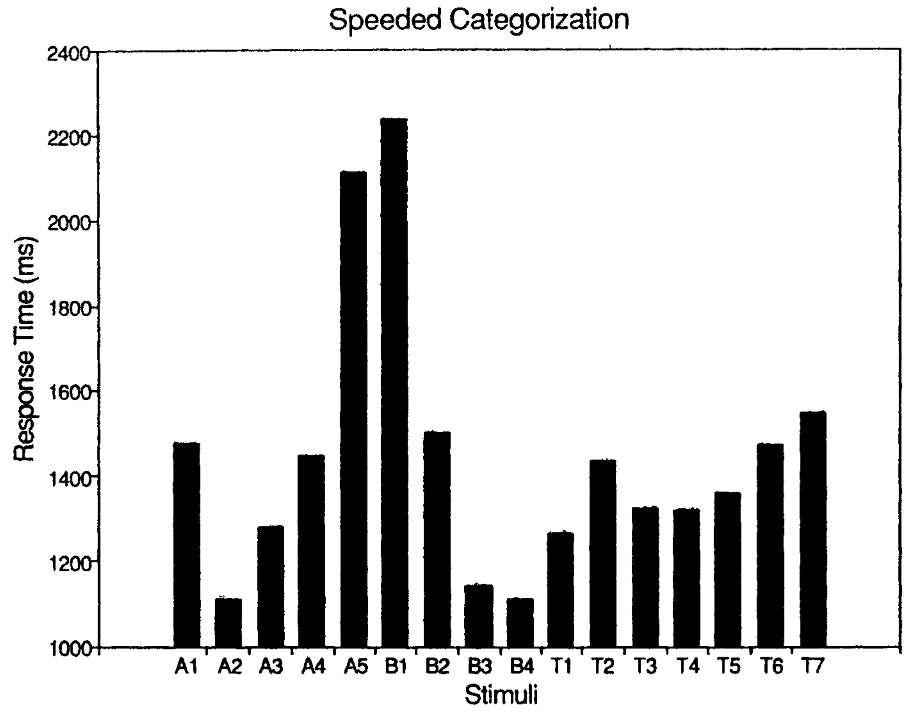

Figure 2. Median speeded categorization response times in Experiment 1.

data by using a hill-climbing algorithm that minimized the summed squared error. The results, shown in column 3 of Table 2, yielded an *SSE* of 0.134, with best fitting similarity parameters of $s_1 = 0.000, s_2 = 0.074, s_3 = 0.066$, and $s_4 = 0.141$. This fit is appreciably worse than that of RULEX, even though the context model has an additional free parameter.

Recognition theoretical analysis. Having obtained evidence that participants in this experiment followed the explicit instructions provided to them, we now turn to the real issue of interest, namely the effect that the rule-plus-exception strategy had on old–new recognition. To address this issue, we fitted versions of RULEX, the context model, and the combined model to the recognition probabilities shown in Figure 1. In all of the fits reported, we found parameters that maximized the linear correlation between the observed recognition probabilities (or ratings) and the predicted summed similarities.[2]

We first fitted the strict RULEX model to the recognition data. In all fits involving RULEX, the parameters *scrit, pstor,* and W_i that best fitted the categorization data were held fixed; only the similarity parameters s_s and s_w were allowed to vary. Each simulation yielded a unique set of rules and exceptions for a hypothetical participant, which was used to generate that participant's summed similarities for each item. The predictions for the individual simulated runs were then averaged to predict the group data. The predictions are shown in Figure 1A along with the observed recognition probabilities. The correlation between the observed and predicted recognition probabilities was $r = .865$. Although this strict RULEX model successfully predicted superior recognition memory for the exceptions, A5 and B1, it had difficulty accounting for the residual recognition of old rule-following items relative to new transfer items; the model overpredicted the recognition of many new items and underpredicted the recognition of many old items. Residual memory for old exemplars was found, even under conditions involving explicit rule-plus-exception instructions.

We next fitted a four-parameter version of the context model to the recognition probabilities. The results are shown in Figure 1B. The correlation between the observed and predicted recognition probabilities was relatively poor, $r = .734$. Although the context model correctly predicted better overall recognition of the old items relative to the new items, it failed to predict many important trends in the data, especially regarding recognition of the exceptions. Foremost, the context model failed to predict high recognition of one of the exceptions, A5; contrary to the data, three old items (A1, A4, and B2) and one new item (T3) were predicted to be better recognized than A5. In fact, as shown in Appendix A, the standard context model cannot predict high recognition of both A5 and B1 compared with A4 and B2, regardless of the values of the similarity parameters. The qualitative pattern of results in this experiment rules out an explanation of old–new recognition in terms solely of the exemplar storage processes allowed in the standard context model.

Taken together, these modeling results suggest that although the exceptions do have a special status in memory, old

[2] The assumption of a linear relation between recognition probabilities (or ratings—see Experiments 3 and 4) and summed similarities is made for simplicity. Nonlinear squashing functions could yield slight improvements in quantitative fit but would not change the qualitative predictions of the models. In every case, for clarity and ease of exposition, we display the predicted recognition probabilities (or ratings) found by regressing the predicted summed similarities onto the observed recognition probabilities (or ratings).

exemplars were still remembered. Hence, we tested a combined model in which recognition decisions were based on similarity to stored exceptions and old exemplars (Equation 7). The best fitting parameters yielded a correlation of $r = .925$, with $s_s = 0.329$, $s_w = 0.050$, $s = 0.000$, and $\omega = 0.889$. The predicted recognition probabilities of this combined model are shown in Figure 1C. The addition of residual memory for old exemplars increased the familiarity of old items relative to new items, but the exceptions were more heavily weighted, as indicated by the high value of ω.[3]

Overall, the combined RULEX model does an excellent job of predicting the overall pattern of recognition data (its only obvious failings were moderate mispredictions of B3 and B4). Furthermore, for RULEX, the fit to the recognition data was entirely dependent on the categorization simulation to determine the rules and exceptions formed—even more impressive fits could be achieved if the categorization data and the recognition data were fitted simultaneously.

Discussion

This experiment provided support for our hypothesis that, if people are instructed to use a rule-plus-exception strategy during learning, then the exceptions to the supplied rule would be the best recognized items. The theoretical analyses revealed that the exceptions to category rules have a special status in memory relative to other category exemplars. A hybrid representational model was required in which recognition decisions were a function of strongly weighted exceptions plus some residual memory for the remaining old exemplars.

Further evidence for the special nature of exceptions was supplied by the speeded phase in which the exceptions were categorized several hundred milliseconds slower than any other item (see Ward and Scott, 1987, for a similar result). Although there does not exist a response time model of categorization that can formally be applied to these data, the results are suggestive that something special may be occurring when people encounter the exceptions. For example, every dimension of an exception must be verified. By contrast, for the remaining items, few dimensions need to be checked because once a single mismatch is found, the rule can be applied. This pattern of results is consistent, therefore, with a rule-plus-exception strategy involving some form of limited capacity, self-terminating comparison process. Clearly, however, further theoretical and empirical work is needed to test other potential response time models.

Although the exceptions were more strongly encoded in memory, there was also residual memory for the nonexception old items. This result is extremely important because it indicates that even under extreme conditions involving explicit rule-plus-exception instruction, there is still some memory for old exemplars. Thus, even if rule-plus-exception processes operate in free-strategy conditions, we should still expect to see evidence of memories for old exemplars.

Residual memory for nonexception old items is consistent with the processing assumptions of RULEX. Before a rule can be applied, a decision must be made about whether the item is an exception. This decision requires the participant to check other dimensions besides the rule dimension. Whereas the exceptions are encoded by an active memorization process, the nonexceptions may be encoded as a byproduct of this exception-verification process. Thus, the nonexceptions will be remembered, but not to the same extent as the exceptions.

Experiment 2

Armed with the knowledge of what to expect when participants are instructed to use a rule-plus-exception strategy, we now looked for evidence of rule-plus-exception processes during free-strategy conditions. In Experiment 2, we provided participants with standard, free-strategy instructions, replicating the classic Medin and Schaffer (1978) experiment. Nosofsky, Palmeri, and McKinley (1994) showed that RULEX successfully predicted the averaged categorization data obtained in this experimental paradigm. Furthermore, beyond predicting averaged transfer data, RULEX provided a fairly good account of the distribution of generalization patterns observed at the individual participant level. The novel contribution of the present experiment involved collecting recognition judgments after category learning to test for superior recognition of the exceptions relative to the other items.

Recall that a crucial assumption underlying RULEX is that different participants form different rules and store different exceptions. According to RULEX, average categorization data for the structure shown in Table 1 can be accounted for by assuming most participants form rules along dimension 1 or dimension 3, with partial exceptions A5 and B1 or partial exceptions A4 and B2, respectively (along with smaller proportions of rules along dimension 4 and a variety of conjunctive rules). Such a mixture of different strategies makes it difficult to examine differential recognition of the exceptions directly because average data obscure any difference in recognition that may be present. Suppose A5 and B1 are the best recognized items for one group of participants (with A4 and B2 poorly recognized), and A4 and B2 are the best recognized items for another group of participants (with A5 and B1 poorly recognized). Averaging across these two sets of participants masks the differences that are present.

Therefore, the idea in this experiment was to first select those participants who most likely formed single-dimension rules along dimensions 1 and 3 and then to examine differential recognition and speeded categorization of the exceptions within these subgroups. As described below in more detail, participants were partitioned into groups on the basis of the pattern of categorization responses made during the transfer phase. For example, a participant who classified T1–T3 into Category A and T4–T7 into Category B likely formed a rule that value 1 on dimension 1 signals Category A. Participants who showed this pattern of dimension 1 generalizations, as well as other patterns, were grouped together for analysis; we

[3] The model fits suggest that these residual exemplar memories were highly distinctive, as indicated by $s = 0.000$. A result for which we have no explanation is why the estimated value of s_w, which measures the similarity between a dimension value and a wildcard, is less than that of s_s, which measures the similarity between mismatching dimension values. The fits are not very sensitive to the precise values of s_w, however, so these results should not be emphasized.

then tested for differential recognition of A5 and B1. A similar subgroup of participants was formed for dimension 3 generalizations, testing for differential recognition of A4 and B2.

Table 3
Categorization Response Probabilities P(A) for All Participants and Those Showing Dimension 3 Generalizations and Dimension 1 Generalizations in Experiment 2

Stimulus				
Category and item	Dimension value	Overall	Dimension 3	Dimension 1
Category A				
A1	1112	.811	.787	.894
A2	1212	.838	.870	.887
A3	1211	.856	.917	.943
A4	1121	.704	.685	.816
A5	2111	.722	.898	.560
Category B				
B1	1122	.322	.157	.532
B2	2112	.306	.500	.106
B3	2221	.200	.167	.149
B4	2222	.113	.056	.071
Transfer item				
T1	1221	.630	.250	.943
T2	1222	.379	.093	.759
T3	1111	.846	.917	.943
T4	2212	.337	.787	.078
T5	2121	.318	.361	.326
T6	2211	.589	.907	.255
T7	2122	.192	.167	.113

Method

Participants. Participants were 198 undergraduates from Indiana University who received partial course credit in an introductory psychology course for their participation.

Stimuli. The stimuli and category structure were the same as those used in Experiment 1.

Procedure. Participants received standard, free-strategy instructions. There were 25 blocks of training trials. Each of the nine training stimuli, A1–A5 and B1–B4, was presented once per block. On each trial, a participant was presented with a stimulus, judged if it was a member of Category A or Category B, and received corrective feedback for 2 s. The next stimulus was shown after an interval of 500 ms. Training ended if the participant completed four consecutive error-free blocks.

Following training, three blocks of transfer-recognition trials were presented. On each trial, a participant first categorized an item and then judged it as new or old. Recognition judgments were made by using a confidence rating scale with responses ranging between (1) *absolutely sure new* and (8) *absolutely sure old.* Responses were made by pressing one of eight labeled numeric keys at the top of a computer keyboard. No corrective feedback was provided after either response.

Following the transfer phase, participants were given a speeded categorization task identical to that used in Experiment 1.

Results

Overall categorization. The categorization data obtained during the transfer phase are shown in Table 3 as the probability that each item was classified as a member of Category A. Notice that most errors were made on A5 and B1, and A4 and B2, the exceptions to rules along dimensions 1 and 3, respectively. The distribution of generalization patterns underlying these average transfer data is shown in Figure 3. As introduced in Nosofsky, Palmeri, and McKinley (1994; see also Nosofsky, Clark, & Shin, 1989; Pavel, Gluck, & Henkle, 1988), we defined a pattern of generalization for an individual participant as the pattern of responses given to each new transfer item. For example, if a participant classified T1–T3 in Category A, and T4–T7 in Category B, this defined the generalization pattern AAABBBB. Because there were seven transfer items and two categories, there was a total of $2^7 = 128$ possible patterns.

Three prominent generalization patterns were observed, along with a number of other patterns. Pattern AAABBBB is consistent with a single-dimension rule along dimension 1, and pattern BBAABAB is consistent with a single-dimension rule along dimension 3. Pattern ABABBAB is consistent with rules along both dimension 1 and dimension 3, as well as being consistent with pure exemplar storage.[4] Overall, the distribution of generalization patterns is similar to the distribution observed by Nosofsky, Palmeri, and McKinley (1994).

Other generalization patterns, besides the prominent ones listed above, are consistent with rules along dimensions 1 or 3. For example, pattern ABABBBB is consistent with a rule on dimension 1 along with the exception 1*22 → B. Pattern BAAABAB is consistent with a rule on dimension 3 along with the exception 1*2* → A. We conducted a simulation of RULEX by using standard parameter settings (see Nosofsky et al., 1994) to determine which patterns of generalization were consistent with rules along dimension 1 or dimension 3, but not both. Table 4 displays the observed patterns that were exclusively predicted by rules along dimension 1 or dimension 3.[5] Using these sets of generalization patterns, we partitioned participants into those showing dimension 1 generalizations (47 participants) or dimension 3 generalizations (36 participants).

In the following analyses, we examine categorization, recognition, and speeded categorization for these two subgroups of participants. We expected participants who displayed dimen-

[4] Pattern ABABBAB is consistent with two different interpretations. First, it mirrors the average transfer data (see also Medin & Schaffer, 1978; Nosofsky, Palmeri, & McKinley, 1994), so the context model predicts it to be the most probable generalization pattern.

Second, pattern ABABBAB is also consistent with at least two different rule-based patterns of generalization. Imagine that a participant forms a rule along dimension 1, such that value 1 signals Category A, along with exceptions 1*22 → B and 2*11 → A. Stimuli T1, T3, T4, T5, and T7 would be classified according to the rule as members of Category A, A, B, B, and B, respectively. Now, exception 1*22 matches T2, so it would be classified as a member of Category B, and exception 2*11 matches T6, so it would be classified as a member of Category A. This yields the generalization pattern ABABBAB for transfer stimuli T1–T7.

Similarly, imagine that a participant forms a rule along dimension 3, such that value 1 signals Category A, along with the exceptions 1*21 → A and 2*12 → B. Stimuli T2, T3, T5, T6, and T7 would be classified according to the rule as B, A, B, A, and B, respectively. Exception 1*21 matches T1, so it would be classified in Category A. Exception 2*12 matches stimulus T4, so it would be classified in Category B.

[5] Some of these generalization patterns might be consistent with other single-dimension or conjunctive rules as well.

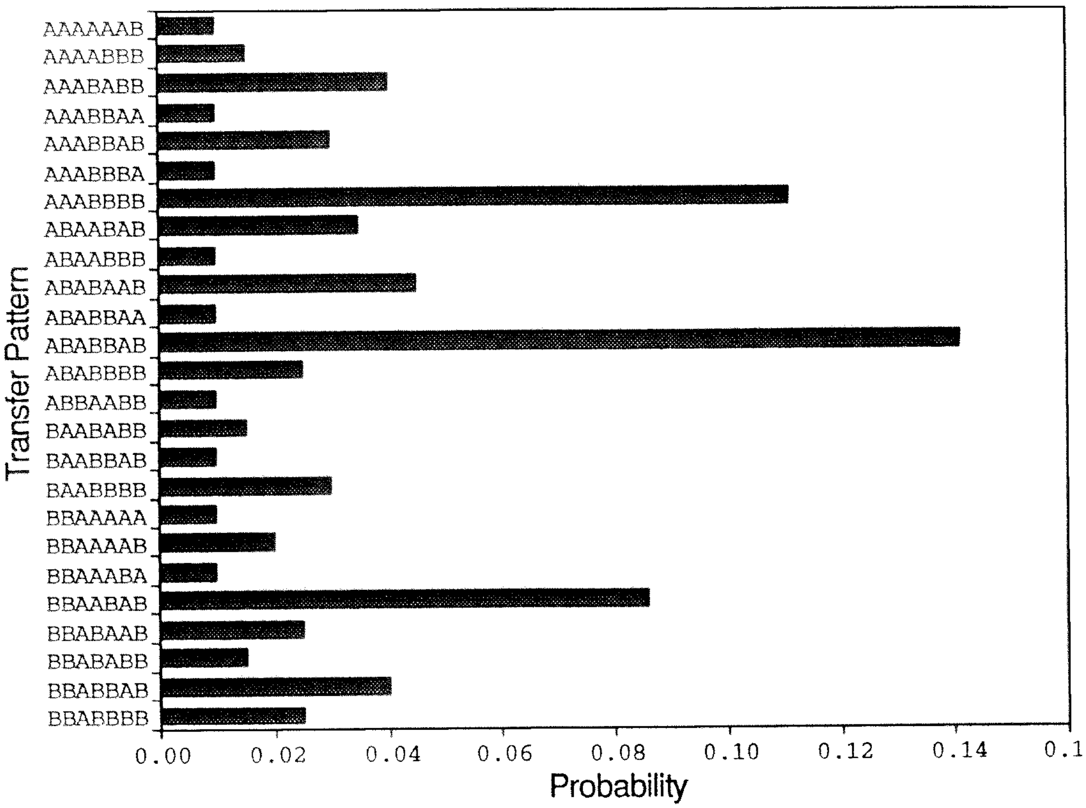

Figure 3. Observed distribution of generalization patterns in Experiment 2. Only those patterns observed in two or more participants are displayed.

sion 1 generalizations to show superior recognition of A5 and B1, and participants who displayed dimension 3 generalizations to show superior recognition of A4 and B2. Furthermore, these items should produce more errors during transfer and be categorized more slowly during the speeded phase. As discussed in Experiment 1, the context model cannot predict superior recognition of both A5 and B1 relative to A4 and B2, nor can it predict the opposite. Therefore, for each group of participants, a critical test will be to compare recognition of A5 and B1 with A4 and B2.

Dimension 3 generalizations. We start by discussing the dimension 3 generalizations because these data provide the clearest evidence in favor of the RULEX predictions. Table 3 displays the categorization data for this subgroup of participants. Most errors were made on the exceptions to the dimension 3 rule, A4 and B2. Furthermore, the tendency to apply the dimension 3 rule to the new transfer items was weakest for those items that were similar to the exceptions, namely T1, T4, and T5. Each of these items differs from one of the exceptions along a single, nonrule dimension. This pattern of results is consistent with the idea that many of these participants used a rule-plus-exception strategy along dimension 3.

Table 4
Generalization Patterns Consistent With Rules Along Dimension 1 and Dimension 3 in Experiment 2 According to RULEX Simulations

Dimension 1	Dimension 3
ABABBBB	BBABBAB
AABBBBB	BBAABBB
AAABBBB	BBAABAB
AAABBBA	BBAABAA
AAABBAB	BBAAAAB
AAABABB	BAAABAB
AAAABBB	ABAABAB

Note. RULEX = rule-plus-exception model.

Figure 4A displays the recognition data for participants who showed dimension 3 generalizations. Consistent with the predictions of a rule-plus-exception strategy, the exceptions to the dimension 3 rule, A4 and B2, tended to be the best recognized items. Also, the average recognition of A4 and B2 was significantly higher than that for A5 and B1, $t(35) = 2.572$, $p < .01$.

Table 5 displays the median speeded categorization response time for participants who made dimension 3 generalizations. Again, consistent with the rule-plus-exception hypothesis, the exceptions to the dimension 3 rule, A4 and B2, were categorized more slowly than any other old exemplar.

Dimension 1 generalizations. Table 3 displays the categorization data for the subgroup of participants who showed patterns of generalization consistent with a logical rule along dimension 1. As expected, the exceptions, A5 and B1, were categorized with the lowest accuracy. Also, the pattern of responses for the transfer items is consistent with participants forming rules along dimension 1 combined with some partial exceptions formation. As in Experiment 1, the tendency to apply the rule along dimension 1 was weakest for those new transfer items that were similar to the exceptions, namely T2, T5, and T6.

Figure 4B displays the recognition ratings for participants who showed dimension 1 generalizations. By comparison to the dimension 3 data, the present recognition data do not provide as clear evidence for the use of rule-plus-exception strategies. That is, some nonexceptions were recognized as well or better than the exceptions. Still, the exceptions received higher recognition ratings than most of the old exemplars. Critically, the average recognition of A5 and B1 was significantly greater than that of A4 and B2, $t(46) = 2.163, p < .05$, a pattern impossible for the context model to predict, regardless of its parameter settings.

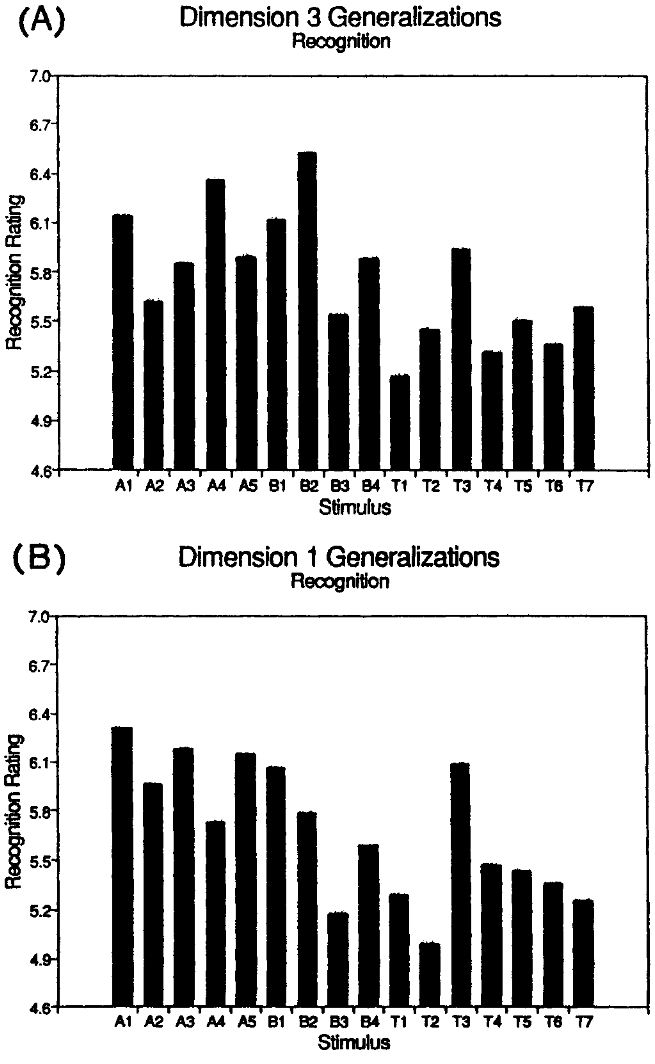

Figure 4. Recognition ratings observed in Experiment 2 for each generalization subgroup. (A): dimension 3 generalizations and (B): dimension 1 generalizations.

Table 5

Speeded Categorization Response Times (in Milliseconds) From Participants Who Showed Dimension 1 and Dimension 3 Generalizations in Experiment 2

Stimulus			
Category and item	Dimension value	Dimension 1	Dimension 3
Category A			
A1	1112	980	957
A2	1212	930	1,004
A3	1211	861	924
A4	1121	938	1,189
A5	2111	1,249	1,139
Category B			
B1	1122	1,327	1,072
B2	2112	1,094	1,316
B3	2221	1,027	1,103
B4	2222	897	1,063
Transfer item			
T1	1221	904	1,176
T2	1222	1,048	1,003
T3	1111	933	981
T4	2212	980	1,010
T5	2121	1,025	1,198
T6	2211	1,131	991
T7	2122	950	1,091

Although the general pattern of recognition data for this subgroup of participants was similar to that found in Experiment 1, A5 and B1 were not the best recognized items. This result could be due to greater exemplar memorization occurring during free-strategy conditions relative to explicit rule-plus-exception conditions. However, it could also be because participants did not adequately store the exceptions to the rules they formed. Unlike in Experiment 1, we did not have enough participants to allow us to discard nonlearners. Hence, many of the participants in this subgroup may indeed have formed a rule along dimension 1, but failed to adequately store the exceptions, so the exceptions were not recognized as well for these participants. Furthermore, although the generalization patterns were consistent with single-dimension rules, some of these participants may have formed various conjunctive rules, causing different stimuli to become the exceptions.

Table 5 displays the median speeded categorization response times for the participants who showed dimension 1 generalizations. As in Experiment 1, the exceptions, A5 and B1, were categorized more slowly than any other item.

Discussion

A major explanation for how people learn ill-defined categories has been that people memorize exemplars and make decisions according to similarity relations among these exemplars (Estes, 1994; Medin & Schaffer, 1978; Nosofsky, 1984, 1986). For the most part, support for these claims has relied on averaged categorization data. However, past and current evidence suggests that averaged categorization data may not always accurately reflect the behavior of individual participants (see Nosofsky, Palmeri, & McKinley, 1994; Trabasso & Bower, 1968). Indeed, others have suggested that data consistent with apparent exemplar-based processes may be the result

Table 6
Category Structure Used in Experiment 3

Category and item	Dimension value
Category A	
A1	21111
A2	11122
A3	12211
A4	11221
A5	12112
A6	11212
A7	12121
A8	11111
Category B	
B1	12222
B2	22211
B3	21122
B4	22112
B5	21221
B6	22121
B7	21212
B8	22222
Transfer item	
T1	22221
T2	22212
T3	22122
T4	22111
T5	21222
T6	21211
T7	21121
T8	21112
T9	12221
T10	12212
T11	12122
T12	12111
T13	11222
T14	11211
T15	11121
T16	11112

of averaging together participants who utilize various idiosyncratic rule-based strategies (cf. Martin & Caramazza, 1980; Ward & Scott, 1987).

In the present experiment, evidence for rule-plus-exception processes under free-strategy conditions came from superior recognition and slower categorization of exceptions. Two subgroups of participants were formed on the basis of rule-based generalizations that were made (as predicted by RULEX). The exceptions to the rule that defined each subgroup tended to be recognized better and categorized more slowly than any other item. In a critical test that bears on the predictive power of the context model, the exceptions for one subgroup were recognized better and categorized more slowly than the items that were the exceptions for the other subgroup. These data were consistent with the predictions of RULEX, replicating and extending the findings of Experiment 1. In contrast, it is impossible for the context model to predict these results (see Appendix A).

One potential way to address the shortcomings of the context model would be to allow individual differences in the similarity parameters for different subgroups of participants. Like the stochastic learning rule in RULEX, which allows different participants to form alternative rules and form varying exceptions, different groups of participants could selectively attend to alternative dimensions. However, as shown in Appendix A, the context model always predicts better recognition of B2 than A5, and better recognition of B1 than A4, regardless of the values of the similarity parameters. Thus, the inability of the context model to predict these results cannot be addressed simply by allowing individual differences in the similarity parameters. Rather, an additional mechanism is required to determine which items are exceptions and to store these items more strongly.

Experiment 3

The previous experiment provided evidence consistent with the use of rule-plus-exception strategies for categorization that led to differential recognition of the exceptions. However, this evidence required us to conditionalize on the patterns of categorization data before generating recognition predictions. The goal of Experiment 3 was to design a category structure that allowed us to predict a priori which items would serve as exceptions to the rule.

The category structure is shown in Table 6. RULEX predicts that nearly all participants will learn this structure by forming a rule along dimension 1 and attempt to remember the exceptions, A1 and B1. Hence, the model predicts that the exceptions, A1 and B1, should be the best recognized items. As we discuss later, this category structure has an interesting property with respect to the predictions of the context model—it is impossible for the context model to predict superior recognition of the exceptions relative to the other old items, regardless of the values of the parameters. In fact, the only qualitative prediction the standard context model can make is that the old items are given higher recognition ratings than the new items.

Inspection of the category structure reveals that a rule-plus-exception strategy is not the only possible strategy that participants could adopt. The categories were generated from two prototypes, A8 and B8, which were presented during training. Thus, dimensions 2–5 are partially diagnostic for determining category membership. That is, a value 1 along each dimension appears more often for Category A exemplars, and a value 2 along each dimension appears more often for Category B exemplars. This arrangement gives the categories a family resemblance structure in addition to a rule-plus-exception structure. Thus, the category structure does not "force" the use of a rule-plus-exception strategy. Alternative strategies, such as prototype or independent-feature strategies, could also be used.

Method

Participants. Participants were 54 members of the Indiana University community who were paid $5.00 for their participation. They could receive up to $3.00 bonus depending on their performance on the three phases of the experiment. All participants were individually tested.

Stimuli. The stimuli were computer-generated drawings of starfish that varied along five binary-valued dimensions: size (large or small), texture (solid or speckled), number of arms (four or six), color (yellow or blue), and outer texture (smooth or spiny; modeled after stimuli used by Ahn & Medin, 1992). All physical dimensions and values along dimensions were randomly assigned to abstract dimensions and values.

The abstract category structure is shown in Table 6. The main characteristic of this structure was that there existed an imperfect rule that value 1 on dimension 1 signals Category A and value 2 on dimension 1 signals Category B. There were two exceptions to this rule, A1 and B1. Furthermore, dimensions 2–5 were somewhat diagnostic; for Category A, the value 1 appeared five eighths of the time along each dimension; for Category B, the value 2 appeared five eighths of the time along each dimension.

Procedure. There was a maximum of 32 blocks of training trials. Each of the 16 training stimuli, A1–A8 and B1–B8, was presented once per block. The order of presentation was randomized for each participant. On each trial, the participant was presented with an item and judged if it was a member of Species A or Species B. After participants responded, corrective feedback was provided for 2 s. The next stimulus was shown after an interval of 500 ms. The training trials were terminated after the participant completed 2 error-free blocks in a row. Participants were paid $1.00 if they reached this criterion before the 32 blocks had been completed.

Following training, participants made old–new recognition judgments about each of the 32 stimuli, using the 8-point confidence rating discussed earlier. There were two blocks of recognition. Participants were told to judge a starfish as old only if it had been seen during training. No corrective feedback was provided. Participants were paid up to $1.00, depending on their performance during the recognition test.

Following the recognition test, participants were given a transfer test using all 32 stimuli and were asked to judge each one as an A or a B. There were two blocks of transfer. No corrective feedback was provided. Participants were paid up to $1.00 depending on how accurately they categorized the old stimuli during the transfer trials.

Results

Categorization and recognition. Only those participants who reached the established criterion of two consecutive error-free training blocks in a row were included in the analyses. Of the 54 participants, 43 reached this criterion (80%). This high criterion was set because a participant who never learned the exceptions could still have achieved approximately 87% accuracy at the end of training. Including participants who never learned the exceptions would have defeated the purpose of testing for superior recognition of the exceptions.

The categorization data obtained during the transfer phase are shown in Table 7. For the nonexceptions, we defined *P*(correct) as the probability of assigning each item to the category dictated by the value along dimension 1. For the exceptions, we defined *P*(correct) as the probability of assigning each exception to the correct category. Because the physical dimensions were randomly assigned to the abstract dimensions for every participant, we did not expect differential categorization or recognition of items that are logically indistinguishable; for example, A2 (11122) and A3 (11221) both differ from the category prototype, A8 (11111), and the exception, A1 (21111), in logically the same way. Thus, we collapsed the data into five distinct item types. A1 and B1 were combined as the exceptions (Exc); A2–A7 and B2–B7 were combined as the old items (Old); A8 and B8 were combined as the prototypes (Pro); T4, T6, T7, T8, T9, T10, T11, and T13 were combined as the new items similar to the exceptions (Sim); the remaining new items were then combined (Dis). As shown in Table 7, more errors were made on the exceptions, Exc, compared with the other old training items. Furthermore, more errors were made on those new items similar to the exceptions, Sim, than the remaining new items, Dis.

Table 7
Categorization Accuracy Observed and Predicted by RULEX and the Context Model in Experiment 3

Stimuli	Observed	RULEX	Context model
Exc	.774	.813	.761
Old	.962	.960	.958
Pro	.989	.994	.997
Sim	.788	.787	.761
Dis	.954	.968	.986

Note. RULEX = rule-plus-exception model; Exc = exceptions; Old = old items; Pro = category prototypes; Sim = new items similar to the exceptions; Dis = remaining new items.

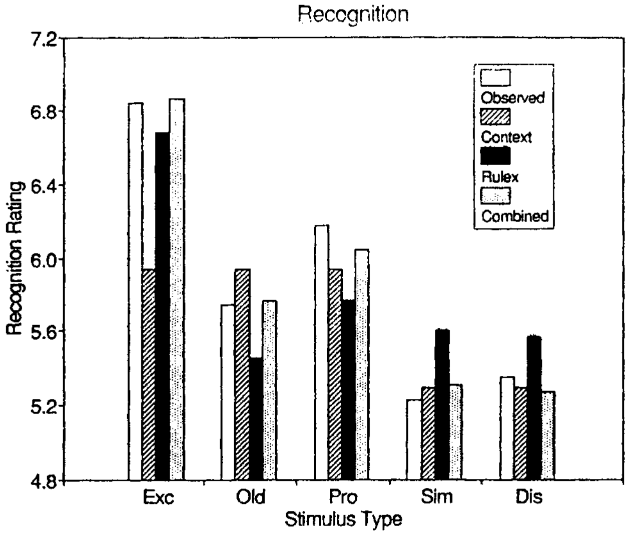

Figure 5. Recognition ratings observed and predicted by the context model, RULEX (rule-plus-exception model), and the combined model in Experiment 3. Exc = exceptions; Old = old items; Pro = prototypes; Sim = new items similar to the exceptions; Dis = remaining new items.

The average recognition ratings are shown in Figure 5. As predicted, the exceptions, A1 and B1 (Exc), were recognized with the highest confidence. Also, the category prototypes, A8 and B8 (Pro), tended to be recognized with higher confidence than the other old stimuli (Old). There was essentially no difference in the recognition ratings given to the two types of new items, Sim and Dis. The old items were recognized with higher confidence than were the new items.

Categorization theoretical analysis. Two-parameter versions of RULEX and the context model each fitted the categorization data quite well, as shown in Table 7. The best fitting parameters for RULEX were *pstor* = 0.610 and *scrit* = 0.790, with *SSE* = 0.0048. The best fitting parameters for the context model were s_1 = 0.002, s_x = 0.228 (where s_x is a common similarity parameter along dimensions 2–5), with *SSE* = 0.0150.

Recognition theoretical analysis. The predicted recognition ratings for the five item types are shown in Figure 5. A two-parameter version of the strict RULEX model fitted the recognition data poorly, *r* = .689. As expected, the strict

RULEX model was able to capture the enhanced recognition of the exceptions, A1 and B1. However, as in earlier simulations, it was unable to predict residual recognition of the old items relative to the new items.

A two-parameter version of the context model also failed to fit the recognition data, $r = .776$. As shown in Figure 5, the only qualitative trend the context model was able to predict was that the old items were recognized with higher confidence than the new items. The context model was unable to predict that the exceptions were the best recognized items nor that the prototypes were better recognized than most of the remaining old items. In fact, this prediction is parameter free. The absolute summed similarity of a given item to the old exemplars is the same for every old item and for every new item, regardless of the parameter settings (see Appendix B).

As shown in Figure 5, the three-parameter combined model predicted superior recognition of the exceptions, A1 and B1 (Exc), relatively high recognition of the category prototypes, A8 and B8 (Pro), and higher recognition of the old exemplars (Old) relative to the new exemplars (Sim and Dis). The quantitative fit was quite good, $r = .987$, with $s_s = 0.000$, $s_w = 0.999$, and $\omega = 0.769$. As expected, the exceptions were more heavily weighted than the other old exemplars, as indicated by the large value of ω. The value of s, the residual similarity parameter, was arbitrary because the absolute summed similarity for every old item and every new item was identical (see Appendix B); thus, the residual exemplar similarity terms merely added a constant amount to the absolute summed similarity of every old item. The model also predicted slightly higher recognition of Sim items relative to Dis items, but a nonsignificant difference was observed in the opposite direction.

Consideration of a rehearsal-borrowing hypothesis. Thus far, we have interpreted our experimental results in terms of rule-plus-exception classification strategies leading to memory representations in which the exceptions have enhanced strength. An alternative view needs to be considered, however, which is not based on rule-plus-exception strategies. In all cases that we have examined, the exceptions to the category rule are also the most difficult items to classify. An exemplar theorist can argue that participants devote special rehearsal to these difficult-to-classify items in an effort to learn them. This rehearsal borrowing leads to increased strength for the exceptions, thus explaining the superior old–new recognition performance observed for these items.

Our main reaction to this rehearsal-borrowing hypothesis is to note that it is extremely similar in spirit to the hypothesis that has motivated our work. Furthermore, whereas the hypothesis that exceptions may have a special status in memory is an a priori prediction stemming from RULEX, the rehearsal-borrowing idea is post hoc. Exemplar models have been used in previous work to predict old–new recognition following category learning, and in no case has an investigator posited increased memory strength for difficult-to-classify items (e.g., Estes, 1986b, 1994; Hayes-Roth & Hayes-Roth, 1977; Medin, 1986; Medin & Florian, 1992; Nosofsky, 1988, 1991, 1992; Shin & Nosofsky, 1992).

Nevertheless, in this section we make some preliminary attempts to address this alternative hypothesis. First, we note that although the hypothesis that difficult-to-classify items have increased memory strength would allow the context model to predict the old–new recognition results, this same hypothesis causes the model trouble in its predictions of categorization. For example, we find that if memory-strength terms are attached to the exceptions so as to enable adequate recognition predictions (see Nosofsky, 1991), then the context model no longer predicts that the exceptions are the worst classified old items.

In another attempt to address the rehearsal-borrowing idea, we tested a model that basically implements a form of rehearsal borrowing in its learning rule. Kruschke's (1992) ALCOVE model is an extended version of the context model incorporating an error-driven learning rule, with association weights learned between stored exemplars and alternative categories. A key property of that model is that, because of its error-driven learning rule, difficult-to-classify items can develop stronger association weights to their respective categories than other items. Indeed, when we fitted ALCOVE to the classification data from the present experiment, we found that it predicted stronger learned association weights for the exceptions than for the other old exemplars. Nevertheless, when the best fitting version of ALCOVE was then used to fit the old–new recognition data, it failed to predict superior recognition of the exceptions. (Old–new recognition predictions are generated in ALCOVE by summing the activations on all category-output nodes, a process that is analogous to the summed-similarity rule in the context model—see Nosofsky and Kruschke, 1992, pp. 225–226, for details.) The key point here is that it is not trivial to find an exemplar model with a learning rule in which exceptions are accorded greater strength that can simultaneously predict the categorization and old–new recognition data.

Discussion

The results of this experiment support the predictions of the rule-plus-exception model of categorization and recognition memory without supplying participants with an explicit strategy, as in Experiment 1, and without partitioning participants on the basis of generalization patterns, as in Experiment 2. The results support the prediction of RULEX with respect to the superior recognition of the exceptions to a rule along dimension 1. In contrast, this fairly straightforward prediction was impossible for the standard context model to predict. Evidence for the use of the rule-plus-exception strategy was obtained despite the fact that the category structure afforded the use of alternative strategies, such as independent-feature, prototype, and exemplar strategies.

Further empirical support for the use of rule-plus-exception processes during free-strategy category learning was supplied by a second closely related experiment that we did not report. In that experiment, the category structure was identical to the one used here, except the prototypes, A8 and B8, were not presented during training. The exceptions, A1 and B1, were by far the best recognized items. Unlike the present experiment, the new transfer items similar to the exceptions (Sim) were given a higher recognition rating than the dissimilar new items (Dis), a result correctly predicted by RULEX. We do not report this second experiment in detail, however, because the context model was also able to predict higher recognition of the exceptions, albeit with highly unusual parameter settings. Nevertheless, although this second experiment was not as clearly diagnostic as the present one, the results add to the generality of the recognition predictions stemming from RULEX.

Experiment 4

In the final experiment we sought to generalize the finding that exceptions to the "category rule" may have a special status in memory. In the previous experiments the stimuli varied along highly separable, binary-valued dimensions. By contrast, the dimensions of natural objects are not always so clearly delineated and they often vary continuously. We used the classic random dot-pattern stimuli used by investigators such as Posner and Keele (1968), Homa (1984), and numerous others. In categorization experiments involving such stimuli, prototypes defining each category are formed by randomly positioning a set of dots. Statistical distortion algorithms are then used to create training exemplars from each of the prototypes. The stimuli created in such experiments are essentially infinitely variable and have a complex dimensional structure, perhaps mimicking the dimensional structure of many natural objects.

In the current experiment, we first formed two prototypes and seven moderate-level distortions of each prototype. Six distortions from one prototype and one distortion from the other prototype formed each of two categories. The distortion created from the prototype of the opposite category can be thought of as the "exception" in its own category. Consistent with previous results, we predicted that, after learning, the exceptions would be better recognized than the other old items. During recognition and transfer, we also included new low-level distortions of the exceptions. We predicted that new items that were similar to the exceptions might have high false-alarm rates.

Unlike the previous experiments, no obvious single-dimension rule was present to classify the items. Instead, each category was defined by a prototype, and the exceptions were those items generated from the opposite prototype. At present, for such continuous-dimension stimuli, there is no model analogous to RULEX that we can compare with the context model at predicting the categorization data. Regardless, we can still focus on the main aim of this research, the role that exceptions play in recognition memory. In the theoretical analyses, we compare the predictions of the recognition data for the standard context model to a version that gives extra weight to the exceptions.

Method

Participants. Participants were 86 undergraduate students at Indiana University who received partial course credit for their participation. Participants who reached a learning criterion received a $1.00 bonus. All participants were individually tested.

Stimuli. Stimuli were random dot patterns similar to those used by Posner and Keele (1968). Each pattern was constructed by randomly placing nine dots on the center 30 × 30 of a 50 × 50 square grid, subject to the constraint that the dots must be at least two units apart. Two prototypes were randomly generated for every participant. Unlike many previous experiments using random dot patterns, every participant was exposed to a different set of randomly generated stimuli. By testing each participant on a different set of stimuli we gain confidence concerning the generality of the results (i.e., they are not due to idiosyncrasies involving any particular set of stimuli).

From each of the two prototypes, Ap and Bp, seven moderate-level distortions were generated (6 bits/dot, see Posner, Goldsmith, & Welton, 1967). Six distorted patterns from one prototype, A1–A6 and B1–B6, and one distorted pattern from the other prototype, Ax and Bx, made up each category. Thus, there was one exception in each category, Ax or Bx, generated from the other prototype, Bp or Ap, respectively. There were also 14 new patterns: two low-level distortions of each exception (3 bits/dot), Axl and Bxl; two moderate-level distortions of each exception (6 bits/dot), Axm and Bxm; two low-level distortions of one of the nonexception patterns, Aol and Bol; and two moderate-level distortions of one of the nonexception patterns, Aom and Bom. Two new moderate-level distortions of the prototype were also created, An and Bn. Along with the two objective prototypes, Ap and Bp, the empirical category prototypes were formed by spatially averaging the six distortions plus the one exception from each category, At and Bt. (Because we found no significant differences in either categorization or recognition of these two types of prototypes, we treat them identically in the *Results* section.)

Procedure. Participants learned to classify each of the old patterns, A1–A6, B1–B6, Ax, and Bx, into Category A or Category B. These 14 stimuli were shown once per block for a total of 20 blocks. Participants were informed at the start of the experiment that they would be paid a $1.00 bonus if at any time during the experiment they correctly classified every stimulus on two consecutive blocks without making an error. On every trial, a dot pattern was shown on the computer screen, and participants were required to respond A or B. Corrective feedback was supplied and the stimulus remained on the screen for 2 s. After an interval of 1 s, the next pattern was shown.

Following the training phase, participants were given a recognition test. The 14 old stimuli and the 14 new stimuli were displayed to the participants once per block for a total of three blocks. The recognition ratings were numerical judgments between (1) *absolutely sure new* and (8) *absolutely sure old.* No feedback was supplied during the recognition task.

Following the recognition phase, participants were given a transfer test. The 14 old stimuli and the 14 new stimuli were displayed once per block for a total of three blocks. Participants judged if each dot pattern was a member of Category A or Category B. No feedback was supplied. Category judgments and response times were recorded.

Results

Categorization and recognition. A criterion of fewer than eight errors on the last four training blocks was established to ensure that only those participants who adequately learned the task were included in the analyses. Because there were two exceptions per block, a participant who never learned the exceptions would make at least eight errors on these last four blocks. A total of 49 of the 86 participants (57%) reached this criterion and were included in all subsequent analyses.

Table 8 displays the average categorization accuracy and the median categorization response times. As in earlier experiments, we converted the categorization response probabilities into P(correct) and combined these data according to item type. As expected, more errors were made on the exceptions than on the other old items. Furthermore, more "errors" were

Table 8
Categorization Accuracy and Median Categorization Response Times (RTs; in Milliseconds) Observed in Experiment 4

Stimuli	*P* (correct)	RT
X	.694	1.405
O	.956	977
Xl	.434	1,294
Xm	.218	1,253
Ol	.925	999
Om	.895	1.147
N	.888	1,065
P	.892	1.101

Note. X = exceptions; O = old items; Xl = new low distortions of the exceptions; Xm = new moderate distortions of the exceptions; Ol = new low distortions of an old item; Om = new moderate distortions of an old item; N = new items; P = true and empirical prototypes.

made on the new distortions of the exceptions. In fact, these items tended to be classified into the opposite category from the exceptions. As we found earlier, the exceptions (X) were also classified several hundred milliseconds slower than other old items (O). Furthermore, those items that were similar to the exceptions (Xl and Xm) were also classified more slowly than the other new items (Ol, Om, N, and P).

Figure 6 displays the average recognition ratings for the eight types of items. The exceptions (X) were better recognized than the old exemplars (O). The old exemplars (O) were given higher recognition ratings than the new exemplars (N). The low distortions of the exceptions (Xl) were given higher recognition ratings than the low distortions of the nonexceptions (Ol), $t(48) = 2.34, p < .01$. Indeed, the low distortions of the exceptions (Xl) were given an average recognition rating that was not significantly different from the old exemplars (O), $t(48) < 1.0$. We regard the high recognition ratings given to the exceptions and low distortions of the exceptions to be consistent with the basic ideas underlying the RULEX model.

Recognition theoretical analysis. Using procedures analogous to ones used in previous work by Homa, Sterling, and Trepel (1981) and Busemeyer, Dewey, and Medin (1984), we defined four parameters to capture the similarity relations among the 28 items in this experiment. These similarity parameters were used to compute the summed similarity of an item to all old exemplars (see Equation 3). The within-category similarity parameter, s_w, defines the similarity between items that surround the same prototype, such as the similarity between A1 and A4, the similarity between an exception and a nonexception of the other category, such as Ax and B4, or the similarity between a new distortion and another item surrounding the same prototype, such as Aol and A2. The between-category similarity parameter, s_b, defines the similarity between an item surrounding one prototype and an item surrounding the other prototype, such as A2 and B4, B3 and Ap, Ax and A4, or Bxl and B5. The prototype similarity parameter, s_p, defines the similarity between an item and the prototype it surrounds, such as A1 and Ap or Bx and Ap. It also defines the similarity between an item and its moderate-level distortion, such as B6 and Bom or Ax and Axm, because the same algorithm was used to create medium-level distortions from prototypes and from old exemplars. Finally, the low-distortion similarity parameter, s_l, defines the similarity between an item and its low-level distortion, such as A6 and Aol or Bx and Bxl. The similarity between an item and itself was set to 1.0.

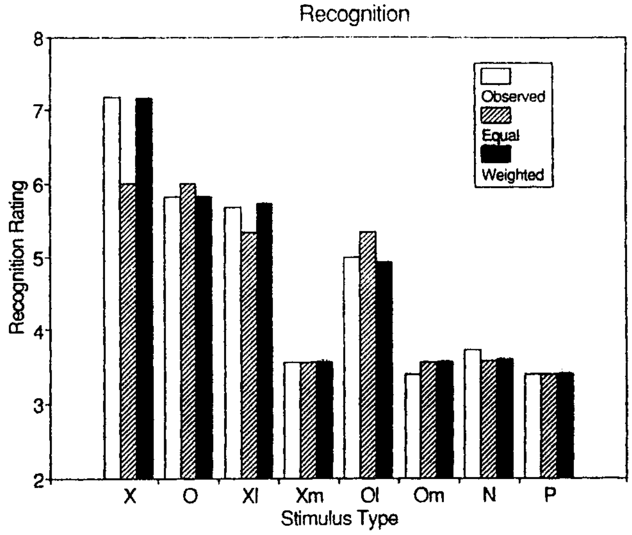

Figure 6. Recognition ratings observed and predicted by the equal and weighted versions of the context model in Experiment 4. X = exceptions; O = old items; Xl = new low distortions of the exceptions; Xm = new moderate distortions of the exceptions; Ol = new low distortions of an old item; Om = new moderate distortions of an old item; N = new items; P = objective and empirical prototypes.

As in the previous analyses we defined the familiarity of a stimulus S_i as $F_i = \omega F_i^X + (1 - \omega)F_i^R$, where F_i^X is the summed similarity of S_i to the exceptions and F_i^R is the residual similarity to remaining exemplars. If $\omega = 1.0$ then only exceptions are remembered; if $\omega = 0.5$ then the exceptions have the same strength as any other item; and, if $0.5 < \omega < 1.0$ then the exceptions have greater strength than the other old items.

We first fitted a four-parameter, equal-weight ($\omega = 0.5$) version of the model to the recognition data. As shown in Figure 6, the fit was fairly good, $r = .950$; however, this model failed to predict the recognition advantage observed for the exceptions (X) and the low distortions of the exceptions (Xl).

We next fitted a four-parameter, strict exception-based ($\omega = 1.0$) version of the model. Again, we were not surprised that the fit was quite poor, $r = .791$. Although the model now captured the superior recognition of the exceptions (X) and the items similar to the exceptions (Xl), there was no way for this model to capture the difference between the old items (O) and the new items (Ol, Om, and N).

Finally, we fitted a five-parameter, weighted (ω free) version of the model to the recognition data. The fit was excellent, $r = .999$, with $s_w = 0.349$, $s_b = 0.267$, $s_p = 0.340$, $s_l = 0.738$, $\omega = 0.616$. As expected, the exceptions were more strongly encoded in memory. As shown in Figure 6, the model accounted for all of the important trends in the data. It predicted that the exceptions were the best recognized items and that the remaining old items were recognized better than most of the new items. Furthermore, it predicted that the new items similar to the exceptions were given recognition ratings comparable to those for the old items.

Discussion

In this experiment we found further evidence for the special role that exceptions play in recognition memory following category learning. We extended the findings from the previous experiments to a situation in which the dimensions were continuous and not readily apparent. The exceptions were those items that were created from the opposite prototype. As in the previous experiments, we found that people demonstrated better recognition for the exceptions than for the other old items. Furthermore, we found that those new items that were similar to the exceptions were also given high recognition ratings. These results are consistent with the predictions stemming from RULEX that exceptions to the "category rule" have a special status in memory.

Although the results of the present experiment are analogous to those observed in our earlier studies, the question arises whether similar processes are involved in forming rules and exceptions for these dot-pattern stimuli as occurred for the stimuli varying along discrete, binary-valued dimensions. We have recently begun investigating a continuous-dimension version of RULEX (Nosofsky, Palmeri, & McKinley, 1993). In the model, single-dimension rules divide a continuous-dimension psychological space into category response regions (cf. Ashby & Townsend, 1986). Exceptions are those items of a category falling outside of the appropriate region defined by the rule. Classification of an item is determined jointly by its similarity to the exceptions and by the category response region in which it falls. As assumed in the baseline version of RULEX that is applicable for stimuli composed of binary-valued dimensions, different participants form alternative single-dimension rules and form different exceptions.

Nosofsky et al. (1993) demonstrated that this continuous-dimension version of RULEX was able to predict a set of classification data involving dot-pattern stimuli nearly as well as did the context model (Shin & Nosofsky, 1992). In the model, the single-dimension rules are defined within a multidimensional scaling solution for the patterns (Carroll & Wish, 1974; Shepard, 1962). The rule dimensions are highly complex and derived, but the same basic principles operate as when participants classify stimuli varying along discrete, binary-valued dimensions.

Further evidence for such rule-based processes in classifying dot-pattern stimuli has been provided by Hock and his colleagues (Hock, Tromley, & Polmann, 1988; Hock, Webb, & Cavedo, 1987). Using a part-parsing procedure, these researchers demonstrated that people are sensitive to large perceptual units, akin to rules, when learning to classify dot patterns. We suggest that once one allows for highly complex, derived perceptual dimensions, the principles underlying RULEX may apply not only to the learning of binary-valued stimuli and complex dot patterns but also to the learning of many natural categories, a point we reprise in our General Discussion section.

General Discussion

The main goal of this research was to examine recognition memory for exceptions to logical rules. In so doing, we provided evidence for the use of rule-plus-exception processes in conjunction with exemplar memorization during category learning. We tested an exemplar-based model, the context model (Medin & Schaffer, 1978), and a rule-plus-exception model, RULEX (Nosofsky, Palmeri, & McKinley, 1994), on categorization and item recognition under a variety of conditions and category structures. In the *Summary* section, we summarize the results of these experiments, and then we discuss possible implications of these results and relations to other research.

Summary

In Experiment 1, participants were supplied with explicit rule-plus-exception instructions before learning in order to control the type of strategy they used. Consistent with the predictions of RULEX, the exceptions were the best recognized items and were the slowest categorized items. This fairly straightforward finding was impossible for the context model to predict. However, even when participants were supplied with an explicit rule-based strategy, residual memory for the other old exemplars was still observed, a result also obtained in all of our subsequent experiments.

In Experiment 2, we found evidence for rule-plus-exception processes during free-strategy conditions. Two subgroups of participants were formed on the basis of rule-based generalizations made when categorizing new transfer items. Mirroring the results from Experiment 1, the exceptions to the rule for each subgroup tended to be the best recognized items overall. Furthermore, as a critical test, the exceptions for one subgroup were recognized better and categorized more slowly than items that were exceptions for the other subgroup. Again, these results provide evidence that the exceptions to simple logical rules have a special status in memory relative to the remaining old items, consistent with predictions of RULEX.

Experiment 3 provided converging evidence for rule-plus-exception processes under free-strategy conditions. In contrast to Experiment 2, the category structure afforded a limited number of possible single-dimension rules that could be formed during learning and so the exceptions could be predicted a priori. Thus, it was not necessary to examine subgroups of participants. As predicted, the exceptions were the best recognized items. These results were impossible for the standard context model to predict; in fact, the context model could only predict that old items were given higher recognition ratings than new items. In contrast, the combined RULEX–exemplar model provided a good account of the data, especially with regard to superior recognition of the exceptions and the category prototypes. Further evidence that exceptions receive stronger memory representations than other items was provided in Experiment 4, which generalized the previous results by using dot pattern stimuli that did not have a clear dimensional structure.

Relations to Other Research

One possible way of viewing the exception formation process is as a form of stimulus-specific selective attention, which has previously been proposed as a necessary extension to the context model (Medin & Edelson, 1988). In the standard context model, selective attention is applied to dimensions uniformly across all items, thereby playing a role similar to that of explicit rules. Under certain conditions, particular dimensions could come to be more highly attended for some items compared with other items. Unfortunately, there have been no published accounts that formalize the processing assumptions of such a model. RULEX, in essence, may be seen to implement one form of stimulus-specific selective attention. For most items, the dimension along which a rule has been formed receives the most attention. For the exceptions, however, other dimensions must also be given some attention so as to distinguish these items from nonexceptions.

To what extent may rule-plus-exception processes govern learning and representation of natural categories? Also, what is the nature of exceptions in other cognitive domains? We now address these questions in several different areas: social expectations, semantic memory, face recognition, perceptual expertise, and language learning.

Research on social expectations has found that, in general, memory is better for expectancy-incongruent (exceptional or atypical) than expectancy-congruent (rule-following or stereotypical) information (Stangor & McMillan, 1992). The expectancies in these experiments derive from previous experiences, whereas the expectancies in our experiments derive from experimental factors. Some researchers have posited that expectancy-incongruent information is stored explicitly, in a separate memory trace, from expectancy-congruent information, hence, expectancy-incongruent information is better recognized (see Stangor & McMillan, 1992, for a review).[6]

Likewise, most semantic memory models posit category exceptions to have a special status relative to other category exemplars (see Chang, 1986, for a recent review). For example, semantic networks and schemas often assume information about exceptional items to be stored directly with those items, whereas information about typical items is stored at a more general category level. Furthermore, classic research on semantic verification found faster response times for typical objects, "a robin is a bird," than for atypical items, "an ostrich is a bird." Similarly, we found rule-following items to be categorized faster than exceptions. Regardless of the particular model, exceptional items are represented explicitly, apart from general category representations.

Exceptions appear to have a special status in perceptual domains as well. Evidence suggests that distinctive or unusual faces are processed differently than typical faces (e.g., Bartlett, Hurry, & Thorley, 1984; Valentine & Ferrara, 1991). Similar to our results, typical faces are classified faster than atypical (exceptional or distinctive) faces, but atypical faces are better recognized than typical faces. Valentine and Ferrara (1991) argued that the standard context model could not account for these "distinctiveness" effects in face recognition.

In another perceptual domain, Biederman and Shiffrar (1987) examined expertise in an extraordinarily difficult perceptual task, sexing day-old chicks. Experts can classify 1,000 chicks per hour at over 98% accuracy, but they require many years to achieve this level of performance. For a naive observer, it is not at all obvious how to tell male and female chicks apart. From interviews with an expert (who had spent 50 years classifying over 55 million chicks), Biederman and Shiffrar discovered a fairly simple perceptual "rule" this expert used to classify most examples. However, following attainment of high accuracy, this expert spent many years learning (memorizing) the rare, exceptional configurations of genitalia. At least for this one case, perceptual expertise can be characterized as a rule-plus-exception process. Research has only recently begun to examine perceptual aspects of expertise in other areas, such as medical diagnosis (e.g., Brooks, Norman, & Allen, 1991; Lesgold, Glaser, Rubinson, Klopfer, Feltovich, & Wang, 1988).

Finally, debate has centered on the use of rules and exceptions in language (e.g., Pinker & Prince, 1988; Rumelhart & McClelland, 1986). There is a posited U-shaped learning curve for the acquisition of the past tense—children initially learn both regular, rule-following verb forms (*walked* or *killed*) and irregular, exceptional verb forms (*went* or *sang*); later, they overgeneralize the rule (*wented* or *goed*); finally, they acquire an adult understanding of both regular and irregular verbs. Traditional linguistic theories posit an explicit notion of rules and exceptions (see Pinker & Prince, 1988)—"not only must the child induce the rules which underlie the use of regular linguistic forms, he [she] must also learn the exceptions to these rules" (Kuczaj, 1977, p. 600). Connectionist models reject explicit rules entirely; however, Rumelhart and McClelland (1986) acknowledged the special status of exceptions by allowing them to acquire stronger representations early in learning (see Kruschke, 1992; Lachter & Bever, 1988).

In conclusion, the emphasis in the present work was on the special role that exceptions to category rules may play in recognition memory. Our model-based analyses of the old–new recognition data provide converging evidence that rule-plus-exception processes may be common strategies in category learning. Nevertheless, our analyses also point to a continued role of old exemplars that are not exceptions to the category rule. Just as rules, exceptions, and old exemplars appear to jointly influence categorization and memory (e.g., Medin & Ross, 1989; Nosofsky, 1992; Regehr & Brooks, 1993), it is critical that continued research explore the co-existence of exemplar-storage strategies and rule-based strategies in other fundamental cognitive tasks, such as similarity (Smith & Sloman, 1994), problem solving (Medin & Ross, 1989; Ross, 1987), and induction (Medin & Ross, 1989; Smith, Langston, & Nisbett, 1992).

[6] Theoretical analyses conducted by Heit (1993) suggest that certain memory effects involving expectancy-congruent and expectancy-incongruent information can be modeled in terms of storage of prior examples of the expectancy-congruent information. Our results suggest, however, that there are at least certain experimental conditions that accord a special memory status to incongruent information as well.

References

Ahn, W.-K., & Medin, D. L. (1992). A two-stage model of category construction. *Cognitive Science, 16,* 81–121.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–179.

Bartlett, J. C., Hurry, S., & Thorley, W. (1984). Typicality and familiarity of faces. *Memory & Cognition, 12,* 219–228.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert system analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 640–645.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120,* 278–287.

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

Carroll, J. D., & Wish, M. (1974). Models and methods for three-way multidimensional scaling. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2, pp. 57–105). San Francisco: Freeman.

Chang, T. M. (1986). Semantic memory: Facts and models. *Psychological Bulletin, 99,* 199–220.

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

Estes, W. K. (1986b). Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology: General, 115,* 155–174.

Estes, W. K. (1994). *Classification and cognition.* London: Oxford University Press.

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

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

Heit, E. (1992). Categorization using chains of examples. *Cognitive Psychology, 24,* 341–380.

Heit, E. (1993). Modeling the effects of expectations on recognition memory. *Psychological Science, 4,* 244–252.

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

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

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

Hock, H. S., Webb, E., & Cavedo, L. C. (1987). Perceptual learning in visual category acquisition. *Memory & Cognition, 15,* 544–556.

Hoffman, J., & Ziessler, C. (1983). Objectidentifikation in kunstlichen begriffshierarchien [Object identification in artistic concept hierarchies]. *Zeitscrift fur Psychologie, 16,* 243–275.

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

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

Kruschke, J. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44.

Kuczaj, S. A. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior, 16,* 589–600.

Lachter, J., & Bever, T. G. (1988). The relation between linguistic structure and associative theories of language learning: A constructive critique of some connectionist learning models. In S. Pinker & J. Mehler (Eds.), *Connections and symbols* (pp. 195–247). Cambridge, MA: MIT Press.

Lesgold, A., Glaser, R., Rubinson, H., Klopfer, D., Feltovich, P., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M. T. H. Chi. R. Glaser. & M. J. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale. NJ: Erlbaum.

Levine, M. (1975). *A cognitive theory of learning: Research on hypothesis testing.* Hillsdale, NJ: Erlbaum.

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

Medin, D. L. (1986). Commentary on "Memory storage and retrieval processes in category learning." *Journal of Experimental Psychology: General, 115,* 373–381.

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

Medin, D. L., & Edelson, S. M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General, 117,* 68–85.

Medin, D. L., & Florian, J. E. (1992). Abstraction and selective coding in exemplar-based models of categorization. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 207–235). Hillsdale, NJ: Erlbaum.

Medin, D. L., & Ross, B. H. (1989). The specific character of abstract thought: Categorization, problem solving, and induction. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5, pp. 189–223). San Diego, CA: Academic Press.

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

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 355–368.

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

Medin, D. L., Wattenmaker, W. D., & Michalski, R. S. (1987). Constraints and preferences in inductive learning: An experimental study of human and machine performance. *Cognitive Science, 11,* 299–339.

Metcalfe, J., & Fisher, R. P. (1986). The relation between recognition memory and classification learning. *Memory & Cognition, 14,* 164–173.

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

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

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

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

Nosofsky, R. M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multidimensional 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, and Cognition, 15,* 282–304.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. T. (1994). Comparing models of rule-based classification learning: A replication of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition, 22,* 352–369.

Nosofsky, R. M., & Kruschke, J. K. (1992). Investigations of an exemplar-based connectionist model of category learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 28, pp. 207-250). San Diego, CA: Academic Press.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1993, November). *Rule-plus-exception model of classification learning.* Paper presented at the 34th annual conference of the Psychonomic Society, Washington, DC.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79.

Omohundro, J. (1981). Recognition vs. classification of ill-defined category exemplars. *Memory & Cognition, 9,* 324–331.

Palmeri, T. J., & Nosofsky, R. M. (1993). Generalizations by rule models and exemplar models of category learning. In W. Kintsch (Ed.), *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 794–799). Hillsdale, NJ: Erlbaum.

Pavel, M., Gluck, M. A., & Henkle, V. (1988). Generalization by humans and multi-layer networks. *Proceedings of the 10th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. In S. Pinker & J. Mehler (Eds.), *Connections and symbols* (pp. 73–193). Cambridge, MA: MIT Press.

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

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Regehr, G., & Brooks, L. R. (1993). Perceptual manifestations of an analytic structure: The priority of holistic individuation. *Journal of Experimental Psychology: General, 122,* 92–114.

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

Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 629–639.

Rumelhart, D. E., & McClelland, J. J. (1986). On learning the past tenses of English verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume I: Foundations* (pp. 216–271). Cambridge, MA: Bradford Books/MIT Press.

Shepard, R. N. (1962). The analyses of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika, 27,* 125-140, 219–246.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75*(13, Whole No. 517).

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General, 121,* 278–304.

Smith, E. E., Langston, C., & Nisbett, R. E. (1992). The case for rules in reasoning. *Cognitive Science, 16,* 1–40.

Smith, E. E., & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory & Cognition, 22,* 377–386.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social development literatures. *Psychological Bulletin, 111,* 42-61.

Trabasso, T., & Bower, G. H. (1968). *Attention in learning: Theory and research.* New York: Wiley.

Valentine, T., & Ferrara, A. (1991). Typicality in categorization, recognition and identification: Evidence from face recognition. *British Journal of Psychology, 82,* 87–102.

Ward, T. B., & Scott, J. (1987). Analytic and holistic modes of learning family-resemblance concepts. *Memory & Cognition, 15,* 42–54.

Appendix A

Context Model Predictions of Recognition in Experiments 1 and 2

In this appendix we prove that the context model cannot predict the observed pattern of recognition data in Experiments 1 and 2. In particular, the context model cannot predict better recognition of A5 than A4 and better recognition of B1 than B2, simultaneously, for the category structure shown in Table 1 in the text (nor can it predict the inverse).

First, define the similarity parameters α, β, γ, and δ for dimensions 1, 2, 3, and 4, respectively.

Second, calculate the absolute summed similarity (familiarity) of stimuli A4, A5, B1, and B2 to all of the old exemplars by using Equations 2 and 3 from the text. (Recall that recognition is a monotonically increasing function of familiarity.)

$$F_{A4} = \gamma\delta + \beta\gamma\delta + \beta\gamma + 1 + \alpha\gamma + \delta + \alpha\gamma\delta + \alpha\beta + \alpha\beta\delta$$

$$F_{A5} = \alpha\delta + \alpha\beta\delta + \alpha\beta + \alpha\delta + 1 + \alpha\gamma\delta + \delta + \beta\gamma + \beta\gamma\delta$$

$$F_{B1} = \gamma + \beta\gamma + \beta\gamma\delta + \delta + \alpha\gamma\delta + 1 + \alpha\gamma + \alpha\beta\delta + \alpha\beta$$

$$F_{B2} = \alpha + \alpha\beta + \alpha\beta\delta + \alpha\gamma\delta + \delta + \alpha\gamma + 1 + \beta\gamma\delta + \beta\gamma$$

Third, compare the summed similarities and cancel common terms, yielding the following predicted relations:

$$F_{A5} > F_{A4} \leftrightarrow \alpha > \gamma \tag{A}$$

$$F_{B1} > F_{B2} \leftrightarrow \gamma > \alpha \tag{B}$$

$$F_{B2} > F_{A5} \quad \text{always} \tag{C}$$

$$F_{B1} > F_{A4} \quad \text{always} \tag{D}$$

It is impossible for Relations A and B to hold simultaneously, regardless of the parameter values. Hence, the context model cannot predict better recognition of A5 than A4 and better recognition of B1 than B2, simultaneously (nor can it predict the inverse).

(*Appendixes continue on next page*)

Appendix B

Context Model Predictions of Recognition in Experiment 3

In this appendix we illustrate that the context model predicts the same recognition ratings for all old items and for all new items in Experiment 3. The category structure is shown in Table 4 in the text.

Define similarity parameters α, β, γ, δ, and ϵ for dimensions 1, 2, 3, 4, and 5, respectively.

Calculate the absolute summed similarity (familiarity) between an item and all of the old items by using Equations 2 and 3 in the text. For illustration, we calculate the familiarity of items A2 and B1.

$$F_{A2} = \alpha\delta\epsilon + 1 + \beta\gamma\delta\epsilon + \gamma\epsilon + \beta\delta + \gamma\delta + \beta\epsilon + \delta\epsilon + \beta\gamma$$
$$+ \alpha\beta\gamma\delta\epsilon + \alpha + \alpha\beta\delta + \alpha\gamma\epsilon + \alpha\beta\epsilon + \alpha\gamma\delta + \alpha\beta\gamma$$

$$F_{B1} = \alpha\beta\gamma\delta\epsilon + \beta\gamma + \delta\epsilon + \beta\epsilon + \gamma\delta + \beta\delta + \gamma\epsilon + \beta\gamma\delta\epsilon + 1$$
$$+ \alpha\delta\epsilon + \alpha\beta\gamma + \alpha\gamma\delta + \alpha\beta\epsilon + \alpha\gamma\epsilon + \alpha\beta\delta + \alpha$$

By rearranging terms, it is easy to see that both equations are identical. In fact, the familiarity of every old item is identical. Hence, the context model must predict the same recognition ratings for every old item.

Similarly, perform calculation for the new items. For illustration, we calculate the familiarity of items T4 and T14.

$$F_{T4} = \beta + \alpha\beta\delta\epsilon + \alpha\gamma + \alpha\beta\gamma\delta + \alpha\epsilon + \alpha\beta\gamma\epsilon + \alpha\delta + \alpha\beta + \alpha\gamma\delta\epsilon$$
$$+ \beta + \beta\delta\epsilon + \epsilon + \beta\gamma\delta + \gamma + \beta\gamma\epsilon + \gamma\delta\epsilon$$

$$F_{T9} = \alpha\beta\gamma\delta + \beta\gamma\epsilon + \delta + \beta + \gamma\delta\epsilon + \beta\delta\epsilon + \gamma + \beta\gamma\delta + \epsilon + \alpha\delta$$
$$+ \alpha\beta\gamma\epsilon + \alpha\gamma\delta\epsilon + \alpha\beta + \alpha\gamma + \alpha\beta\delta\epsilon + \alpha\epsilon$$

By rearranging terms, we see that the sums are identical. Hence, the context model must predict the same recognition ratings for every new item.

Received January 10, 1994
Revision received May 26, 1994
Accepted June 17, 1994 ■

Mentors for Journal Authors Needed

APA's Committee on International Relations in Psychology is encouraging publication of international scholars' manuscripts in U.S. journals. To accomplish this initiative, the committee is asking for U.S. "mentors" who are willing to work with non-English-language authors to bring the manuscripts into conformity with English-language and U.S. publication standards. The committee is looking for both translators and those with APA journal experience. Interested individuals should contact

APA International Affairs Office
750 First Street, NE
Washington, DC 20002
Telephone: (202) 336-6025
FAX: (202) 336-5919
Internet: jxb.apa@email.apa.org