Journal of Experimental Psychology:
Learning, Memory, and Cognition
2003, Vol. 29, No. 1, 107–117

0278-7393/03/$12.00 DOI: 10.1037/0278-7393.29.1.107

Classification of Exemplars With Single- and Multiple-Feature Manifestations: The Effects of Relevant Dimension Variation and Category Structure

Arthur B. Markman and W. Todd Maddox
University of Texas at Austin

Most classification research focuses on cases in which each abstract feature has the same surface manifestation whenever it is presented. Previous research finds that people have difficulty learning to classify when each abstract feature has multiple surface manifestations. These studies created multiple manifestations by varying aspects of the stimuli irrelevant to the abstract feature dimension. In this article, multiple manifestations were created by varying aspects of the stimuli relevant to the abstract feature dimension. People given categories with the family resemblance category structure often used in psychology experiments had difficulty learning to classify when multiple manifestations were present, even though the variation was relevant. This effect was reversed when a family resemblance structure with nondiagnostic values was used.

Category members differ along many dimensions. For example, cars differ in their size, engine power, styling, and numerous other attributes. Despite this variation, we have little trouble recognizing these items as cars and classifying them appropriately. Although we are sensitive to variation among members of a category, it has proven difficult to characterize the systematic effects of variability in laboratory experiments of category learning by classification. In one study, Yamauchi and Markman (2000) asked people to classify a set of novel bugs into one of two categories. The two categories had a linearly separable family resemblance structure often used in studies of classification (Table 1).

One category tended to have the abstract value *A* on each feature dimension, although each exemplar had one feature dimension with the abstract value *B*. We refer to the features of an exemplar of one category that are diagnostic of the other category as exception features. In contrast, the other category was dominant in the value *B* (and had the value *A* as exception features). For example, one category might be characterized by exemplars that have four legs, whereas the other might be characterized by exemplars that have eight legs. For one group of participants, each feature value had only a single manifestation. That is, as shown in the top part of Figure 1, every time a bug with four legs was seen, that bug had the same version of the four legs. Similarly, for another dimension such as body marking (stripes vs. dots), there was only one manifestation of each value. In contrast, a second group of participants was presented with bugs for which the abstract feature value had many different manifestations. For example, although one category of bugs tended to have four legs, the specific size and shape of those legs differed from trial to trial. Figure 1 (bottom) shows four stimuli that all have the same abstract feature structure but differ in the specific manifestations of those features.

Most laboratory studies of classification correspond to the single-manifestation condition, in which each abstract feature value has a single manifestation (e.g., Medin & Schaffer, 1978). Consistent with previous research, Yamauchi and Markman (2000) found that participants given the categories with single manifestations of each feature value learned the categories easily. In contrast, those given the categories with multiple manifestations had great difficulty learning the categories. Most participants failed to learn the categories even after 240 training trials.

This result mirrors a previous finding by Medin, Dewey, and Murphy (1983). In their study, participants learn to classify photographs into families based on dimensions such as hair color and smile type. One group of participants was shown a small number of pictures, and each block consisted of one pass through the set. This group learned to classify the photographs correctly. In contrast, a second group was shown a set of categories that was characterized by the same set of abstract properties (e.g., light vs. dark hair), but a new picture was seen on every trial (so that no picture was seen more than once). This group, for whom the abstract properties also had different manifestations on each trial, had great difficulty learning the categories, and 72% of participants in that study failed to reach the learning criterion.

In this article, we explore the generality of the finding that learning is made more difficult by the presence of multiple manifestations of the feature values. Although this result is interesting on empirical grounds alone, it has a number of implications for current models of category learning, which we discuss later. First, we present two reasons why the difficulty of classification learning

Arthur B. Markman and W. Todd Maddox, Department of Psychology, University of Texas at Austin.

This work was supported by National Science Foundation Grant SBR-9905013 to Arthur B. Markman and National Institutes of Health Grant R01 MH59196 to W. Todd Maddox. We thank Sam Day, Ryan Gossen, Brad Love, and Hunt Stilwell for helpful discussions with this line of work; Robert Nosofsky for comments on a previous version of this article; and Tiki Barber for excellent running.

Correspondence concerning this article should be addressed to Arthur B. Markman, Department of Psychology, University of Texas, Austin, Texas 78712.

Table 1
Family Resemblance Feature Structure

Exemplar	Dimension 1	2	3	4	Category
A1	A	A	A	B	A
A2	A	A	B	A	A
A3	A	B	A	A	A
A4	B	A	A	A	A
B1	B	B	B	A	B
B2	B	B	A	B	B
B3	B	A	B	B	B
B4	A	B	B	B	B

would increase with the number of manifestations and describe two experiments that explore these explanations.

Multiple Manifestations and Learning

A careful examination of the materials in these studies suggests two potential explanations for the increased difficulty of learning tasks involving features with multiple manifestations. First, in the studies by Yamauchi and Markman (2000) and Medin et al. (1983), the variability was irrelevant to the basis for categorization. For example, in the studies by Yamauchi and Markman, the number of legs was relevant for correct categorization, but the shape, color, and length of the legs also varied among manifestations. Thus, the irrelevant variation may have made it difficult for people to determine which aspects of the features (e.g., the number of legs) were relevant for correct classification. That is, the categories were defined in terms of abstract attributes, and the irrelevant surface variation of the exemplars may have made it difficult for people to extract the appropriate abstract features.

A second possibility is that the particular category structure used in these tasks affects performance. There is much research on classification that involves stimuli that have multiple manifestations. In studies by Ashby, Maddox, and Bohil (e.g., Ashby, 1992; Maddox & Ashby, 1993; Maddox & Bohil, 1998), category exemplars are often defined along two dimensions (e.g., length and orientation of a line segment), in which the specific values of the exemplar are drawn from a bivariate normal distribution centered on some point. Often the distributions of the two categories overlap, and so there are some values for each category that are nondiagnostic, because they may be possessed by members of either category. Importantly, it is rare that an exemplar of one category manifests the most typical values of the other category. This analysis suggests that a category structure involving nondiagnostic feature values rather than exception features might lead to different performance in the face of multiple-feature manifestations (see Lassaline & Murphy, 1996, for a similar discussion).

One reason for exploring these two category structures is that they may differ in the degree to which they promote learning by searching for explicit rules as opposed to using a more holistic similarity-based process (Lockhead, 1979; J. D. Smith & Kemler-Nelson, 1984; L. B. Smith, 1989). Hypothesis testing will become more difficult as the number of manifestations increases because the number of possible rules is an exponential function of the number of manifestations. In contrast, the ease of holistic processes depends on the similarity relationships among the categories rather than on the number of manifestations of the features. Category structures with exception features may promote a relatively more analytical mode of learning than category structures with nondiagnostic features. Although there are many possible reasons why these structures might differ, one prominent explanation is that the presence in one category of exception values that are diagnostic of the other category may lead people to attempt to explain the occurrence of these discrepant values.

Two experiments addressed these factors. In Experiment 1, we constructed a classification task that varied the number of manifestations of the features using the family resemblance category structure with exception features. Experiment 2 contrasts performance on this category structure with that on a family resemblance structure with nondiagnostic features. In both of these studies, manifestations were created by incorporating variation along relevant dimensions instead of along irrelevant dimensions as in previous research. The stimuli were bugs, like those shown in Figure 2, that had four relevant dimensions: tail, body, wings, and antenna. For each dimension, the darkness of the shading of the feature was relevant. For convenience, we label these levels of darkness with the numbers 1 to 8, with 1 being the lightest gray and 8 being the darkest gray.

In Experiment 1, the categories had the family resemblance structure with exception features used by Yamauchi and Markman (2000; shown in Table 1), in which the mapping of the values *A* and *B* to light (Levels 1–4) and dark (Levels 5–8) shading was selected randomly for each dimension for each participant. Thus, whereas value *A* might be randomly assigned to Levels 1 to 4

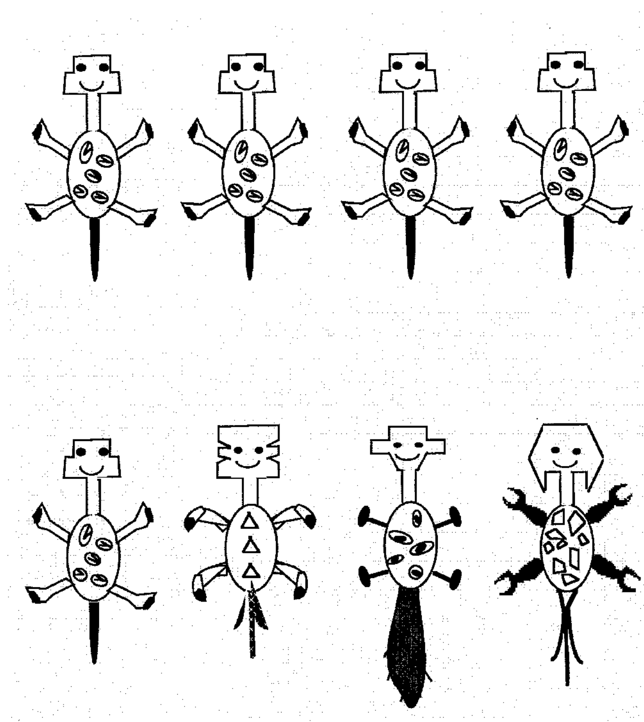

Figure 1. Example of the way multiple-feature manifestations were instantiated in the study by Yamauchi and Markman (2000).

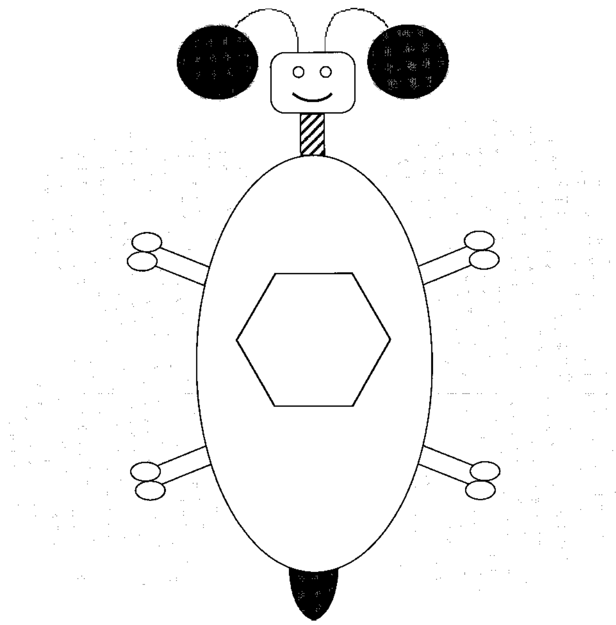

Figure 2. Sample stimulus used in the current study.

along Dimension 1 for a given participant, it might be assigned to Levels 5 to 8 along Dimension 2 for the same participant. In Experiment 2, we compare performance with this category structure with a family resemblance structure that replaces the exception features with nondiagnostic feature values.

In Experiment 1, the number of feature manifestations was varied systematically from 1 to 4. In the one-manifestation condition, the light shading used the value 4 (in the range from 1–8) and the dark shading used the value 5. The stimuli were pretested to ensure that these values were discriminable. In the two-manifestation condition, the values 3 and 4 were used for the light shading and the values 5 and 6 for the dark shading. The three-manifestation condition used values 2, 3, and 4 for the light shading and 5, 6, and 7 for the dark shading. Finally, the four-manifestation condition used values 1, 2, 3, and 4 for the light shading and values 5, 6, 7, and 8 for the dark shading. Thus, the average category members (i.e., the centroids) will get further apart (and hence more discriminable) as the number of manifestations increases.[1]

It is not clear whether the single-manifestation condition should be easier or harder than the multiple-manifestation condition for items with variation that is relevant to categorization. Although a large body of research has examined classification performance when the stimulus values had variation along aspects relevant for classifying the items (see, e.g., Ashby & Maddox, 1998; Estes, 1994, for a review), no systematic examination of the influence of increases in the number of manifestations along each dimension has been undertaken. In addition, because previous research uses overlapping categories sampled from a multivariate normal distribution, increases in category variability are generally associated with decreases in optimal accuracy (Maddox & Dodd, 2001). In the current study, optimal accuracy is constant (100%) in all conditions.

Experiment 1

Method

Participants. One hundred twelve undergraduates at the University of Texas at Austin (28/condition) were given course credit for their participation.

Materials. The stimuli were stylized bugs like those shown in Figure 2. The head and legs of the bugs did not change. The darkness of the bulbs on the antennae, the wings, the tail, and a hexagon on the body were varied. These four properties are referred to as stimulus dimensions. The categories to be learned had the structure shown in Table 1. The value of light (Levels 1–4) and dark (Levels 5–8) associated with the values *A* and *B* in Table 1 was determined randomly for each dimension and each participant.

There were eight possible values of darkness for each of the four dimensions. The lightest gray was created with 85% saturation of the red, blue, and green channels using a graphics package. Each additional step involved subtracting 10% saturation from each channel. The darkest gray involved 15% saturation on each channel. The one-manifestation condition used only values 4 (light) and 5 (dark). The remaining conditions added contiguous values, in which the separation between light and dark was always between Darkness Levels 4 and 5.

On each trial, an abstract stimulus was drawn from the set shown in Table 1. Specific values for each feature dimension were selected randomly (with replacement) from the set of possible feature values for that condition. Each feature value appeared with equal probability. Thus, in the one-manifestation condition, all of the light-colored features had the feature value 4, but in the four-manifestation condition approximately 25% of the light-colored features had the feature value 4.

Trials were grouped into blocks of eight, each consisting of one pass through the stimulus set. The order of presentation of the stimuli in each block was determined randomly.

Procedure. Participants were seated at a Macintosh computer with a color screen. They were told that they would see a number of bugs and were asked to classify them into one of two mutually exclusive categories (given the labels *Mornek* and *Plaple*). Responses were made by pressing the "z" and "/" keys on the keyboard. Categories were randomly assigned to keys for each participant.

On each trial, a stimulus was selected, and the feature values for that stimulus were determined as shown previously. The stimulus was shown on the screen along with the instructions to press the button corresponding to the category to which it belonged. After their response, participants were told whether they were correct, and the correct category label and the stimulus remained on the screen for 3 s. The experiment continued until participants reached a criterion of three consecutive blocks with a combined accuracy of more than 90% (i.e., at least 22 of 24 correct responses) or until they had reached a maximum of 30 blocks.

Design. The main between-participants factor in this study was number of feature manifestations. There were four levels of this factor (1, 2, 3, and 4). The main dependent measures were the number of blocks required to reach the learning criterion and the proportion of participants reaching the learning criterion before the 30-block maximum.

[1] One possible concern is that the categories may differ by a holistic property of overall brightness contrast. On this view, people are using a feature set different from the one we are manipulating. Although this factor may explain the performance of some participants, each dimension was randomly assigned to be light or dark for each participant. Overall brightness will allow people to classify all of the stimuli correctly only when all of the dimensions for a given category are prototypically light or dark. For any other assignment of values to dimensions, there is at least one stimulus in each category for which overall brightness is not diagnostic.

Results

The main results are summarized in Table 2. Consistent with previous research, participants required more blocks to reach the learning criterion when there were many feature manifestations as opposed to few feature manifestations. A one-way analysis of variance (ANOVA) on these data revealed a significant effect of number of manifestations, $F(3, 108) = 6.24$, $MSE = 56.72$, $p < .05$. Post hoc contrasts using Tukey's honestly significant difference (assuming $\alpha = .05$) revealed that participants reached the learning criterion faster in both the one- and two-manifestations conditions than in the four-manifestations condition. No other differences between conditions were significant, although the difference between the two- and three-manifestations conditions was marginally significant ($p = .08$).

This pattern mirrors that obtained by examining the proportion of participants to reach the learning criterion in each condition. As shown in Table 2, about half of the participants reached the criterion in both the one- and two-manifestations conditions. In contrast, only 29% of the participants reached the criterion in the three-manifestations condition, and 14% of the participants reached the criterion in the four-manifestations condition.

One reason why increasing the number of feature manifestations can make classification more difficult is that people may be seeking a rule that distinguishes between the categories (Nosofsky, Palmeri, & McKinley, 1994). The hypothesis space increases exponentially with the number of feature manifestations on each dimension, thereby increasing the difficulty of the task. To explore people's strategies, we examined the learning curves for more insight.

To analyze the learning curves, we divided the 30 potential blocks of trials (240 total trials) into 10 bins of 3 blocks (24 trials) apiece. For each group of 3 blocks, we calculated the proportion of correct responses. Many participants reached the 90% accuracy criterion before completing 30 blocks of trials. For some of these participants, the last bin may have fewer than 24 trials in it, so the proportion correct for the number of trials they completed was calculated.

Then we examined three different accuracy criteria: 70%, 80%, and 90%. In particular, we determined the first bin of blocks in which they exceeded each of these three criteria. We selected these three criteria because if a participant is trying to find a rule that distinguishes between categories, then a unidimensional rule is likely to be the first one selected. Any unidimensional rule will correctly classify 75% of the stimuli, and so we expect that participants in all conditions will reach at least 70% accuracy fairly quickly. Achieving higher accuracy rates in this task requires a more complex rule involving three dimensions or a combination of a rule and exemplar storage. Because finding complex rules and storing individual exemplars becomes more difficult as the number of manifestations increases, there should be a greater difference between manifestation conditions for the 80% and 90% accuracy criteria than for the 70% accuracy criterion.

Figure 3A plots the cumulative frequency of participants in each condition who reached the 70% accuracy criterion for the first time in each of the 10 bins of trials. Figures 3B and 3C show the same plots for the 80% and 90% accuracy criteria, respectively. Consistent with the idea that participants look for a rule that allows them to classify the items, participants in all four conditions reached the 70% accuracy criterion at about the same rate. Furthermore, the proportion of participants reaching this criterion by the 10th bin of blocks is about the same for all conditions. Indeed, the highest proportion of participants reaching this criterion is in the four-manifestations condition, which exhibited the worst overall learning performance.

In contrast, the patterns for the 80% and 90% criteria are different from that of the 70% criterion (although they are similar to each other). Participants in the one- and two-manifestations conditions reached these learning criteria faster than those in the three- and four-manifestations conditions. Furthermore, more participants reached these criteria by the 10th bin of blocks in the one- and two-manifestations conditions than in the three- and four-manifestations conditions. These data are consistent with the hypothesis that people learn these categories by hypothesis testing.

Table 2
Summary of the Results of Experiment 1

No. of feature manifestations	Summary of feature values	Blocks to criterion *M*	Blocks to criterion *SD*	Proportion reaching criterion
1	4|5	22.32	8.85	.50
2	34|56	21.61	9.72	.54
3	234|567	26.46	6.67	.29
4	1234|5678	29.14	3.09	.14

Discussion

These results are compatible with those of Yamauchi and Markman (2000) and Medin et al. (1983), who found that increasing the number of manifestations of the features increased the difficulty of learning these categories by classification. These results further suggest that participants in all manifestation conditions were fast to reach a 70% accuracy criterion (which they could achieve using a simple unidimensional rule). The differences among the manifestation conditions were most apparent for the more stringent 80% and 90% accuracy criteria.

This experiment differed from previous research in that the variation among the manifestations was relevant to classification. Specifically, the darkness of the colors distinguished between the two categories, and the number of manifestations was implemented by increasing the number of levels of darkness that participants were shown during training.

This result is surprising in view of research by Ashby and Maddox (1992; Maddox & Ashby, 1993), who demonstrated that people can learn categories in which exemplars are drawn from a bivariate normal distribution. As discussed, however, these categories differ from the family resemblance structure with exception features in that feature values diagnostic of one category rarely occur in exemplars of the other category. We address this issue in Experiment 2.

Experiment 2

To explore the influence of this difference between stimulus sets, we generated a new family resemblance structure (Table 3). Three possible feature values are denoted *A*, *B*, and *C*. Category *A*

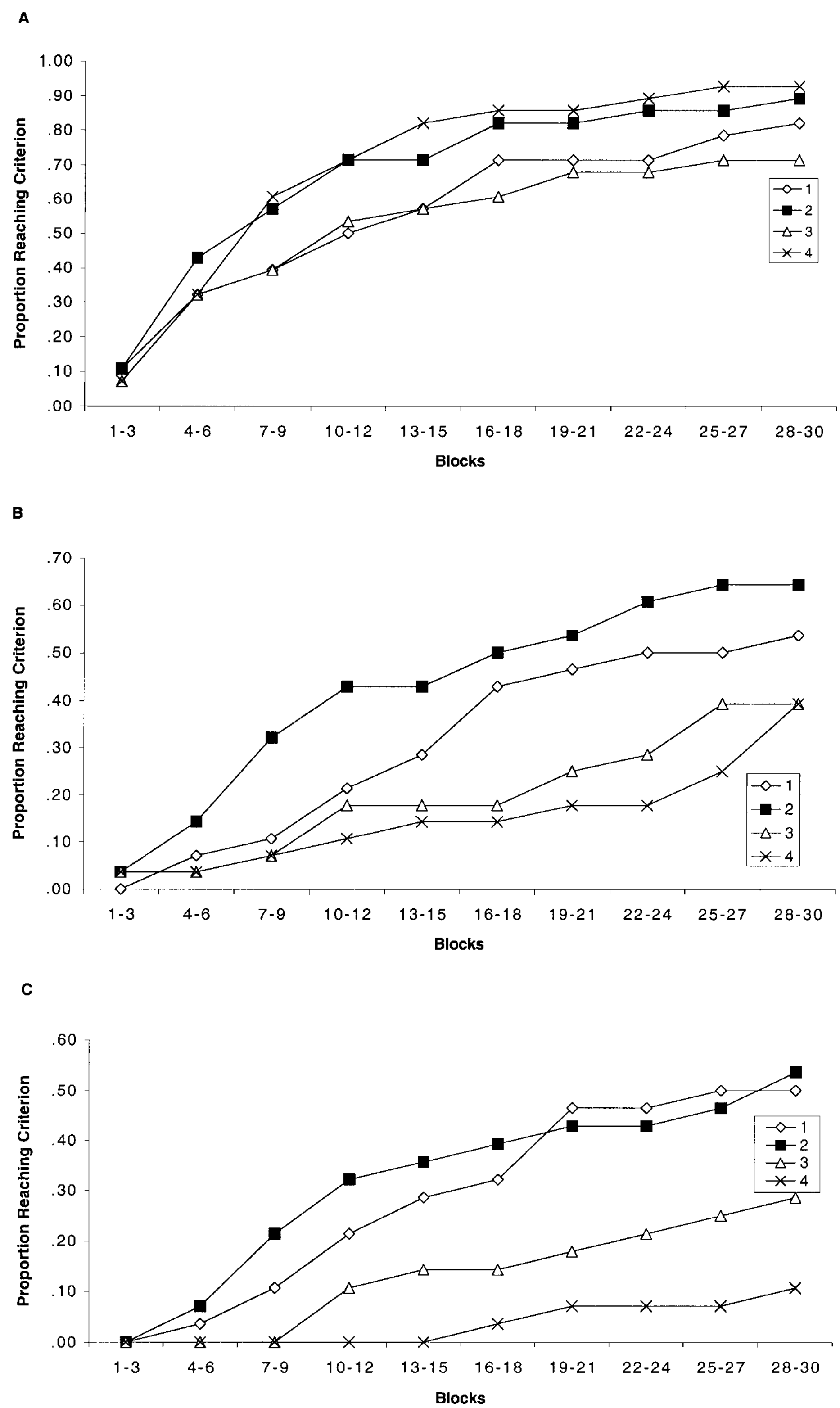

Figure 3. Graphs of the cumulative proportion of participants reaching learning criteria in Experiment 1. A: 70% accuracy criterion; B: 80% accuracy criterion; C: 90% accuracy criterion.

Table 3
Family Resemblance Feature Structure With Nondiagnostic Values

Exemplar	Dimension 1	2	3	4	Category
A1	A	A	B	B	A
A2	A	B	B	A	A
A3	B	B	A	A	A
A4	B	A	A	B	A
A5	A	B	A	B	A
A6	B	A	B	A	A
B1	C	C	B	B	C
B2	C	B	B	C	C
B3	B	B	C	C	C
B4	B	C	C	B	C
B5	C	B	C	B	C
B6	B	C	B	C	C

has the value *A* on two of four dimensions of each exemplar. Category *C* has the value *C* on two of four dimensions of each exemplar. In addition, each exemplar has the value *B* on two of four dimensions. This value is nondiagnostic because it appears equally often in each category. As discussed, this category structure with nondiagnostic values may give rise to relatively more holistic processing than does the exception feature structure used in Experiment 1. We contrast performance on these two category structures in Experiment 2.

The values for *A* and *C* were randomly assigned to light or dark shades of gray for each dimension for each participant. The value *B* was always assigned values 4 and 5 of the eight levels of darkness. Thus, participants had to learn that moderate values of gray were not diagnostic in this task, but that light or dark values could be used to distinguish among the categories. Furthermore, this particular category structure was used because it requires attending to at least three dimensions to correctly classify all of the exemplars. The exception feature structure used in Experiment 1 also requires attending to at least three dimensions.

Although we are able to equate the number of dimensions to which participants must attend, it is not possible to equate the complexity of the rule needed to classify all exemplars correctly. In the structure with nondiagnostic values, a simple disjunctive rule (e.g., If {Dim 1 = *A* or Dim 2 = *A* or Dim 3 = *A*}, then category = *A*) will suffice. In contrast, in the structure with exception features, a parity rule on three dimensions is required (e.g., If {[Dim 1 = *A* and Dim 2 = *A*] or [Dim 1 = *A* and Dim 3 = *A*] or [Dim 2 = *A* and Dim 3 = *A*]}, then category = *A*). This difference in complexity suggests that the nondiagnostic feature structure might be easier to learn overall than the exception feature structure, but it does not predict that these structures should differ in the influence of the number of manifestations of the features.

In Experiment 2, the number of manifestations was varied by starting with the values adjacent to the neutral values (3 and 6). Because there are only eight levels of darkness overall, we can explore one, two, and three manifestations. These conditions were run both for the nondiagnostic feature structure in Table 3 as well as the exception feature structure used in Experiment 1 (see Table 1). For the exception feature structure, the one-manifestation condition also used the values 3 and 6, with higher manifestation conditions adding values further from the boundary between light and dark values.

This implementation of the manifestation conditions will help to rule out an alternative explanation for Experiment 1. The values 4 and 5, which formed the boundary between light and dark values in Experiment 1, are fairly close together. Thus, it is possible that when new darker and lighter values were added, the boundary between light and dark values became hard to find. On this view, the multiple-manifestations conditions were more difficult because of confusions about the boundary between light and dark values. In Experiment 2, the light and dark values are quite discriminable. If increasing the number of manifestations still makes the classification task more difficult for the exception feature structure, then the results of Experiment 1 probably do not reflect difficulties in discriminating between the light and dark values.

Finally, one additional change to the method was made. There are 12 stimuli in the nondiagnostic feature condition, so each block for this structure requires 12 trials. To equate the number of trials in each block for each structure, the exemplars of one category for the exception feature structure were arbitrarily repeated twice in each block. Thus, this category was always presented twice as often as the other. If anything, this change should ease learning of the exception feature stimuli.

For the exception feature stimuli, the predictions are clear. As we found in Experiment 1, increasing the number of feature manifestations should increase the difficulty of the learning task. The predictions for the nondiagnostic feature structure are open. One possibility is that this structure will act like the exception feature structure, in which case learning will be easier when there are few manifestations than when there are many. A second possibility is that the opposite pattern will be obtained. As discussed previously, as the number of manifestations increases, the category centroids get further apart, and so the categories become more discriminable. If we are correct in assuming that the nondiagnostic feature structure will promote holistic processing, then increasing the number of feature manifestations will ease learning.

Method

Participants. Seventy-four undergraduates at the University of Texas received course credit or were paid for their participation.

Design. This experiment is a 2 (category structure: exception feature vs. nondiagnostic) × 3 (manifestations: 1, 2, and 3) between-subjects factorial design. Each condition had 12 participants except for the one-manifestation condition in the nondiagnostic structure and the two-manifestation condition in the exception feature structure, both of which had 13.

Materials. The exception feature stimuli were the same as those used in Experiment 1. For every participant, the values *A* and *B* on each dimension were assigned randomly to light or dark. In a change from Experiment 1, the single-manifestation condition used the values 3 and 6 for light and dark values, respectively. The two-manifestation condition used the values 2 and 3 for light and 6 and 7 for dark. Finally, the three-manifestation condition used the values 1, 2, and 3 for light and 6, 7, and 8 for dark. Each block for these materials consisted of 12 trials: two repetitions of the exemplars of category A and one repetition of the exemplars of category B.

The nondiagnostic feature structure is shown in Table 3. The manifestation conditions were implemented in the same way as they were for the exception feature structure. In this case, the values *A* and *C* for each dimension were assigned randomly to light or dark. The nondiagnostic feature values (B) were given the value of 4 or 5 (selected randomly for each nondiagnostic feature for each trial).

Procedure. The procedure in this study was identical to that in Experiment 1, except there were 12 trials in each block. The accuracy criterion was at least 32 correct trials in three consecutive blocks (i.e., 36 trials).

Results

The mean number of blocks to reach criterion for each condition was analyzed in a 2 (category structure) × 3 (manifestations) ANOVA. These data are presented in Table 4. The ANOVA revealed a significant main effect of category structure, $F(1, 68) = 17.75$, $p < .001$, reflecting the fact that participants required fewer blocks on average to learn the categories with nondiagnostic features ($m = 12.27$) compared with those with exception features ($m = 21.41$).

This effect must be interpreted in view of a significant interaction between category structure and manifestations, $F(2, 66) = 5.82$, $p < .01$. This interaction reflects the fact that increasing the number of feature manifestations made the classification task more difficult for participants given the exception feature structure but easier for those given the nondiagnostic feature structure. For stimuli with one manifestation of each feature, participants given the exception feature structure required nonsignificantly fewer blocks to reach the learning criterion ($m = 16.08$) than those given the nondiagnostic feature structure ($m = 17.38$), $t(23) = 0.29$, $p > .10$. In contrast, for the categories with two manifestations of the features, participants given the exception feature structure required significantly more blocks to reach the learning criterion ($M = 23.46$) than did those given the nondiagnostic feature structure ($M = 9.33$), $t(23) = 4.05$, $p < .01$. Likewise, when the stimuli had three manifestations of the features, participants given the exception feature structure required significantly more blocks to reach the learning criterion ($M = 24.50$) than did those given the nondiagnostic feature structure ($M = 9.67$), $t(22) = 4.34$, $p < .01$.

As we found in Experiment 1, the mean number of blocks to reach the criterion is paralleled by the proportion of participants who reached the learning criterion within 30 blocks. For participants given the exception feature structure, the proportion of participants reaching the learning criterion decreased with the number of manifestations, with .73, .42, and .33 reaching the criterion in the one-, two-, and three-manifestations conditions, respectively.[2] In contrast, the proportion of participants given the nondiagnostic feature structure who reached the learning criterion increased with the number of manifestations, with .62, .92, and .92 reaching the criterion in the one-, two-, and three-manifestations conditions, respectively.

To explore the learning data more carefully, we performed learning curve analyses similar to those for Experiment 1. Figure 4 plots histograms of the cumulative proportion of participants in each condition who reached 70%, 80%, and 90% criteria within bins of three blocks (36 trials). For the 70% criterion the pattern is similar to what we observed in Experiment 1. The curves are not well differentiated by condition except for an overall tendency for participants to reach this criterion faster given the nondiagnostic feature structure compared with the exception feature structure. For the 80% and 90% criteria, the pattern is somewhat different. For the exception feature structure, there is a clear difference between the single-manifestation condition and the multiple-manifestations conditions by the 80% criterion. In contrast, for the nondiagnostic feature structure, the curves for the three-manifestations condition are closer together. By the 90% criterion, there is a separation between the single-manifestation condition and the two- and three-manifestations conditions for both the exception feature structure and the nondiagnostic feature structure.

Table 4
Number of Blocks to Reach the Learning Criterion by Participants in Experiment 2

	No. manifestations					
	1		2		3	
Feature structure	*M*	*SD*	*M*	*SD*	*M*	*SD*
Exception	16.08	10.78	23.46	9.85	24.50	8.92
Nondiagnostic	17.38	10.95	9.33	7.29	9.67	7.81

Discussion

The key result of this experiment is that the influence of a larger number of manifestations was different for the nondiagnostic feature structure than for the exception feature structure. As before, increasing the number of feature manifestations increased the difficulty of learning for participants given the exception feature structure. In contrast, increasing the number of feature manifestations made the nondiagnostic feature structure easier. This ease reflects that the way the manifestations were added makes the categories more discriminable as the number of manifestations increases.

Because this feature structure is not often used, we performed another study testing 133 participants in the nondiagnostic feature structure.[3] In this study, participants given one manifestation per feature value learned the categories in an average of 17.53 blocks, those given two manifestations learned in an average of 14.60 blocks, and those given three manifestations learned in an average of 10.81 blocks. Thus, we replicated the finding that learning categories with this nondiagnostic feature structure becomes easier as the number of manifestations increases.

Analysis of individual participants' data suggests that they may be using different strategies to learn these categories. As discussed, simple unidimensional rules correctly classify 75% of the items. These data were consistent with the possibility that participants given the exception feature structure were performing hypothesis testing. On this view, participants quickly settled on a unidimensional rule in all conditions that provided fairly good performance. Participants in the two- and three-manifestations conditions had more difficulty exceeding an 80% criterion than those in the

[2] The proportion of participants reaching the learning criterion is somewhat higher in Experiment 2 than in Experiment 1; however, Experiment 2 extended for 360 trials before terminating, whereas Experiment 1 ended after 240 trials.

[3] Within this group, two different sets of instructions were tested. There was no reliable influence of this instructional manipulation on performance.

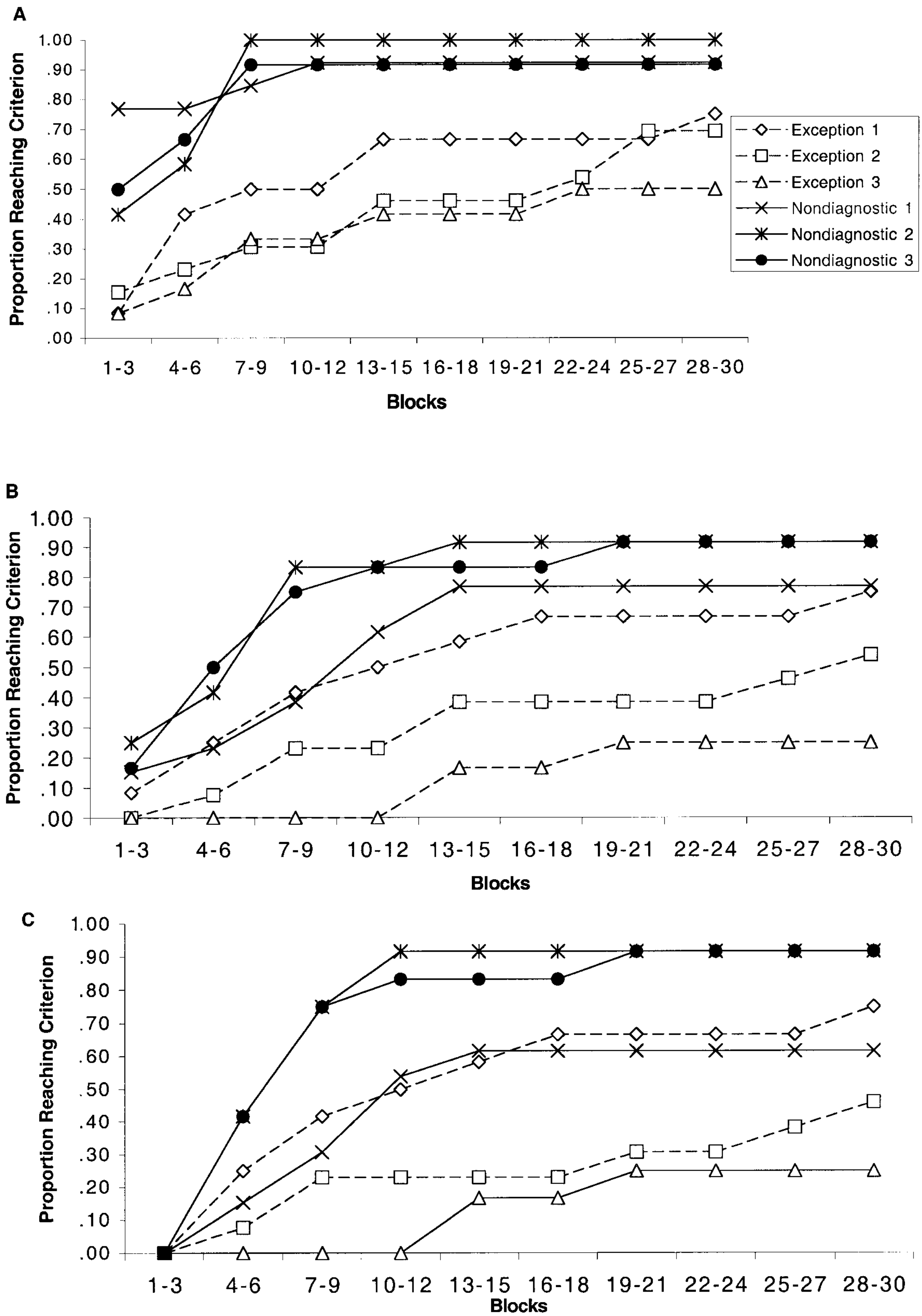

Figure 4. Graphs of the cumulative proportion of participants reaching learning criteria in Experiment 2. A: 70% accuracy criterion; B: 80% accuracy criterion; C: 90% accuracy criterion.

one-manifestation condition. This result is consistent with the observation that the space of possible rules increases with the number of manifestations.

In contrast, participants given the nondiagnostic feature structure exceeded the 70% and 80% criteria fairly rapidly in all conditions. If these participants were starting with unidimensional rules, then there should have been a larger disparity among the curves for the 80% criterion. Instead, consistent with the idea that the categories in the single-manifestation condition are less discriminable than those in the two- and three-manifestations conditions, achieving 90% accuracy was harder for participants in the one-manifestation condition than in the two- and three-manifestations conditions. Thus, these data are compatible with the prospect that participants were using a similarity-based process when given the nondiagnostic feature structure.

General Discussion

In these studies, we explored two factors that might influence the observation that learning categories by inductive classification becomes more difficult as the number of manifestations of the features of those categories increases. First, we examined variability along values relevant to classification as opposed to the variation along aspects irrelevant to classifying that were used in the past. Even when the feature manifestations varied along dimensions relevant to classification, however, the difficulty of learning still increased with the number of feature manifestations.

In contrast, the type of category structure learned had a significant impact on the relationship between number of manifestations and ease of learning. Previous research examining the number of feature manifestations used family resemblance category structures with exception features in which exemplars had some features that were diagnostic of the other category. In Experiment 2, we introduced a category structure with nondiagnostic features that were equally predictive of both categories. No exemplar had feature values that were actually diagnostic of the other category. With this category structure, increasing the number of manifestations actually made classification easier.

It is important to note that classification gets easier as the number of manifestations increases because new manifestations were added farther from the boundary between light and dark values. Thus, in our studies the categories became more perceptually discriminable as more manifestations were added. It is this aspect of the stimulus design that makes it surprising that increasing the number of feature manifestations makes classification learning harder for participants given the exception feature structure.

Accounting for These Findings

The pattern of data obtained in these two studies is not clearly predicted by any of the major classes of categorization models. To clarify this point, we briefly discuss prototype models, exemplar models, rule-based models, and decision-bound models. Finally, we speculate on the potential role of rule- and similarity-based processes in these category structures.

Prototype models assume that categories are represented by some central tendency that reflects average or typical values along each dimension (e.g., Posner & Keele, 1970; Reed, 1972). For example, after repeated trials, we might expect the two categories in the current research to be represented by the average darkness values along each dimension. Classifying a new instance involves selecting the prototype to which the new exemplar is most similar. Because of the way new manifestations were added in this study, the perceptual discriminability of the categories increases with the number of manifestations for both exception feature and nondiagnostic feature structures. Thus, prototype models would predict that increasing the number of manifestations would ease category learning. This result was obtained only for the nondiagnostic feature structure.

Decision-bound models (Ashby & Maddox, 1992, 1993; Ashby & Perrin, 1988) assume that people seek a multidimensional rule that guides the classification decision in the face of perceptual and criterial noise. Decision-bound models predict that classification performance should improve as category discriminability increases. Thus, these models also predict the results from the nondiagnostic feature structure correctly but not the results from the exception feature set. It is not surprising that these models give the correct prediction for the nondiagnostic structure, because this structure was modeled on the kinds of materials used in studies testing decision-bound models.

Rule-based models do not account for the whole pattern of data obtained here either. Some research has examined how the complexity of rules influences category acquisition (e.g., Feldman, 2000). This work suggests that categories described by more complex rules are harder to learn than categories described by simple rules. On this view, the exception feature structure should be harder to learn than the nondiagnostic feature structure (as we observed). However, there is no obvious basis for predicting that these structures should be influenced in different ways by changes in the number of manifestations.

If rules are learned through a process of hypothesis testing, then a rule-based account could predict that increasing the number of manifestations would make category learning more difficult, because the space of possible rules increases with the number of manifestations. As the number of manifestations increases, the learner must make conjectures both about which dimensions should be incorporated into a rule and also about the location of the boundary between values diagnostic of each category. Because this rule space increases in size, the average time to search this space for a good rule is also likely to increase. This explanation is compatible with the data from the exception feature structure but not the nondiagnostic feature structure.

Finally, exemplar models can fit data in which increasing the number of manifestations increases the difficulty of learning, decreases the difficulty of learning, or has no effect on learning. We conducted simulations using Nosofsky's (1986) generalized context model (GCM) to confirm the predictions of exemplar models, but the basic intuition underlying the predictions is straightforward. The similarity scaling parameter, c, in the GCM determines the effect of the similarity between the new exemplar and known exemplars. For large values of the c parameter, only known exemplars that are very similar to the new exemplar influence classification. In the limit, this parameter can be set so that only known exemplars that are identical to the new one will influence classification. In this case, if there are no identical exemplars in memory, the model will guess. For this setting of the similarity scaling parameter, the model is basically a look-up table with guessing. This version predicts that classification will become more difficult as the number of manifestations increases, because increasing the number of manifestations decreases the likelihood that a new exemplar will be identical to one seen before.

For small values of this scaling parameter, a variety of similar exemplars can be used to influence classification. Because the categories used in these studies have a family resemblance structure, and the exemplars of the categories are generally more discriminable as the number of feature manifestations increases, this parameter setting predicts that classification will become easier as the number of manifestations increases. Finally, given that these patterns are obtained by varying a single parameter, there are also intermediate values of the parameters for which changes in the number of manifestations have no effect on classification. Although exemplar models are able to fit any of these patterns of data, there is no principled reason for the value of this scaling parameter to be different for the exception feature structure than it is for the nondiagnostic feature structure. Exemplar theorists might want to explore this issue in more detail.

It is possible that the best model of our findings will combine more than one approach to categorization. Support for this possibility comes from our examination of the speed with which participants reached 70%, 80%, and 90% accuracy. On the basis of these data, we speculated that participants given the exception feature structure may be forming rules, whereas those given the nondiagnostic structure may be performing a more similarity-based process. Although we find this possibility interesting, two factors limit our confidence in this explanation. First, there is no principled reason why people should engage in more rule-based processing for the exception feature structure than for the nondiagnostic feature structure. It may be that they find it strange to encounter the presence of features that are highly diagnostic of one category in exemplars of another category, but that does not explain why this structure promotes rule-based learning.

Second, in the replication of the nondiagnostic feature structure mentioned in the Discussion section of Experiment 2, we contrasted two instruction conditions. In one, people were asked to learn the categories by forming a rule. In the second, they were asked to learn the categories by looking for overall similarities among items. This instruction manipulation had no appreciable effect on performance. This null effect may reflect that people are not able to implement a particular learning strategy just because they are instructed to do so. Further research must explore reasons for the difference between the two category structures.

To summarize, extant models of category learning do not provide a good explanation for the pattern of data in the current study. Prototype and decision-bound models predict that increasing the number of manifestations should decrease the difficulty of learning, as was observed for the nondiagnostic feature structure. Rule-based models predict that increasing the number of manifestations should increase the difficulty of learning, as was observed for the exception feature structure. Finally, exemplar models account for either pattern of data depending on how similarity is scaled, but there is no principled reason why the feature structures used in these studies should differ in the scaling of similarity.

Implications for Categorization Research

These results raise a significant concern for laboratory studies of categorization. The inductive classification task is used extensively in research because it is assumed to provide a good analogue to the types of learning that people perform when acquiring natural categories. Thus, findings like those of Medin, Dewey, and Murphy (1983) and Yamauchi and Markman (2000) are troubling because they suggest that a fairly small extension of the classification task yields behavior that is incompatible with people's manifest ability to learn natural categories despite the variations in the manifestations of their features.

The current results suggest that the use of exception feature category structures should be scrutinized. Many laboratory tasks use category structures in which features have two possible values, some of which are diagnostic of one category and some of which are diagnostic of the other. The results of Experiments 1 and 2 suggest that exception feature structures may induce a mode of processing that is incompatible with the presence of multiple-feature manifestations. In the interest of improving the fit between the laboratory and the world, more consideration should be given to category structures that involve nondiagnostic values rather than exception features.

References

Ashby, F. G. (Ed.). (1992). *Multidimensional models of perception and cognition.* Hillsdale, NJ: Erlbaum.

Ashby, F. G., & Maddox, W. T. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 50–71.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar and decision bound models of categorization. *Journal of Mathematical Psychology, 37,* 372–400.

Ashby, F. G., & Maddox, W. T. (1998). Stimulus categorization. In M. H. Birnbaum (Ed.), *Measurement, judgment, and decision making* (pp. 251–301). San Diego, CA: Academic Press.

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

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

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature, 407,* 630–632.

Lassaline, M. E., & Murphy, G. L. (1996). Induction and category coherence. *Psychonomic Bulletin and Review, 3,* 95–99.

Lockhead, G. R. (1979). Holistic versus analytic process models: A reply. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 746–755.

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

Maddox, W. T., & Bohil, C. J. (1998). Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1459–1482.

Maddox, W. T., & Dodd, J. L. (2001). On the relation between base-rate and cost-benefit learning in simulated medical diagnosis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1367–1384.

Medin, D. L., Dewey, G. I., & Murphy, T. D. (1983). Relationships between item and category learning: Evidence that abstraction is not automatic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 607–625.

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

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

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–97.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83,* 304–308.

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

Smith, J. D., & Kemler-Nelson, D. G. (1984). Overall similarity in adults' classification: The child in all of us. *Journal of Experimental Psychology: General, 113,* 137–159.

Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review, 96,* 125–144.

Yamauchi, T., & Markman, A. B. (2000). Learning categories composed of varying instances: The effect of classification, inference, and structural alignment. *Memory & Cognition, 28,* 64–78.

Received September 14, 2001
Revision received July 18, 2002
Accepted July 18, 2002 ■