Journal of Experimental Psychology:
Learning, Memory, and Cognition
2002, Vol. 28, No. 3, 585–593

0278-7393/02/$5.00 DOI: 10.1037//0278-7393.28.3.585

Learning Nonlinearly Separable Categories by Inference and Classification

Takashi Yamauchi
Texas A&M University

Bradley C. Love and Arthur B. Markman
University of Texas at Austin

Previous research suggests that learning categories by classifying new instances highlights information that is useful for discriminating between categories. In contrast, learning categories by making predictive inferences focuses learners on an abstract summary of each category (e.g., the prototype). To test this characterization of classification and inference learning further, the authors evaluated the two learning procedures with nonlinearly separable categories. In contrast to previous research involving cohesive, linearly separable categories, the authors found that it is more difficult to learn nonlinearly separable categories by making inferences about features than it is to learn them by classifying instances. This finding reflects that the prototype of a nonlinearly separable category does not provide a good summary of the category members. The results from this study suggest that having a cohesive category structure is more important for inference than it is for classification.

Categories are crucial for a variety of cognitive processes, including analogy, causal reasoning, memory, imagination, creativity, generalization, and prediction. Most research on category acquisition has focused on how category representations are formed, often using a variant of an inductive classification task. A typical assumption in this research is that the category representations acquired through classification learning support the many uses of categories. There has been a growing movement, however, to examine how the variety of category uses influence what is learned (Brooks, 1978; A. B. Markman & Makin, 1998; A. B. Markman, Yamauchi, & Makin, 1997; Ross, 1996, 1997, 1999; Yamauchi & Markman, 1995, 1998, 2000a, 2000b, 2000c). This work suggests that the representations formed during learning depend in important ways on the task carried out during learning.

Two central uses of categories that have been explored in previous research are classification and predictive inference. Classification is clearly an important function of categories. An item must be classified before category knowledge can be applied. Accordingly, human classification performance has been studied extensively (e.g., Medin & Schaffer, 1978; Nosofsky, 1986; Shepard, Hovland, & Jenkins, 1961). Inference is also a critical function of categories. One reason we have categories is to support predictions. For example, once we know a politician's party affiliation we can infer his or her view on a number of issues. Inference is so central to categorization that Anderson (1991) based his rational model of category learning on the assumption that categories are formed to maximize people's ability to make accurate predictive inferences. Consistent with the importance of inference, a number of studies have been directed at the way categories are used to make predictions (e.g., Heit & Rubinstein, 1994; Lassaline, 1996; Malt, Ross, & Murphy, 1995; Murphy & Ross, 1994; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Rips, 1975; Ross & Murphy, 1996; Yamauchi & Markman, 2000a).

In this article, we explore differences between categories learned by classification and categories learned by making predictive inferences (see Yamauchi & Markman, 1995, 1998, 2000b). In the next section, we describe previous research on this topic, and isolate a prediction that has not yet been tested. In brief, the results of previous research with cohesive, linearly separable category structures are consistent with the hypothesis that inference focuses participants on summary information (such as prototypes), whereas classification learning focuses participants on information that discriminates between the categories (Yamauchi & Markman, 1998). Experiments 1 and 2 provide further support for this hypothesis using nonlinear category structures. Whereas previous research with linear categories demonstrated advantages for inference learning (further discussed below), the nonlinear category structures explored here should (and do) display advantages for classification learning.

Inference, Classification, and Category Labels

Categories are formed in the process of interacting with category instances. The information that is salient for carrying out a particular task may influence what is learned about a category. Nevertheless, it may seem surprising that inference and classification differ in the category representation that they generate, because these two tasks are very similar.

Takashi Yamauchi, Department of Psychology, Texas A&M University; Bradley C. Love and Arthur B. Markman, Department of Psychology, University of Texas at Austin.

Part of the study introduced in this article was presented at the 22nd Annual Conference of the Cognitive Science Society, Philadelphia, August 2000. This research was supported by Texas A&M University Faculty Mini-Grant FMG-R1-037 awarded to Takashi Yamauchi and by National Science Foundation Grant SBR-9905013 awarded to Arthur B. Markman. We thank Rob Goldstone, Ryan Gossen, and Hunt Stilwell. We also thank Douglas Medin for comments on part of this article. Special thanks go to Tiki Barber and Ron Dayne for making up a super backfield tandem, as well as to Mary T. for teaching Bradley C. Love Pat's trick.

Correspondence concerning this article should be addressed to Takashi Yamauchi, Department of Psychology, Texas A&M University, College Station, Texas 77843. E-mail: tya@psyc.tamu.edu

In the classification task typically used in studies of category acquisition, the participant is shown an exemplar consisting of values along a set of feature dimensions. For example, each exemplar might have four dimensions, and the value along each binary dimension might be denoted as a *1* or a *0* (e.g., a *1* could denote a triangle, whereas a *0* could denote a circle). Thus, the configuration *1 0 0 0* might describe a particular exemplar with four feature dimensions (e.g., form, size, color, position). The participant's task is to predict the category label of this stimulus. If there are two possible labels, then these may also be denoted as *1* or *0*. Thus, the sample exemplar discussed here could be schematized as a five-dimensional stimulus with the values *1 0 0 0 ?* (e.g., form, size, color, position, category label), where the question mark signals an unknown value. On all trials in a classification task, the category-label dimension is unknown and must be predicted. After category membership is predicted, feedback that displays all stimulus features (e.g., *1 0 0 0 0*) is provided.

The predictive inference task used in previous research has a similar structure (Yamauchi & Markman, 1995, 1998, 2000b). Participants are told the category label and the values of all feature dimensions but one, and are asked to predict the value of the missing dimension. For example, on one trial, the participant might see a stimulus with the abstract structure *1 ? 0 0 0* and predict the value of the missing feature (in this case, the category label is the fifth dimension and the second dimension is being predicted). After predicting the unknown feature value, feedback is provided with all stimulus features (e.g., *1 0 0 0 0*). Across trials, the unknown stimulus dimension that must be predicted varies, though the category label is always known. One should notice that classification and inference learning present the same stimulus information to the learner (i.e., the complete stimulus description is shown after feedback).

Assuming the category label is just another feature dimension, each trial of the inference task is equivalent to the classification task. On this view, the primary difference between these tasks is that classification requires a prediction of the category label on each trial, whereas inference requires a prediction of a value for a different feature dimension on each trial. In both cases, the feedback at the end of each trial presents a complete stimulus with the category label.

Despite this similarity in task structure, there is some reason to expect that the two tasks will differ. In particular, category labels may play a larger role in organizing conceptual information than other features (Yamauchi & Markman, 2000a). For example, Gelman and Markman (1986) found that four-year-old children are capable of using category labels for induction even when two items that bear the same label appear very different. E. M. Markman and Hutchinson (1984) also found that two- to three-year-old children are able to grasp categorical implications associated with a count noun. When an object is referred to with an arbitrary noun (e.g., "This is a dax. Find another dax."), children tend to group objects from the same category. In contrast, when an object is referred to with an indexical expression (e.g., "Find another one that is the same as this."), children are likely to group objects thematically. Yamauchi and Markman (2000a) obtained a similar result and found that a description associated with a category is treated like a label when the value is related to the whole exemplar (e.g., dax means that a bug is poisonous), but like a feature when it is related to a specific dimension of an exemplar (e.g., dax means that a bug has a poison needle). Finally, Gelman and Heyman (1999) found that children treat novel personality traits ascribed to a person in the form of a label (e.g., "He is a carrot-eater") as more enduring than traits ascribed to a person in the form of a feature (e.g., "He eats carrots").

These psychological differences between category labels and category features lead to systematic distinctions between what is learned about categories during inference and classification. On classification trials, because the category label must be predicted, participants seek information that divides items into groups. This information might be simple rules (and exceptions; e.g., Nosofsky, Palmeri, & McKinley, 1994), whole exemplars (e.g., Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986), or prototypes (e.g., Reed, 1972). Because the goal is to separate items into different classes, people may seek the type of information that helps segregate exemplars with different category labels, even when there is little within-category similarity.

In contrast, predictive inference is strongly influenced by the consistency between the category label and the category features. Because category labels may point to some deep attribute that is shared by the members of that category (Medin & Ortony, 1989), people do not want to make inferences that violate the value most strongly associated with the category. Thus, inference directs people's attention away from individual exemplars and towards the features common to the category.

Previous Research Related to Inference and Classification Learning

Yamauchi and Markman (1998) used linearly separable categories and investigated the impact of inference and classification on category acquisition. Their categories consisted of four exemplars each. Each exemplar differed on one stimulus dimension from an underlying category prototype (see Table 1). Thus, the categories had a linearly separable family resemblance structure (i.e., exemplars could be classified with an additive summation of feature values). Yamauchi and Markman's (1998) experiments consisted of a learning phase and a transfer phase. During the learning phase, the participant learned these categories by responding either to classification questions or to inference questions (see Figure 1). The participant received feedback immediately after responding, and was expected to learn the categories incrementally by a process of trial and error. The participant given classification ques-

Table 1
Linearly Separable Categories

	Set A						Set B				
	F	S	C	P	Ca		F	S	C	P	Ca
A1	1	1	1	0	1	B1	0	0	0	1	0
A2	1	1	0	1	1	B2	0	0	1	0	0
A3	1	0	1	1	1	B3	0	1	0	0	0
A4	0	1	1	1	1	B4	1	0	0	0	0
A0	1	1	1	1	1	B0	0	0	0	0	0

Note. A1–A4 are the exemplars of Set A, and B1–B4 are the exemplars of Set B. These exemplars are obtained from two prototypes, A0 and B0. F = form; S = size; C = color; P = position; Ca = category label.

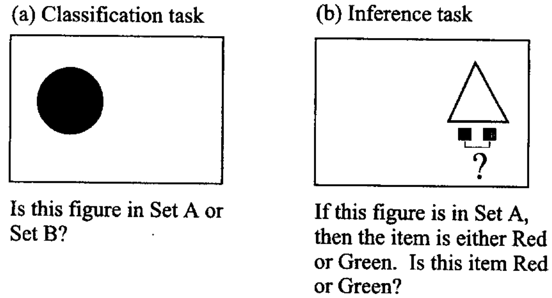

Figure 1. Panel A: A stimulus frame for a classification trial. In a particular classification trial, a participant is given a figure whose form, size, color, and position are specified. Then, the participant is asked to predict the category label (Set A or Set B) of the stimulus. Panel B: A stimulus frame for an inference trial. In a particular inference trial, a participant is given a figure whose form, size, position, and category label are specified. Then, the participant is asked to predict the color of the item. The participant predicts different feature values in different trials.

tions predicted the category label of a stimulus. The participant given inference questions predicted a missing feature value, given the category label and other stimulus feature information. The learning phase continued until the participant met an accuracy criterion. After the learning phase, all participants carried out the same transfer tasks involving both classification questions and inference questions for old stimuli that appeared during the learning phase and for new stimuli that did not appear during the learning phase.

There were three critical findings from this work. First, given the linearly separable category structure, participants given inference learning required significantly fewer trials to reach the learning criterion than did participants given classification learning. Second, for the inference transfer questions, participants given inference learning were more likely to respond with the feature values that were consistent with the prototype of the target category than were participants in classification learning. Third, a final experiment was conducted in which participants were given a block of classification learning and a block of inference learning with the order varied between participants. Participants who learned categories starting with inference followed by classification reached the learning criteria faster than did participants who did the tasks in the opposite order (with classification learning first). This result suggests that category representations formed through inference learning helped classification (presumably because participants were learning the category prototype), but that category representations learned by classification did not help inference to the same extent (presumably because participants were only learning enough about each category to distinguish between the categories).

All of these results indicate that inference and classification differ in the strategies they evoke, and hence in what participants learn when carrying them out. Classification leads people to extract diagnostic attributes that can divide items into groups; inference guides people to find some abstract attributes that can integrate the members in a category. This distinction suggests that categories that lack integral structure will be poorly learned by inference. In contrast, these categories should still be learnable by classification as long as people can memorize specific exemplars (e.g., Medin & Schaffer, 1978).

We present two studies to explore this hypothesis. In each study, we ask people to learn a pair of nonlinearly separable categories that have a low level of within-category similarity. The two different structures used in the two experiments allow us to address specific issues about the type of information incorporated into category representations. In the case of nonlinear categories, prototypes provide a poor summary of categories. Thus, the general prediction is that inference learning will be very difficult with the nonlinearly separable categories, but that classification learning will be relatively easy.

Experiment 1

To test the ease of learning nonlinearly separable categories by inference and classification, we used geometric figures as stimuli (see Figure 1). All the stimuli varied along five binary feature dimensions: four perceptual dimensions—size (large, small), form (circle, triangle), position (left, right), and color (red, green)—and one category-label dimension (Set A, Set B). These geometric stimuli were like those used by Medin and Schaffer (1978) and by Yamauchi and Markman (1998). The category structure is shown in Table 2. The prototypes of these categories are Stimuli A1 (with the value of 1 on all the dimensions) and B1 (with the value 0 on all the dimensions). The category structure (shown in Table 2) was identical to the one used by Medin and Schaffer.

In Experiment 1, the participants learned these two categories in one of two conditions: classification or inference. In the classification learning condition, the participant saw a whole stimulus and had to predict the category label. In the inference learning condition, the participant saw the category label and the values of three feature dimensions and had to predict the value of the missing

Table 2
The Category Structure of Experiment 1

	Set A						Set B				
	F	S	C	P	Ca		F	S	C	P	Ca
Learning											
A1	1	1	1	1	1	B1	0	0	0	0	0
A2	1	0	1	0	1	B2	1	0	1	1	0
A3	0	1	0	1	1	B3	0	1	0	0	0
A0	1	1	1	1	1	B0	0	0	0	0	0
Transfer											
A4	0	1	1	1	1	B4	1	0	0	0	0
A5	1	1	0	1	1	B5	0	0	1	0	0
A6	1	1	1	0	1	B6	0	0	0	1	0
						B7	0	0	1	1	0
						B8	1	1	0	0	0

Notes. A1–A3 are the exemplars of Set A, and B1–B3 are the exemplars of Set B. A0 is the prototype of Set A and B0 is the prototype of Set B. The correct answers for the inference questions are the responses that correspond to the values of the table. For example, given the inference question of the form of Stimulus A4, 0 is the correct response and 1 is the incorrect response. F = form; S = size; C = color; P = position; Ca = category label.

dimension. Initially, no information about the categories was given to participants in this study, so that they had to learn the two categories by trial and error on the basis of the feedback they received after their response. The learning phase continued until participants reached a criterion of 90% accuracy in three consecutive combined blocks (18 trials) or until they completed 30 blocks (180 trials).

Following the learning phase, we tested the nature of the acquired category representation using transfer trials. Transfer trials consisted of classification and inference questions involving stimuli that appeared during the previous learning phase as well as novel stimuli that did not appear during learning (i.e., the novel stimuli had new combinations of feature values). In the transfer phase, all the participants received the same trials.

The predictions for inference transfer are straightforward. The prototypes do not provide a good summary of these categories, and so participants should have difficulty making inferences involving the novel items introduced in the transfer phase. The predictions for classification transfer are more complex. The transfer stimuli were designed to test the hypothesis that inference focuses learners on the category prototype and that classification focuses learners on diagnostic information (and in this study, exemplar memorization). The transfer stimuli, A4–A6 and B4–B6, deviate equally from the prototype of each category, but these items differ in their similarity to individual exemplars of the categories.[1] Stimuli B4–B6 are highly similar to one exemplar in Set A and one exemplar in Set B. In contrast, Stimuli A4–A6 are highly similar to two exemplars in Set A but not similar to any of the exemplars in Set B. If participants in inference learning are extracting the prototype of the category, then they should classify these stimuli equally well. In contrast, participants in classification learning are expected to engage in exemplar memorization. Thus, they should classify Stimuli A4–A6 more accurately than Stimuli B4–B6. These manipulations and predictions are taken directly from Medin and Schaffer (1978).

A similar set of predictions can be made for Stimuli B7 (*0 0 1 1*) and B8 (*1 1 0 0*). These two stimuli are neutral with respect to the two prototypes; both have two dimension values consistent with Set A and two dimension values consistent with Set B. These items are highly similar to at least one of three exemplars of Set B (B7 is similar to B2, and B8 is similar to B3), but they are not similar to any of the exemplars of Set A. As a consequence, Stimuli B7 and B8 should be accurately classified into Set B as a function of exemplar storage during learning.

Method

Participants. Participants were 49 undergraduates at Columbia University who were paid for their participation. The data from 1 participant were lost because of an error in recording. In total, the data from 48 participants (24 in each condition) were analyzed.

Materials. Each category consisted of three exemplars that were shown during learning and transfer trials. In addition, there were eight new stimuli that were given only in the transfer phase. Two versions of the dimension value assignment were introduced in this experiment. In one version, the value of 0 was triangle and the value of 1 was circle for form. For color, the value of 0 was green and the value of 1 was red. For size, the value of 0 was small and the value of 1 was large. For position, the value of 0 was right and the value of 1 was left. In the other version, the values of form and size were reversed. Each stimulus was bounded by a 20.3-cm × 17.4-cm rectangular frame drawn with a solid black line on the computer screen.

Procedure. The experiment involved three phases: a learning phase, a filler phase, and a transfer phase. In the learning phase, participants were randomly assigned to one of two conditions: classification or inference. In the two conditions, participants continued in the learning phase until they performed three consecutive blocks with a combined accuracy of 90% or until they completed 30 blocks (180 trials). A classification block consisted of presentations of the six exemplars. In each block, the order of stimulus presentation was determined randomly.

Each inference block consisted of one inference question taken from each of the six stimuli. In each block, every exemplar in the two categories was presented exactly once. A specific question was assigned to a particular stimulus in a block to balance the number of dimension questions given during the learning phase. For example, in Block 1, questions about position, color, and size were assigned to Stimuli A1 and B1, A2 and B2, and A3 and B3, respectively. In Block 2, questions about color, size, and form were assigned to Stimuli A1 and B1, A2 and B2, and A3 and B3, respectively. Thus, there were exactly three different dimension questions in one block, which were shifted systematically in a set of four blocks. In each set of four blocks, every dimension of every stimulus was queried exactly once (24 questions in total), and each stimulus was queried exactly four times. Within each block, the order of stimulus presentation was determined randomly. Within the set of four blocks, the presentation order of each block was determined randomly.

In classification learning, participants saw one of the six stimuli and indicated the category to which it belonged by clicking a button with the mouse (see Figure 1A). In inference learning, participants inferred a value for one of the four dimensions, whereas its category label and the remaining three dimension values were depicted in the stimulus frame (see Figure 1B). Different perceptual dimensions were predicted on different trials. Participants responded by clicking one of two labeled buttons with the mouse. For each stimulus, the location of the correct choice was randomly determined. Following each response, feedback and the correct stimuli were presented on the screen for 3 s. The stimuli presented during feedback were identical in both the classification and inference tasks.[2]

After the learning trials, there was a brief filler task, and then all participants carried out the same transfer tasks. In the transfer phase, participants were first given classification transfer followed by inference transfer. The transfer stimuli consisted of six old stimuli and eight new stimuli (see Table 2). All of these stimuli were shown both in the classification transfer task and in the inference transfer task. The order of stimulus presentation for each task was determined randomly. In total, there were 14 classification transfer questions (6 questions from old stimuli and 8 questions from new stimuli), and 56 inference transfer questions (24 questions from old stimuli and 32 questions from new stimuli). The classification transfer questions asked participants to predict the category

[1] For the purposes of this discussion, we will assume that exemplars with the prefix *A* should be classified into Set A and exemplars with the prefix *B* should be classified into Set B. Thus, in this discussion, we will refer to accuracy of transfer performance. This phrasing is a simple way of talking about the likelihood of classifying an item into the category specified by the prefix. Obviously, because these are transfer items, there is no objectively "correct" response.

[2] The inference for the size of Stimuli B1 and B3 has two right answers. Given the inference questions (*0 ? 0 0*), the response of the feature value 1 corresponds to Stimulus B3 and the response of the feature value 0 corresponds to Stimulus B1. We gave participants a positive feedback irrespective of their responses for this question. This treatment should make inference learning faster, and thus functions against our hypothesis that inference learning requires more trials than classification learning for this category structure.

label of a stimulus, given information about its features. The inference transfer questions asked participants to predict the value of an unknown feature, given the category label of the stimulus and the information about the other features. The procedures for the transfer questions and the learning questions were analogous except that no feedback was given during transfer.

Results and Discussion

The basic results of Experiment 1 are consistent with our hypotheses (see Table 3). As predicted, given nonlinearly separable categories, inference learning was much more difficult than classification learning. This finding contrasts with our previous research with linearly separable categories, in which inference was easier than classification (Yamauchi & Markman, 1998).

In all, 17 participants reached the learning criterion in the inference learning condition, and 22 participants reached the criterion in the classification learning condition. Considering only those who reached the learning criterion, participants in the inference learning condition (M = 15.8) required significantly more blocks during the learning phase than did participants in the classification learning condition (M = 10.5), $t(37) = 3.32$, $p < .01$. For the classification transfer of old stimuli, participants given classification learning were significantly more accurate than participants given inference learning; $t(37) = 5.28$, $p < .01$.[3]

Consistent with the prediction that classification learning would lead to the use of exemplars, participants given classification learning classified Transfer Stimuli A4–A6 more accurately than the Transfer Stimuli B4–B6. In contrast, participants given inference learning did not show this trend. There was a marginally significant interaction between the two learning procedures (inference learning vs. classification learning) and the two sets of stimuli (A4–A6 vs. B4–B6), $F(1, 37) = 3.30$, $MSE = 0.073$, $p < .08$. Planned comparisons revealed that participants in classification learning were more accurate in classifying Stimuli A4–A6 than in classifying Stimuli B4–B6, $t(21) = 4.18$, $p < .01$ (Bonferroni adjustment). In contrast, participants in inference learning did not differ in their classification accuracy for these stimuli, $t(16) = 0.75$, $p > .10$. As previously discussed, Stimuli A4–A6 and B4–B6 are equally similar to the prototype of each category. These stimuli diverge with respect to their similarity to individual exemplars. Stimuli B4–B6 are highly similar to one exemplar in Set A and one exemplar in Set B, whereas Stimuli A4–A6 are highly similar to two exemplars in Set A, but are not similar to any of the exemplars in Set B. This finding suggests that exemplar information played a bigger role in classification learning than it did in inference learning.

We were concerned that this interaction might reflect a floor effect for transfer performance by participants given inference learning. However, participants given inference learning were able to classify the transfer items A4–A6 and B4–B6 reasonably well as compared with a chance level, $t(17) = 1.77$, $.05 < p < .10$. Furthermore, their overall classification transfer performance for these items (M = 0.59) was not significantly different from that of the participants given classification learning (M = 0.61) as evidenced by the absence of a main effect of learning condition in the analysis of these transfer items, $F(1, 37) < 1.0$.

Finally, we predicted that participants given classification learning would be more likely than participants given inference learning to classify items B7 and B8 into Set B, because these items are similar to one exemplar from Set A but none of the exemplars in Set B. As shown in Table 3, classification performance for these items was better for classification learning participants than for inference learning participants, but this difference was not statistically significant, $t(40) = 1.04$, $p > .10$.

For the inference transfer questions, the correct answers were those shown in Table 2. Participants in the two conditions were about equally accurate in making inferences for old stimuli ($t < 1$). Their performance was much lower for the new stimuli than for the old stimuli. Participants given classification learning were at chance for these items ($t < 1$), and participants given inference learning were actually significantly below chance, $t(16) = -2.36$, $p < .05$.

We investigated the source of this anomaly by grouping every new inference transfer question into two types: consistent type and inconsistent type. The consistent inference questions required a response that was in accord with the prototype of the corresponding category (e.g., the question about the form of Stimulus A5 in Table 2). The inconsistent inference questions required a response that was inconsistent with the prototype of the corresponding category (e.g., the question about the form of Stimulus A4). For participants in inference learning, the average performance for consistent questions was 0.51, whereas that for inconsistent questions was 0.26, $t(16) = 4.24$, $p < .01$. Obviously, participants in inference learning found the prediction of inconsistent-type dimensions particularly difficult.

Taken together, the results of Experiment 1 support our view that it is difficult to make inferences for nonlinearly separable categories. Furthermore, the results indicate that inference and classification, two of the main functions of categories, make use of different types of category information in their tasks. In Experiment 2, we investigated the generality of this hypothesis using a different nonlinear category structure.

Table 3
The Main Results From Experiment 1

Learning	Old	New: Average	New: A4–A6	New: B4–B6	Neutral: B7 & B8
		Classification transfer			
Classification	0.92	0.61	0.76	0.45	0.61
Inference	0.69	0.59	0.63	0.55	0.44
		Inference transfer			
Classification	0.75	0.50	0.50	0.50	0.50
Inference	0.79	0.46	0.40	0.51	0.36

Note. For the neutral stimuli B7 and B8, we measured the proportion that participants classified the two stimuli into Category B. These data are based on the participants who reached the learning criterion. "Old" means the old stimuli that were shown during the learning phase and the transfer phase. "New" means the new stimuli that were shown during the transfer phase only.

[3] The analyses in this section involve only those participants who reached the learning criterion. Analyses that include all participants yield the same pattern of data.

Experiment 2

Table 4 illustrates the structure of the two categories used in Experiment 2. These categories each consist of three exemplars. This category structure is useful for distinguishing the extent to which participants assess a summary of the category as opposed to individual exemplars. Stimulus A0 (*1 1 1 1*) summarizes Set A, and Stimulus B0 (*1 1 0 0*) summarizes Set B because these dimension values are dominant for the two categories. Participants in inference learning should have difficulty acquiring these two categories because Stimulus A2 is the prototype of Set B, but is actually a member of Set A. Given the classification test of these stimuli, participants in inference learning should be able to classify the prototype A1 but not A2. In contrast, participants in classification learning should not have trouble correctly classifying these stimuli.

Another important aspect of this experiment is an examination of the category information stored through classification learning. We have proposed that inference focuses on summary information. In contrast, there is evidence that people who are trying to classify a set of items tend to focus on diagnostic information that reliably distinguishes between categories (e.g. Nosofsky et al., 1994). This focus on stimulus information that is diagnostic for predicting category membership is also embodied in the attentional mechanisms of exemplar models (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986). The hypothesis that classification tends to focus on diagnostic stimulus dimensions, whereas inference tends to focus on summary information, is supported by some previous work (Yamauchi & Markman, 1998, 2000b). Experiment 2 examines this issue in another way.

There are two strategies that participants might use to learn these categories when classifying. One possibility is that they will focus their attention only on dimensions that are diagnostic. Given the category structure shown in Table 4, the first two dimensions (form and size in Table 4) are not useful for distinguishing the two categories because the same feature values are dominant in the two categories (the dominant values of Set A are *1 1 1 1* and the dominant values of Set B are *1 1 0 0*). In contrast, the last two dimensions (color and position in Table 4) are more informative for distinguishing between the categories. Thus, if classification promotes attention to the dimensions that differentiate the two categories, participants in classification learning should attend more to color and position than to form and size.

A second possibility is that participants who classify the stimuli will memorize the specific exemplars. In this case, they will attend to the specific combination of values of the stimuli seen during learning. In this case, we should not see a difference in the attention given to the diagnostic and nondiagnostic dimensions during classification. Instead, transfer performance should be based on similarity to the exemplars seen during learning.

Finally, inference is assumed to focus on relations among dimensions within a category, and so participants given inference learning should be equally sensitive to the four stimulus dimensions. For inference participants, however, they should be more sensitive to the prototype of the category than to the specific values of the particular exemplars seen during learning. This prediction would manifest itself as more of an advantage for consistent inferences (i.e., those that are consistent with the prototype) than for inconsistent inferences.

Table 4
The Category Structure Used in Experiment 2

	Set A						Set B				
	F	S	C	P	Ca		F	S	C	P	Ca
A1	1	1	1	1	1	B1	1	1	0	*1*	0
A2	1	1	*0*	*0*	1	B2	*0*	1	*1*	0	0
A3	*0*	*0*	1	1	1	B3	1	*0*	0	0	0
A0	1	1	1	1	1	B0	1	1	0	0	0

Notes. The dimension values that are not consistent with the prototype of the corresponding category (inconsistent questions) are shown in italics. A0 is the prototype of Set A, and B0 is the prototype of Set B. The exemplar A1 is the prototype of Set A, and the exemplar A2 is the prototype of Set B. It should be noted that Stimulus A2 is included in Set A, so that the two categories are not linearly separable. F = form; S = size; C = color; P = position; Ca = category label.

Method

Participants. Participants were 48 members of the Columbia University community who were paid for their participation.

Materials. The materials used for Experiment 2 were analogous to those used for Experiment 1 except that the categories in this experiment had a different nonlinearly separable structure (see Table 4). For every feature dimension, there were two exemplars that had a dimension value in common, and there was one exemplar that had a different dimension value from the rest of the members of the category. The prototype of Set A was *1 1 1 1*, which was also a member of the category (exemplar A1 in Table 4). The prototype of Set B was *1 1 0 0*, which was actually a member of Set A (exemplar A2 in Table 4). The six exemplars from Table 4 were used for classification learning and a subsequent classification test. Inference learning and a subsequent inference test consisted of inferences of all the stimulus dimensions of the six exemplars (in total 24 different questions).

Procedure. The basic procedure of this experiment was analogous to that described in Experiment 1. In Experiment 2, however, the transfer phase in Experiment 1 was replaced with a test phase, in which the stimuli that were shown in both inference and classification learning were presented. During the test phase, all participants answered 6 classification questions and 24 inference questions. There were no new transfer stimuli.

Results and Discussion

Consistent with our predictions as well as the results of Experiment 1, learning these categories was particularly difficult for participants given inference learning. All 24 participants in classification learning, but only 8 participants in inference learning reached the learning criterion. Looking only at participants who met the learning criterion, participants in classification learning required 10.4 blocks, and participants in inference learning required 22.1 blocks during the learning phase. The same pattern also holds if we include participants who failed to reach the learning criterion (inference: M = 27.4, classification: M = 10.4). Because the number of participants who reached the learning criterion differed considerably between the two learning procedures, the test data from each learning condition were analyzed separately.

Participants given classification learning exhibited accurate performance for all six stimuli. For the six test stimuli, the accuracy ranged from 88% to 96%. Participants were also accurate in the classification of Stimulus A1 (M = 0.88), which is the prototype of Set A (and a member of Set A), as well as Stimulus A2 (M = 0.92), which is the prototype of Set B, but is actually a member of Set A. During the test phase, participants' classification performance for these two stimuli were indistinguishable; $Z = -0.02$, $p > .10$ (see Table 5).

Participants in classification learning were also accurate in the inference test. Consistent with our prediction, classification learning revealed a tendency to direct participants to focus on the stimulus dimensions that were useful for distinguishing between categories. There was a tendency for the dimension inferences of color and position (M = 0.86) to be better than those for form and size (M = 0.80), although this difference was only marginally significant, $t(23) = 1.83$, $.05 < p < .10$. Consistent with this analysis, among 24 participants in classification learning, 12 participants showed higher performance for color and position than for form and size, 6 participants exhibited higher performance for form and size than for color and position, and the remaining 6 participants were equally accurate in their performance for these two sets of dimensions. In inference learning, the data from all participants were analyzed first, because only 8 of 24 participants met the learning criterion. First, the average performance for the classification test by participants in inference learning was M = 0.70. Unlike classification learning, there is a wide disparity between accuracy in classifying Stimulus A1 and the accuracy in classifying Stimulus A2. Participants in inference learning accurately classified the prototype stimulus of Set A—A1 (*1 1 1 1*), M = 0.83—but not the prototype stimulus of Set B—A2 (*1 1 0 0*), M = 0.46, $Z = 2.41$, $p < .01$. As mentioned earlier, Stimulus A2 (*1 1 0 0*) is the prototype of Set B, but it is placed in Set A. This result agrees with the view that inference learning promotes acquisition of a summary category representation.

Consistent with our prediction, participants in inference learning did not differ in the inferences of form and size, as compared with the inferences of color and position (form and size: M = 0.70; color and position: M = 0.70). This result, combined with the results from classification learning, suggests that inference and classification make use of different types of stimulus information.

The data taken exclusively from the eight participants who reached the learning criterion in inference learning show that these participants

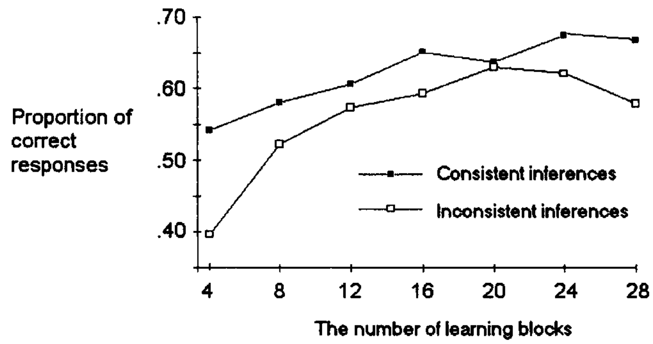

Figure 2. Performance on learning trials in inference learning. The proportions of correct responses were shown with respect to consistent inferences and inconsistent inferences over every four blocks of learning trials. Consistent inferences required the prediction of the value consistent with the prototype, and inconsistent inferences were those that required predicting the value opposite to that of the prototype.

adopted a strategy different from those who did not reach the criterion. They were accurate in their performance for the inference questions (M = 0.86) as well as for the classification questions (M = 0.81). In addition, they classified Stimuli A1 (M = 0.80) and A2 (M = 0.80) equally well, and they inferred form and size (M = 0.88) as accurately as color and position (M = 0.84), $t(7) = 0.57$, $p > .10$ (paired t test). These eight participants might have learned the categories with an exemplar-based strategy.

Because only 8 participants reached the learning criterion in inference learning, we also analyzed participants' learning performance for individual stimuli. As in Experiment 1, consistent inferences were defined as those that required the value possessed by the prototype, and inconsistent inferences were those that required predicting the value opposite to that of the prototype. If participants form a summary representation for inference, consistent inferences should be more accurate than inconsistent inferences. As predicted, participants' learning performance was significantly more accurate for consistent dimensions (M = 0.63) than for inconsistent dimensions (M = 0.56) throughout the learning phase, $t(22) = 3.46$, $p < .01$ (see Figure 2). Thus, this result suggests that participants' difficulty learning the nonlinearly separable categories stems from their inferences for dimensions that did not correspond to the category prototype.

In summary, the results of Experiment 2 support the hypothesis that nonlinearly separable categories are more difficult to learn through inference than through classification. These results are also consistent with the idea that inference and classification diverge in the category information that they promote. In particular, classification learning tends to accentuate feature dimensions that distinguish between categories.

Table 5
The Main Results of Experiment 2

	Classification test			
Learning	A1	A2	All exemplars	
Classification	0.88	0.92	0.94	
Inference	0.83	0.46	0.70	
	Inference test			
Learning	F	S	C	P
Classification	0.81	0.80	0.88	0.85
Inference	0.72	0.68	0.73	0.68

Note. These data are taken from all participants including those who reached the learning criterion and those who did not reach the criterion. In the classification learning condition, all 24 participants reached the criterion, and in the inference learning condition, 8 of 24 participants met the criterion. Stimulus A1 is the prototype of Set A and Stimulus A2 is the prototype of Set B, but this stimulus is included in Set A. F = form, S = size, C = color, P = position.

General Discussion

These studies demonstrate that it is easier to learn categories through classification than through inference when the categories are not linearly separable. This finding contrasts with earlier research using the same stimulus materials but with linearly separable categories, which found that inference learning was easier than classification learning (Yamauchi & Markman, 1995, 1998, 2000b). The distinction between inference and classification was also demonstrated in a previous study that did not require an incremental learning procedure (Yamauchi & Markman, 2000a). If we consider the results of the present studies in light of those done before, it is clear that the disparity between classification and inference did not stem from some superficial dissimilarity between the two tasks, such as the number of different dimensions queried (i.e., participants in classification learning answered only one type of question whereas participants in inference learning answered four different types of feature questions). Instead, the present results support the contention that inference focuses on summary category information regardless of the category structure, whereas classification opportunistically exploits information that allows accurate prediction of the category label. Although some researchers argue that inference and classification are different manifestations of the same task (e.g., Anderson, 1991), our results reveal that people use different strategies to carry them out.

Why do people look for summary information (e.g., prototypes) for inference, but seek information about specific exemplars or diagnostic dimensions for classification for the same category structure? This difference may follow from an intricate link between category representation and category function. Classification is related to object identification and recognition (Nosofsky, 1986). Thus, linking category labels with individual exemplars becomes salient. Once an object is identified, the particular dimension values of that exemplar are less relevant than its category membership. In contrast, inference involves the prediction of the value of a missing feature dimension. In this case, the category identity of the object is known, and so linking the category label to individual objects is not important. Instead, the relationship between the category label and individual features is pertinent to predicting the value of missing dimension values. As a consequence, inference must relate the category label to the features that are typical of category members.

Category Representation and Category Use

We have demonstrated that two interrelated tasks—inference and classification—lead to different category representations. Classification tends to focus on information that distinguishes between categories, whereas inference focuses on information common among exemplars within a category. This observation is part of a more general trend that is concerned with how category use influences category representation.

Some previous research has suggested that different kinds of category representations may emerge even within classification when learners have different beliefs about the type of category being learned. Goldstone (1996) found that categories considered interrelated were represented in a manner that accentuated contrasts between categories. Categories considered isolated did not selectively enhance contrasting information. In other work, Wattenmaker (1995) found that people in a sorting task were more likely to create family resemblance sorts when given items from a social domain (where the different dimension values can be interrelated) than when given artifacts (where participants tend to focus on one stimulus dimension). This finding suggests that people's beliefs about the domain also influence their category representations.

Most of the recent research has been concerned with how different ways of interacting with instances affects what is learned about them. For example, Brooks (1999) found that categories learned as an incidental part of carrying out another task were more likely to be judged as having an underlying essence than were categories learned through explicit classification. Thus, the importance of rules to categories formed in laboratory classification studies may reflect aspects of this task rather than something fundamental about the way categories are learned in general.

A. B. Markman and Makin (1998) explored the influence of communication on category acquisition. They found that having to fix common reference on a set of objects had two influences on representation. First, because similar labels were used for similar objects, but distinctions had to be drawn among objects with the same label, people tended to focus on the commonalities and differences of related items. Second, communication tended to synchronize category structures across individuals. Thus, communication may help to ensure that different people end up with the same category structures (see also Garrod & Doherty, 1994).

Ross (1996, 1997, 1999) has explored how the way one interacts with a set of instances affects subsequent performance with those instances. In one set of studies, people classified a set of diseases. Later, they learned about various drugs that could be prescribed for the diseases. Only a subset of the diagnostic features were actually relevant to the treatment decision. Those features that were relevant to the treatment task were ultimately treated as more important to the diagnoses of the disease than were the features that were equally diagnostic of the disease, but not relevant to the treatment decision. Love (2001) found that even when items only vary on a single stimulus dimension, how one interacts with the instances has a large effect on which category structures are easy or difficult to acquire. Finally, Smith and Minda (1998; Minda & Smith, 2001; Smith, Murray, & Minda, 1997) examined prototype-based and exemplar-based category learning and suggested that what is acquired in category learning depends on the factors that organize categories, such as the number of exemplars and of feature dimensions in a category, and the extent to which two categories overlap.

This brief survey of recent work reveals an emerging consensus that category use is a crucial determinant of category representation. Despite this trend, there are no general conclusions that can be drawn about how category use is related to category learning. We believe that extracting order from this chaos will require a recognition that there are a limited number of ways that people typically interact with instances. Classification, predictive inference, communication, and causal reasoning are among the most important functions of concepts. Understanding the interactions among these tasks as categories are acquired will go a long way toward helping us to understand the structure of people's natural categories. Modeling will also prove useful in uniting the different ways of learning about categories into a common theoretical rubric. Future research must seek a modeling framework that can be applied across learning tasks (see Love, Markman, & Yamauchi, 2000).

Conclusions

The studies presented here provide additional evidence for the hypothesis proposed by Yamauchi and Markman (1998) that categories learned by classification tend to focus on information that distinguishes between the categories whereas categories learned by making predictive inferences tend to focus on information shared across instances of the category. The present studies demonstrate that nonlinearly separable categories, which have little information that is common across members of a category, are much easier to learn through classification than through inference. Obviously the linearity of category structure is only one of many variables that affect inference and classification. Many other factors, such as the number of feature dimensions and the number of exemplars in a category, are likely to interact with the way people learn categories by inference or by classification (see Minda & Smith, 2001; Smith & Minda, 1998; Smith et al., 1997). Future research must address these issues and explore how categories that are learned through a combination of inference and classification compare with those learned by inference and classification alone.

References

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

Brooks, L. (1978). Non-analytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Erlbaum.

Brooks, L. R. (1999). Orientation to meaning suppresses explicit knowledge of surface cues in category learning. *Abstracts of the Psychonomic Society, 4*, 51.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition, 53*, 181–215.

Gelman, S., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science, 10*, 489–493.

Gelman, S., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183–209.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition, 24*, 608–628.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 411–422.

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

Lassaline, M. E. (1996). Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 754–770.

Love, B. C. (2001). *The multifaceted nature of unsupervised category learning*. Manuscript submitted for publication.

Love, B. C., Markman, A. B., & Yamauchi, T. (2000). Modeling inference and classification learning. In H. Kantz & B. Porter (Eds.), *Proceedings of the 17th National Conference on Artificial Intelligence* (pp. 136–141). Cambridge, MA: MIT Press.

Malt, B., Ross, B. H., & Murphy, G. L. (1995). Predicting features for members of natural categories when categorization is uncertain. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 646–661.

Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General, 127*, 331–354.

Markman, A. B., Yamauchi, T., & Makin, V. S. (1997). The creation of new concepts: A multifaceted approach to category learning. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Conceptual structures and processes: Emergence, discovery, and change* (pp. 179–208). Washington, DC: American Psychological Association.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology, 16*, 1–27.

Medin, D., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York: Cambridge University Press.

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

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 775–799.

Murphy, G. L., & Ross, B. H. (1994). Predictions from uncertain categorizations. *Cognitive Psychology, 27*, 148–193.

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.

Osherson, D. N., Smith, E. D., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category based induction. *Psychological Review, 97*, 185–200.

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

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14*, 665–681.

Ross, B. H. (1996). Classification and the effects of interacting with instances. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1249–1265.

Ross, B. H. (1997). The use of categories affects classification. *Journal of Memory and Language, 37*, 240–267.

Ross, B. H. (1999). Postclassification category use: The effects of learning to use categories after learning to classify. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 743–757.

Ross, B. H., & Murphy, G. L. (1996). Category-based predictions: Influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 736–753.

Shepard, R. N., Hovland, D. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75*(13, Whole No. 517), 1–42.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1411–1436.

Smith, J. D., Murray, M. J., & Minda, J. P. (1997). Straight talk about linear separability, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 659–680.

Wattenmaker, W. D. (1995). Knowledge structures and linear separability: Integrating information in object and social categorization. *Cognitive Psychology, 28*, 274–328.

Yamauchi, T., & Markman, A. B. (1995). Effects of category learning on categorization—An analysis of inference-based and classification-based learning. In J. D. Moore & J. F. Lehman (Eds.), *The Proceedings of the 17th Annual Meeting of the Cognitive Science Society* (pp. 786–790). Pittsburgh, PA: Erlbaum.

Yamauchi, T., & Markman, A. B. (1998). Category-learning by inference and classification. *Journal of Memory and Language, 39*, 124–148.

Yamauchi, T., & Markman, A. B. (2000a). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 776–795.

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

Yamauchi, T., & Markman, A. B. (2000c). Making inferences and classifications using categories that are not linearly separable. In J. D. Gleitman & A. K. Joshi (Eds.), *The Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 565–570). Hillsdale, NJ: Erlbaum.

Received February 23, 2001
Revision received October 30, 2001
Accepted October 30, 2001 ■