JOURNAL OF MEMORY AND LANGUAGE **39,** 124–148 (1998)
ARTICLE NO. ML982566

Category Learning by Inference and Classification

Takashi Yamauchi and Arthur B. Markman

Columbia University

The nature of category formation is linked to the tasks applied to learn the categories. To explore this idea, we investigated how three different methods of category learning—Classification Learning, Inference Learning, and Mixed Learning (a mixture of the two)—affect the way people form categories. In Classification Learning, subjects learned categories by predicting the class to which an individually presented exemplar belonged given feature information about the exemplar. In Inference Learning, subjects learned categories by predicting a feature value of a stimulus given the class to which it belonged and information about its other features. In Mixed Learning, subjects received the Classification task on some trials and the Inference task on other trials. The results of two experiments and model fitting indicate that inference and classification, though closely related, require different strategies to be carried out, and that when categories are learned by inference or by classification, subjects acquire categories in a way that accommodates these strategies.

Categories serve a variety of purposes including classification, inference, communication, visual perception, and complex reasoning (Biederman, 1987; Gelman, 1986, 1988; Gentner, 1989; Glucksberg & Keysar, 1990; Harnad, 1987; Heit & Rubinstein, 1994; Holyoak & Thagard, 1995; Lassaline, 1996; Osherson, Smith, Wilkie, Lopes, & Shafir, 1990; Rips, 1975; Smith & Medin, 1981). How do we acquire categories rich enough to subserve these functions? Research on categorization has been primarily concerned with the study of classification and has often neglected to address this question. Central to this approach is the assumption that classification learning is a chief vehicle for forming categories. Categories, however, are used in widely different circumstances and incorporate a variety of information. Thus, it may be more appropriate to assume that categories are formed in relation to specific tasks at hand. From this perspective, the nature of category formation can be examined with respect to the tasks involved in learning (Markman, Yamauchi, & Makin, 1997; Ross, 1996; Whittlesea, Brooks, & Westcott, 1994).

The purpose of this article is to examine the link between the function of categories and the formation of categories. We will address this problem by contrasting two of the fundamental functions of categories—inference and classification—in the context of category learning (Smith, 1994). Inference and classification play a critical role in the formation of natural categories. For example, the family resemblance structure of basic level categories is said to emerge in the process of balancing specificity and generality associated with feature prediction (i.e., inference) and object classification (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Additionally, inference and classification are functionally related and can be treated as identical if category labels and category features are compatible (Anderson, 1990, 1991). In this regard, elucidating the mechanisms involved in the two tasks and their impact on category formation would help us to understand the relationship

This research was supported by Career Award SBR-95-10924 to Arthur Markman from the National Science Foundation. We thank John R. Anderson, Dendre Gentner, Dave Krantz, Trisha Linderman, Douglas Medin, Tomislav Pavlicic, Brian Ross, Yung-Cheng Shen, and Shi Zhang for their helpful suggestions.

Address correspondence and reprint requests to Arthur B. Markman, University of Texas, Department of Psychology, Austin, TX 78712. E-mail: markman@psy.utexas.edu.

0749-596X/98 $25.00

between category learning and category formation and would provide insight into grasping the nature of category formation in an experimental setting. In the following studies, we will examine (1) how inference and classification are carried out by using categories and (2) how the different mechanisms associated with the two tasks alter the way people form categories when categories are learned by inference or by classification.

In this paper, we first review several empirical studies that highlight the distinction between inference and classification. Next, we describe three learning procedures—Inference Learning, Classification Learning, and Mixed Learning and lay out how inference and classification differ in the context of category learning and how these differences affect the way people form categories. Then, we present two studies that investigate the impact of the three learning procedures on category formation. Finally, we fit two mathematical models of classification—Medin and Schaffer's (1978) context model and Anderson's (1990, 1991) rational model—to examine further the distinction between the two tasks.

Throughout this paper, we use the term *category label* to refer to a symbol that denotes a particular group of stimuli and the term *category feature* to mean a symbol that denotes a characteristic of a stimulus. *Classification,* which involves the prediction of the category label of a stimulus, is characterized in our experiments as a practice in which a stimulus is placed into one of two groups when the attributes of the stimulus are known. *Inference,* which involves the prediction of the value of a category feature, is characterized in our experiments as a practice in which an attribute of a stimulus (i.e., a category feature) is predicted when the group to which the stimulus belongs (i.e., the category label) and other attributes of the stimulus are known (for similar descriptions of inference, see Estes, 1994; Murphy & Ross, 1994; Yamauchi & Markman, 1995). For example, classification as we defined it is akin to the situation in which people predict a category to which a person belongs (e.g., Democrat) by observing his attributes (e.g., supports affirmative action and favors reducing defense spending). In contrast, inference as we defined it is akin to the situation in which people predict an attribute of a person (e.g., supports affirmative action) based on a category label to which the person belongs and his other attributes (e.g., is a Democrat and favors reducing defense spending). Finally, we define *category representation* as the mental structure that specifies the information that was acquired through interaction with the members of categories and assume that the specified information is obtained in the process of making classifications and inferences using categories.

INFERENCE AND CLASSIFICATION

Despite the close relationship between inference and classification, several empirical findings reveal that people adopt different strategies to carry out the two tasks. In inference, subjects tend to pay particular attention to relationships between exemplars within a category (e.g., family resemblance among exemplars within a category or typicality information about exemplars in a category) (Lassaline & Murphy, 1996; Rips, 1975; Rosch et al., 1976), while in classification subjects focus on feature information useful for dividing exemplars into groups (Ahn & Medin, 1992; Medin, Wattenmaker, & Hampson, 1987). In one study, for example, Lassaline and Murphy (1996) asked subjects to predict feature values of category exemplars given other feature values of the exemplars. Following this inference task, subjects sorted a set of exemplars into categories. Subjects in this task were much more likely to sort the stimuli on the basis of family resemblance than were subjects who sorted the stimuli after making other judgments (who generally sorted the exemplars based on the values of a single feature dimension). As further support, Rips (1975) found that the likelihood that people predict that subordinate category members have a particular feature value is correlated with the typicality of that category member, suggesting that people make inferences based on family resemblance between exemplars (see also Malt, Ross, & Murphy, 1995; Murphy & Ross, 1994; Ross & Murphy, 1996; for the argument that people focus on a single target category to make inferences).

In contrast to inference, people tend to focus on a small number of diagnostic features in classification (Medin et al., 1987; Nosofsky, Clark, & Shin, 1994; Nosofsky, Palmeri, & Mckinley, 1994; Tversky, 1977). Sorting tasks, which are quintessential classification tasks, provide evidence that subjects generally attend to a limited number of diagnostic features that distinguish between categories when they classify stimuli, as subjects in these tasks tend to group stimuli with a single salient feature even in the presence of a clear family-resemblance structure (Ahn & Medin, 1992; Medin et al., 1987). Nosofsky and his colleagues (1994) also demonstrate that a computational model based on simple rules and exceptions can account for people's performance on a wide variety of classification tasks. Other research suggests that different types of diagnostic features become salient in classification depending on the way that the stimuli are grouped (Tversky, 1977). Although people may carry out a classification task in a number of different ways, it seems reasonable to assume that focusing on diagnostic features is one of many strategies that people adopt in classification. In the following studies, we will investigate why people use different strategies to make inferences or classifications and how these differences affect the way people form categories.

OVERVIEW OF EXPERIMENTS

We developed an inference-based learning task (i.e., the Inference Learning task; see Fig. 1 and Estes, 1994; Yamauchi & Markman, 1995 for descriptions of similar inference tasks) and compared it with a standard classification-based learning task (i.e., the Classification Learning task) in order to investigate the distinction between inference and classification and their impact on category formation. In the standard classification-based learning task, subjects acquire categories incrementally by predicting the category label of individually presented stimuli and receiving feedback after each response (Posner & Keele, 1968, 1970; Medin & Schaffer, 1978; Malt, 1989; Nosofsky, 1986; Shepard, Hovland, & Jenkins, 1961). Similarly, in the inference-based learning task (i.e., Inference Learning), subjects acquire categories incrementally by predicting feature values of individually presented stimuli and receiving feedback after each response. For example, in the Classification task (see Fig. 1a), subjects are presented with a stimulus depicting the values of the form, size, color, and position of the geometric figure and they predict the category label of that stimulus. In the Inference task (see Fig. 1b), subjects are presented with the values of the size, shape, and position of the geometric figure along with the category label to which the stimulus belongs (e.g., Set A), and they predict the value of a missing feature (e.g., the color). On different trials, subjects in Inference Learning predict the values of different features. In addition to these two conditions, we also included a ''Mixed Learning'' condition in Experiment 1, in which subjects classified stimuli on some trials, and made feature inferences on others.

Initially, no information about the categories was given to subjects in our studies, so that they had to learn the two categories by trial and error. The learning phase continued until subjects reached a criterion of 90% accuracy in three consecutive blocks (24 trials) or until they completed 30 blocks (240 trials).[1] Following the learning phase, the nature of the category representation is probed on transfer trials, which consisted of classifications and inferences of old stimuli that appeared during learning and new stimuli that did not appear during learning. In the transfer phase, all subjects received the same trials.

In our experiments, the stimuli were divided into two classes such that every exemplar shared three feature values with its corresponding prototype (A0 or B0) and one feature

[1] This 90% accuracy criterion was introduced to keep the experiment to a reasonable length (about 30 to 40 min).

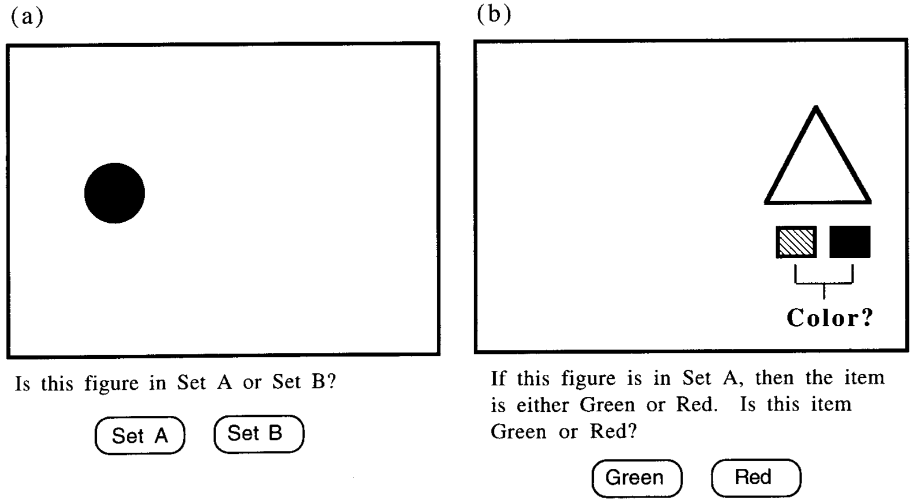

FIG. 1. (a) A stimulus frame for a classification trial; in a particular classification trial, a subject is given a figure whose form, size, color, and position are specified. Then, the subject is asked to predict the category label (Set A or Set B) of the stimulus. (b) A stimulus frame for an inference trial; in a particular inference trial, a subject is given a figure whose form, size, position, and the category label are specified. Then, the subject is asked to predict the color of the item.

value with the prototype of the other category (Table 1). We used simple stimuli consisting of geometric figures varying in their size, form, position, and color in order to focus on the effect of the learning procedures (see Medin & Schaffer, 1978).

In Classification Learning, subjects classified the eight exemplars but not the prototype stimuli. In Inference Learning, subjects inferred all the feature values of stimuli except for the ''Exception-features.'' The Exception-features, shown in bold italics in Table 1, are the feature values of a category that are consistent with the prototype of the other category. For example, the values of all the features in Set A are 1 except the values of the exception features which are 0. We did not include Exception-feature inferences in the learning phase and presented them only in the transfer phase for two reasons. First, we excluded them to keep the Classification Learning condition and the Inference Learning condition as equivalent as possible. On each classification question, subjects predicted the value of the category label given the values of all the four feature dimensions (e.g., the stimulus A1 in Table 1). This question has a schematic structure (1, 1, 1, 0, ?) = (form, size, color, position, category-label) in the exemplar A1, assuming that the category label is just another feature (see Anderson, 1990). Analogously, on each inference question (e.g., a question about the form of the stimulus A1), subjects predicted the value of a missing feature (e.g., the value of form) while the values of the other three features and the category label were shown (e.g., the values of size, color, position, and the category label). This question has a schematic structure (?, 1, 1, 0, 1) = (form, size, color, position, category-label) and is formally equivalent to the classification question (e.g., (1, 1, 1, 0, ?)), provided that the prediction of category labels and the prediction of category features are in principle compatible. The Exception-feature trials have a different structure. For example, the position inference for the stimulus A1 yields a schematic structure (1, 1, 1, ?, 1), and is analogous to the classification of a prototype—(1, 1, 1, 1, ?). Because prototype stimuli were not presented in Classification Learning, it is necessary to exclude Exception-feature inferences from Inference Learning to keep the two learning conditions equivalent.

TABLE 1

Stimulus Structure Used in Experiment 1

Category A	Form	Size	Color	Position	Category B	Form	Size	Color	Position
A1	1	1	1	***0***	B1	0	0	0	***1***
A2	1	1	***0***	1	B2	0	0	***1***	0
A3	1	***0***	1	1	B3	0	***1***	0	0
A4	***0***	1	1	1	B4	***1***	0	0	0
A0 (prototype)	1	1	1	1	B0 (prototype)	0	0	0	0

Note. Exception-features are shown in bold italics.

Second, we excluded Exception-feature questions from Inference Learning in order to examine the nature of feature information used for making inferences in the transfer phase. On an Exception-feature inference trial, there are two possible choices, one that is consistent with the value of the prototype of the category, and the other that is consistent with exception values of each category. For example, on the Exception-feature question for the stimulus A1 (Table 1), subjects see a stimulus with form, size, color and the category label with a value of 1, and they infer the value of position (e.g., (1, 1, 1, ?, 1)). If they respond with the value of 0 (e.g., right), then they are making a response consistent with the stimulus A1 (e.g., (1, 1, 1, 0, 1)), which is presented in feedback during learning. If they respond with the value of 1 (e.g., left), then they are making a response consistent with the prototype (i.e., A0(1, 1, 1, 1, 1)) of the category, which is not given in feedback during learning. Thus, the choice of feature values on these trials may provide some insight to assess the degree to which subjects use either family resemblance information or exception-feature information for inference.

HYPOTHESES AND PREDICTIONS

In the previous section, we described several empirical studies that are consistent with the idea that inference and classification involve different mechanisms; in inference subjects assess relationships between exemplars within a category while in classification they focus on features that distinguish between categories. In this section, we would like to discuss how these two strategies can be translated into our experimental setting and how they would influence the way people form categories when categories are learned by inference or by classification.

As a general rule, we assume that inferences guide subjects to focus on the target category (see Malt, et al., 1995; Murphy & Ross, 1994; Ross & Murphy, 1996), while classification often leads subjects to focus on a small number of diagnostic features that are useful to divide exemplars into groups. This distinction might have arisen because the two tasks are associated with two different purposes of categories. Inference often requires the identification of an unknown property or the internal structure that is not readily apparent (Gelman, 1986). Thus, focusing on the commonalities among exemplars within a category might be advantageous for inference. In contrast, classification is related to the operation of object recognition and identification (see Nosofsky, 1986). For this purpose, finding a salient feature that differentiates between exemplars is useful. This focus is evident in sorting tasks, in which subjects consistently use a single salient feature to sort stimuli if no intervening tasks are given prior to sorting (see Lassaline & Murphy, 1996; Markman & Makin, in press). Although it is not clear exactly how sorting tasks speak to the classification task, it seems plausible to assume that this focus on a single salient feature will occur in classification tasks as well.

In sum, we assumed that categories learned for inference or for classification will embody the characteristics that accommodate these two strategies. We argue that inference promotes the acquisition of category representations characterized with the prototypes of the categories, while classification facilitates the formation of categories consistent with rules and exceptions or concrete exemplars. The following normative models illustrate these processes.

In Classification Learning, subjects would fulfill the classification task by assessing at least one of two conditional probabilities— the probability that the response ''Set A'' is likely given the feature information about Stimulus_i (i.e., *P(SetA|Stimulus_i)*) and the probability that the response ''Set B'' is likely given the feature information about Stimulus_i (i.e., *P(SetB|Stimulus_i)*). For example, subjects may choose the response ''Set A'' if *P(SetA|Stimulus_i)* is larger than *P(SetB|Stimulus_i)*, or vice versa. It is also possible that subjects make a classification judgment by setting a decision criterion. For example, they may choose ''Set A'' if *P(SetA|Stimulus_i)* is larger than, say, 0.5, and choose ''Set B'' if *P(SetA|Stimulus_i)* is not larger than 0.5. In either case, at least one of the following two conditional probabilities would be assessed in the classification question of the stimulus A1

$$P(C_1|F_{f1}, F_{s1}, F_{c1}, F_{p0}) = \frac{P(C_1, F_{f1}, F_{s1}, F_{c1}, F_{p0})}{P(F_{f1}, F_{s1}, F_{c1}, F_{p0})} \quad [1a]$$

$$P(C_2|F_{f1}, F_{s1}, F_{c1}, F_{p0}) = \frac{P(C_2, F_{f1}, F_{s1}, F_{c1}, F_{p0})}{P(F_{f1}, F_{s1}, F_{c1}, F_{p0})}, \quad [1b]$$

where C_1 and C_2 stand for the category labels with the values 1 and 2, respectively, and F_{f1}, F_{s1}, F_{c1}, and F_{p0} stand for the feature values of form, size, color, and position—(1, 1, 1, 0), respectively.

Analogously, subjects in Inference Learning may carry out the inference task by assessing at least one of two conditional probabilities—the probability that the feature value 1 is likely given the values of other features and its category label [2a], and the probability that the feature value 0 is likely given the values of other features and its category label [2b]. As in classification, they may also choose the feature value ''1'' if one of the two probabilities exceeds a particular threshold. In either case, subjects would assess at least one of the two conditional probabilities in an inference question of the stimulus A1[2]

$$P(F_{f1}|C_1, F_{s1}, F_{c1}, F_{p0}) = \frac{P(C_1, F_{f1}, F_{s1}, F_{c1}, F_{p0})}{P(C_1, F_{s1}, F_{c1}, F_{p0})} \quad [2a]$$

$$P(F_{f0}|C_1, F_{s1}, F_{c1}, F_{p0}) = \frac{P(C_1, F_{f0}, F_{s1}, F_{c1}, F_{p0})}{P(C_1, F_{s1}, F_{c1}, F_{p0})} \quad [2b]$$

As the four equations show, the four conditional probabilities that may be assessed in inference and classification are in principle identical if the category labels (C_1 and C_0) and the feature form (F_{f1} and F_{f0}) are identical.

The assumption that subjects at the beginning of the learning phase attend to a diagnostic feature in classification suggests that subjects obtain a classification judgment primarily based on the value of the target feature, while ignoring information about other features that are not attended. To translate this process, the feature values F_{s1}, F_{c1}, F_{p0} can be removed from Eqs. [1a] and [1b], resulting in Eqs. [3a] and [3b], if, for example, subjects focus on the feature form

$$P(C_1|F_{f1}) = \frac{P(C_1, F_{f1})}{P(F_{f1})} \quad [3a]$$

$$P(C_0|F_{f1}) = \frac{P(C_0, F_{f1})}{P(F_{f1})} \quad [3b]$$

[2] As the two equations [1] and [2] show, the classification task is related to *cue validity* (i.e., how likely the category given a feature) and the inference task is related to *category validity* (i.e., how likely a feature given a category). If the category label is equivalent to other features, then the two equations are identical.

Similarly, the assumption that subjects focus on the target category in inference can be translated into Eqs. [4a] and [4b] provided that the focus on the target category is made by the focus on the category label (see Murphy & Ross, 1994, for the argument of a focus on a single target category in inference, and see Yamauchi & Markman, in preparation; for an argument for a focus on the category label in inference)

$$P(F_{f1} | C_1) = \frac{P(C_1, F_{f1})}{P(C_1)} \qquad [4a]$$

$$P(F_{f0} | C_1) = \frac{P(C_1, F_{f0})}{P(C_1)}. \qquad [4b]$$

As in classification, the feature information F_{s1}, F_{c1}, F_{p0} is unattended or not used for the inference judgment.

Our argument is that the difference in focus between inference and classification will ultimately lead to the acquisition of distinct category representations, even if the stimuli presented in each learning procedure convey roughly the same amount of information about the relationship of the features to the categories (see the previous section). In this category structure, none of the features are perfectly correlated with the category division. Therefore, the focus on any single feature is not sufficient to predict the category division more than 75% of the time. Thus, subjects in Classification Learning need either to store some specific cases, such as the case in which the feature value F_{f0} is linked to the category label C_1, or to employ a disjunction rule (e.g., subjects make the response ''Set A'' if at least two of three features have the values 1, otherwise they make the response ''Set B''). In either case, this would induce subjects to attend to concrete exemplars or exception-features in Classification Learning along with a limited number of diagnostic features. Consequently, Classification Learning facilitates the acquisition of category representations characterized by the information about salient features along with some exceptions or with a number of concrete exemplars (Medin & Schaffer, 1978; Nosofsky, 1986; Nosofsky et al., 1989).

In contrast to classification, the focus on the target category promotes the acquisition of categories consistent with family resemblance or prototypes of the two categories. In this setting, the two category labels represent the two groups unambiguously, so that subjects can focus on the category labels and link them directly to feedback C_1 and F_{f1} in this example). No Exception-feature questions are presented to subjects in Inference Learning, so that subjects may associate the category label with the prototype (e.g., A0(1, 1, 1, 1)). As a result, Inference Learning should facilitate the acquisition of category representations consistent with prototypes or family resemblance between members of each category.

In natural settings, if inferences are made to a large number of exemplars and to a variety of feature dimensions, feature values that people associate with the category label should be close to the average feature values of all the exemplars within the category as the number of inferences increases (see Hintzman, 1986). If the exemplars of a category are clustered by a family resemblance structure, the average feature values that subjects link with the category label can be approximately the prototype of that category.[3] As a consequence, making inferences would promote the formation of categories consistent with the prototypical values of the categories. In contrast, the focus on a diagnostic feature, which may be adopted in classification, would impede the extraction of prototypes even when classification is made to a large number of exemplars. If the focus on a diagnostic feature is effective for prediction, then there is no need to attend to other features. If the focus is not very effective for prediction, then one can look for another diagnostic feature or employ some decision rules (e.g., conjunction or disjunction rules). Thus, classification would obscure information about other features that

[3] Because we used the binary values (1 and 0) to distinguish the two feature values, we employed modes to represent the prototypes of the two categories.

are not focused, which would in turn deter the extraction of the prototype of a category.

This reasoning leads to three basic predictions in our experiments. First, the idea that inference is in general linked to the assessment of family resemblance within a category will be examined by observing subjects' response-patterns in Exception-feature inferences. In particular, we predict that subjects in the three learning conditions should respond with prototype-feature values more often than with Exception-feature values. In the category structure employed in our experiments (Table 1), the two prototypes—A0(1, 1, 1, 1) and B0(0, 0, 0, 0)—recapitulate the family resemblance structure of the two categories. If the judgment involved in inference requires assessing the feature values that exemplars of a category have in common (e.g., family resemblance among category exemplars), then subjects in all three learning conditions will exhibit a tendency to select the feature values that are shared by many of exemplars in a category (e.g., prototype-feature values) rather than the features values that are shared by exemplars of the other category (e.g., Exception-feature values).

Second, this tendency may be reduced for subjects in Classification Learning as compared to subjects in Inference Learning. If classification induces attention to a small number of diagnostic features and some specific exceptions, then subjects in Classification Learning may be less likely to respond with prototype-feature values than may subjects in Inference Learning given Exception-feature questions.

Third, the hypothesis that the two learning procedures produce different category representations can be tested by examining subjects' overall performance for the transfer tasks. Given the assumption that Classification Learning promotes the acquisition of category representations consistent with rules and exceptions or concrete exemplars, subjects in Classification Learning should be the best on classification transfer tasks, assuming that categories formed in Classification Learning facilitate classification judgment. Similarly, given the assumption that Inference Learning promotes the formation of categories congruent with family resemblance information, subjects in Inference Learning should be the best on the inference transfer task. Analogously, subjects in Mixed Learning should be better both in classification transfer and in inference transfer to the extent that they receive Inference Learning trials and Classification Learning trials. This pattern of data should emerge regardless of subjects' familiarity with Inference or Classification Learning tasks.

In the two experiments reported below, we will test these predictions. In Experiment 1, subjects learn the two categories with one of three learning procedures—Inference Learning, Classification Learning, or Mixed Learning, and we examine the distinction between inference and classification by focusing on subjects' performance on transfer questions. In Experiment 2, we will test directly if the different strategies employed in the two learning tasks can specify the formation of categories.

EXPERIMENT 1

Method

Participants. Participants were 77 undergraduates at Columbia University who participated in the experiment for course credit. The data from 4 subjects were removed from the analyses because these subjects failed to follow the instructions, and the data from 1 subject were lost due to a coding error. In all, the data from 72 subjects (24 per condition) were analyzed.

Materials. Stimuli used for this experiment were like those used in the first experiment of Medin and Schaffer's (1978) studies. They were geometric figures having four feature dimensions—form (circle, triangle), color (red, green), size (large, small), and position (left, right) (Fig. 1). Each stimulus was bounded by a 20.3 $\times$ 17.4 cm rectangular frame drawn with a solid black line on the computer screen.

The structure of the two categories is illustrated in Table 1 (see Medin et al., 1987). A single stimulus set was drawn containing an arbitrary assignment of dimension values of 0 and 1 in the stimulus design. For form, the value of 0 was triangle, and the value of 1 was circle. For size, the value of 0 was small, and the value of 1 was large. For color, the value of 0 was green, and the value of 1 was red. For position, the value of 0 was right, and the value of 1 was left. All subjects saw the same stimulus set. The eight stimuli (A1–A4, B1–B4) were divided into two categories. The category exemplars share three features with the prototype of that category and one feature with the prototype of the other category. Thus, no single feature can unambiguously determine the category division.

Procedure. The basic procedure of the experiment involved three phases—a learning phase, a filler phase, and a transfer phase. In the learning phase, subjects were randomly assigned to one of three experimental conditions—Classification, Inference, and Mixed. For all the three conditions, subjects continued in the learning phase until they performed three consecutive blocks with a combined accuracy of 90% or until they completed 30 blocks (240 trials). A classification block consisted of presentations of eight exemplars. One inference block included inferences of all four feature dimensions. One block of the Mixed condition was either a classification block or an inference block. In the three conditions, every exemplar appeared once in the feedback of each block. The order of stimulus presentation was determined randomly.

In the Classification Learning condition, subjects were shown one of the eight stimuli and were asked to indicate the category to which it belonged by clicking a button with the mouse (Fig. 1a). In the Inference Learning condition, subjects made inferences of one of four features while its category label and the remaining three feature values were depicted in the stimulus frame. For instance, in Fig. 1b, subjects were given a stimulus frame containing the form, size, and position of the item as well as its category label, and the color of the item was left unspecified. They were then asked to select one of the two values of the unspecified feature—color. On other trials, they predicted other dimensions (form, size, or position). Subjects responded by clicking one of two labeled buttons with the mouse. For each stimulus, the location of the correct choice was randomly determined. The Mixed condition was a mixture of classification and inference blocks. Half of the blocks in this condition were classification and half were inference. The order of the blocks was determined randomly for each subject.[4]

Initially, no information about the category division was given to subjects, and so subjects had to guess. Following each response, feedback was provided in a stimulus frame that depicted the correct response; the stimulus and the feedback remained on the screen for 3 s after their response. The stimulus frames that depicted correct responses were identical in both classification and inference tasks. In Classification Learning, subjects saw all eight exemplars but not the two prototypes (i.e., A0(1, 1, 1, 1) and B0(0, 0, 0, 0)). In Inference Learning, subjects answered all the feature questions for each stimulus except for the Exception-feature questions.

Following the learning trials, all subjects in the three learning conditions participated in the same transfer tasks, which followed a 10-min filler task, where subjects judged the pronounceability of nonsense words. In the transfer phase, subjects were first given classification transfer tasks followed by inference transfer tasks. In this phase, the instructions specifically asked subjects to make their decisions based on the categories learned during the learning phase when the values of the four features were given. In the classification transfer task, as in the classification learning task, subjects were asked to indicate the category label of a stimulus based on the categories they learned. In the inference transfer task, as in the inference learning task, subjects were asked to indicate the value of the missing fea-

[4] Three consecutive blocks used to assess the learning criterion could differ one subject from another and could be any combination of classification and inference blocks in the Mixed condition because the order of the presentation of the blocks was determined randomly.

ture of the stimulus based on the categories they learned when the category label to which the stimulus belonged and the values of other features of the stimulus were shown. No feedback was given during transfer. First, subjects classified the eight exemplars that appeared in the learning phase as well as two new prototype stimuli (A0 and B0) that were not presented during the learning phase. Immediately after each classification, subjects indicated whether they had seen the stimulus during the learning trials.[5] The order of stimulus presentation for the ten stimuli was determined randomly. Following the classification task, subjects proceeded to the inference transfer task. They performed all possible feature inferences including Exception-feature inferences (32 inferences in total). The order of stimulus presentation for inference transfer was determined randomly. The entire experiment took 30 to 40 min.

Design. There were three between-subjects learning conditions: Inference, Classification, and Mixed. Five dependent measures served for our analyses. First, we examined the number of subjects who reached the 90% accuracy criterion, and the number of blocks needed to reach the criterion in the learning phase. The rest of the measures encompassed the transfer tasks: the proportion of correct classifications of old exemplars, the proportion of correct classifications of the prototypes, the proportion of correct inferences to old exemplars, and the proportion of inferences to Exception-features consistent with the prototype features of the category.

Results and Discussion

All dependent measures were analyzed with one-way ANOVAs. For these analyses, we used the data from only those subjects who reached the 90% accuracy criterion before the 30 block maximum.[6] In all, 22 subjects reached the criterion in the Inference Learning condition, 23 in the Classification Learning condition, and 20 in the Mixed Learning condition. First, we measured the number of blocks that were required to reach the criterion to examine relative difficulty of the three learning conditions. In this measure, the three learning conditions were significantly different: $F(2, 62) = 10.62$, $MSE = 32.1$, $p < 0.001$.[7] In particular, subjects in Inference Learning ($m = 6.5$) required fewer blocks to reach the criterion than did subjects in Classification Learning ($m = 12.3$), or in Mixed Learning ($m = 14.2$); for both comparisons, $t > 4.0$, $p < 0.001$(Bonferroni). The difference between subjects in Classification Learning and subjects in Mixed Learning was not statistically significant, $t(41) = 0.96$, $p > 0.10$.

In the transfer phase, the proportions of correct responses exceeded a chance level in every dependent measure of the three learning conditions; $t > 2.5$, $p < 0.05$, implying that the Classification Learning task and the Inference Learning task were capable of producing category representations flexible enough to be used with the transfer task that was not given during learning. As predicted, performance of subjects in each condition was generally better when the learning task matched the transfer task. The three learning procedures differed in the classification transfer of old stimuli; Classification, $m = 0.92$; Mixed, $m = 0.88$; Inference, $m = 0.77$; $F(2,62) = 8.42$, $MSE = 0.02$, $p < 0.001$ (Fig. 2a). Subjects in Classification Learning were more accurate than subjects in Inference Learning; $t(43) = 4.22$, $p < .001$. Subjects in Mixed Learning also performed better than did subjects in Inference Learning in the classification transfer of old stimuli, although this difference was only mar-

[5] We collected the recognition performance data of the subjects on an exploratory basis. Because this experiment was not designed to survey recognition performance (i.e., there were only 2 new stimuli out of 10 stimuli), we will not discuss this task further.

[6] Since analyses of the data from all the subjects (including subjects who did not reach the 90%-above accuracy criterion) showed basically the same patterns as observed in the subjects who reached the criterion, we report only the data obtained from the subjects who reached the criterion in the following two experiments.

[7] The number of blocks shown in these results includes three consecutive blocks used to assess the criterion.

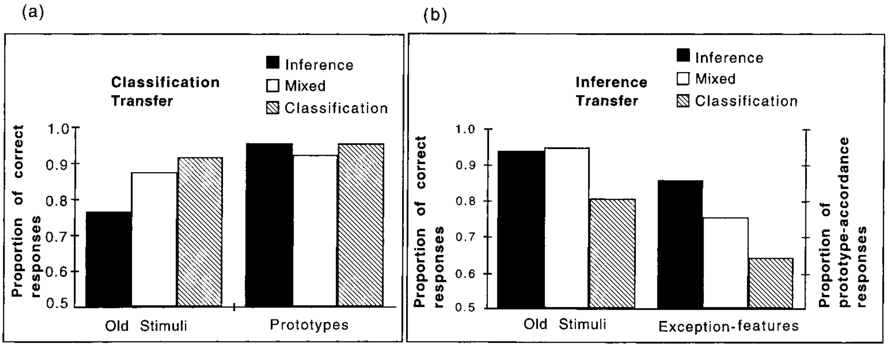

FIG. 2. (a) The classification transfer performance for old stimuli and prototype stimuli of Experiment 1. (b) The inference transfer performance for old stimuli and Exception-feature stimuli of Experiment 1. For the Exception-feature inferences, the proportion that subjects responded with the prototype stimuli was reported (i.e., prototype-accordance responses).

ginally significant; $t(43) = 2.49$, $p < 0.06$. The performance for classification of prototypes did not differ reliably between conditions; Classification, $m = 0.96$; Inference, $m = 0.96$; Mixed, $m = 0.93$; $F(2,62) < 1$, $MSE = 0.04$. In inference transfer of old stimuli, subjects in the three learning conditions differed in their performance (Fig. 2b); Inference, $m = 0.94$; Mixed, $m = 0.95$; Classification, $m = 0.81$; $F(2,62) = 11.3$, $MSE = 0.01$, $p < 0.001$. Subjects in Inference Learning and subjects in Mixed Learning made higher proportions of correct responses than did subjects in Classification Learning; respectively, $t(43) = 3.45$, $p < 0.01$; $t(43) = 3.92$, $p < .001$. These results imply that the category representations obtained by each learning procedure are specific to each learning condition to some extent while maintaining some level of generality.

The response-patterns observed on Exception-feature transfer trials were also consistent with the view that inference promotes the assessment of family resemblance while classification increases attention to a few diagnostic features and exceptions. First, subjects in all the three conditions typically responded with feature values that were in accord with the category prototypes (i.e., prototype-accordance responses); Inference, $m = 0.86$; Mixed, $m = 0.76$; Classification, $m = 0.64$; for all conditions, $t > 2.5$, $p < 0.03$, confirming the view that assessing family resemblance information is critical for inference. The three conditions differed significantly in their preference for prototype-accordance responses, $F(2, 62) = 3.34$, $MSE = 0.08$, $p < 0.01$. As predicted, the tendency to respond consistent with prototype-feature values (i.e., prototype-accordance responses) was reduced significantly in subjects in Classification Learning as compared to subjects in Inference Learning, $t(43) = 2.73$, $p < 0.05$, implying that classification tends to promote a focus on exceptions or concrete exemplars to a larger degree than does inference.

To summarize, three aspects of the results of Experiment 1 are in accord with the hypothesis that subjects employ different strategies to make inferences or classifications. First, on Exception-feature questions, subjects in all the three conditions responded with prototype-feature values more often than with exception-feature values. The results imply that checking family resemblance information is critical in inference. Second, this tendency was reduced in subjects given Classification Learning, suggesting that classification tends to induce attention to exception-feature values to a larger degree than does inference. Third, subjects' performance on the transfer tasks was generally better when the learning task and the transfer task matched, indicating that the representation acquired in each learning procedure may be specific to the corresponding learning task, while these representations retain a minimum level of flexibility to cope with both inference and classification transfer tasks. In particular, the finding that subjects in Mixed Learning excelled both in inference and classification transfer tasks seems to suggest that the combination of the two learning tasks produces a category representation rich enough to deal with both inference and classification transfer tasks at a high level of accuracy.

These results are consistent with the hypothesis that subjects employ different strategies to deal with the inference task and with the classification task, though they are not conclusive to rule out alternative interpretations. In particular, subjects' familiarity with each transfer task might have contributed to the observed results of Experiment 1. For example, the different levels of prototype-accordance responses might have emerged in subjects of Inference Learning and of Classification Learning because they differ in terms of their familiarity with the inference transfer task.

The purpose of Experiment 2 is to rule out this possibility and to contrast directly the characteristics of category representations produced in Inference Learning and in Classification Learning. In this experiment, we had subjects participate both in Inference Learning and in Classification Learning in sequence but in different orders. One group of subjects learned the categories by the Inference Learning task first, and then they learned the same categories by the Classification Learning task (i.e., Inference-first condition). The other group of subjects learned the same categories by the Classification Learning task first, and then they learned the same categories by the Inference Learning task (i.e., Classification-first condition). As in Experiment 1, subjects continued in one learning task until they reached the 90% accuracy criterion or they spent 30 blocks (240 trials) in total. In this setting, we predict that subjects will find it easier to learn the categories in the Inference-first condition than in the Classification-first condition, provided that Inference Learning leads to the acquisition of a category representation consistent with family resemblance information, and that Classification Learning results in the acquisition of a category representation consistent with rules and exceptions or concrete exemplars.

We hypothesized earlier that Classification Learning facilitates the formation of categories involving to rules and exceptions or concrete exemplars, while Inference Learning induces the acquisition of categories in accordance with the information about family resemblance within a category. Because a small number of features assessed in Classification Learning would not be sufficient to answer the inference questions on all four feature dimensions, subjects in the Classification-first condition may need to store extra rules and exceptions or concrete exemplars to cope with the subsequent Inference Learning trials. By contrast, prototype information (e.g., A0(1, 1, 1, 1) and B0(0, 0, 0, 0)) processed in Inference Learning is also useful for dividing the exemplars into the two classes in this setting, so that the representation obtained from Inference Learning can be applied in the subsequent Classification Learning task without modifying its characteristics extensively. As a consequence, we predict that learning the categories starting from Inference and followed by Classification should be less cumbersome than learning the categories in the reverse order. In other words, if Inference Learning produces a category representation consistent with family resemblance information, and if Classification Learning produces a category representation consistent with rules and exceptions or concrete exemplars, subjects should require fewer trials to reach the two learning criteria in the Inference-first condition than in the Classification-first condition. Following the same line of logic, given Exception-feature questions, subjects in the Inference-first condition should choose prototype-accordance features more often than should subjects in the Classification-first condition because subjects in the Inference-first condition would form categories according to family resemblance information while subjects in the Classification-first condition would form categories corresponding to rules and exceptions or concrete exemplars.

EXPERIMENT 2

Method

Participants. Participants were 57 members of the Columbia University community, who participated in the experiment for the payment of $6.00.[8] The data from 9 subjects were removed from the analyses—6 for failing to complete the experiment, 2 due to a coding error, and 1 for not following the instructions. Thus, we were left with 48 subjects (24 per condition).

Materials. The stimuli used for this experiment were identical to those used in Experiment 1.

Procedure. The procedure for this experiment was identical to that employed in Experiment 1 except for the following key manipulation. In the present experiment, all the subjects went through both Classification and Inference Learning, but in different orders. Half of the subjects were assigned to the Inference-first condition performing the Inference Learning task first until they reached the learning criterion—above 90% accuracy over three successive blocks (24 trials)—or they completed 30 blocks (240 trials). After reaching the criterion in Inference Learning they performed Classification Learning until they reached the same learning criterion or completed 30 blocks of trials. The other half of the subjects (i.e., the Classification-first condition) received the two tasks in the reverse order.

Design. The experiment was designed with a between-subject factor consisting of two levels: Inference-first (Inference learning followed by Classification learning) and Classification-first (Classification learning followed by Inference learning). The same dependent measures as used in Experiment 1 served for analyses.

[8] Originally, the data from 81 subjects were collected for this experiment. Due to an experimenter error, 24 subjects who participated in one condition were dropped.

TABLE 2

The Number of Learning Blocks Spent in Each Learning Task of Experiment 2

Learning order	Inference	Classification	Total
Inference-first	7.9	7.8	15.7
	Classification	Inference	Total
Classification-first	12.5	9.2	21.7

Note. One block contains 8 trials.

Results and Discussion

To assess the relative difficulty of the two learning orders, we first examined the number of blocks required to reach the learning criterion. The data are summarized in Table 2 and were analyzed with a 2(Learning order—Inference-first vs Classification-first) $\times$ 2(Learning type—Inference vs Classification) ANOVA. In all, 20 subjects in the Inference-first condition and 18 subjects in the Classification-first condition reached both criteria. As predicted, learning the categories starting from Inference was easier than learning the same categories starting from Classification. Subjects in the Inference-first condition required fewer learning blocks to reach the criterion in both tasks ($m = 15.7$) than did subjects in the Classification-first condition ($m = 21.7$); $F(1,36) = 5.01$, $MSE = 34.8$, $p < 0.05$. The results are consistent with the idea that Inference Learning and Classification Learning produce distinct category representations. The interaction between Learning order (Inference-first vs Classification-first) and Learning type (Inference vs Classification) was not significant, $F(1,36) = 1.41$, $MSE = 39.64$, $p > 0.10$. To examine the effect of the two learning orders further, we compared the number of blocks required to reach the learning criterion as a function of each learning task. In Classification Learning, subjects required significantly fewer

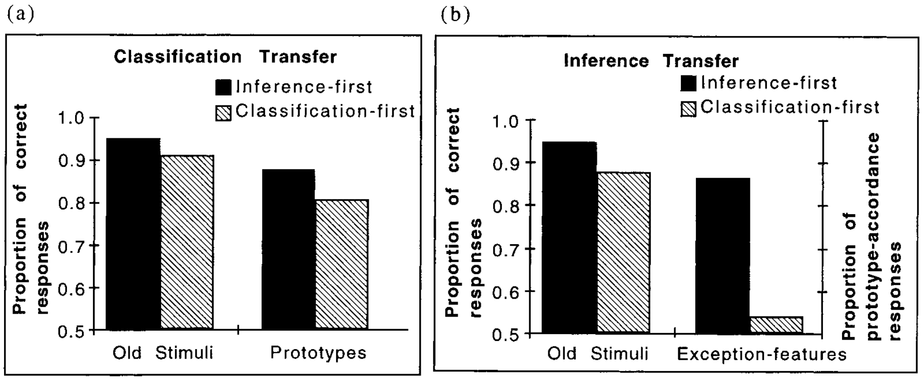

FIG. 3. (a) The classification transfer performance for old stimuli and prototype stimuli of Experiment 2. (b) The inference transfer performance for old stimuli and Exception-feature stimuli of Experiment 2.

blocks to reach the learning criterion if Classification Learning was given after Inference Learning (m = 7.8) than if it was given before Inference Learning (m = 12.5); $t(36) = 2.64$, $p < 0.05$. In contrast, in Inference Learning, the number of blocks required to reach the learning criterion did not differ significantly whether Inference Learning was given before Classification Learning (m = 7.9) or it was given after Classification Learning (m = 9.2); $t(36) = 0.62$, $p > 0.10$. The results suggest that Inference Learning helped subjects to cope with the subsequent Classification Learning task while Classification Learning did not help the subsequent Inference Learning task.

The results obtained from transfer performance, which are summarized in Fig. 3, were also consistent with the idea that Classification Learning and Inference Learning yield distinct categories. Although subjects in the two learning orders did not differ in the classification of old stimuli—Inference-first (m = 0.95), Classification-first (m = 0.91), $t(36) = 1.36$, $p > 0.10$—and in the classification of prototype stimuli—Inference-first (m = 0.88), Classification-first (m = 0.81), $t(36) = 0.64$, $p > 0.10$—, subjects in the Inference-first condition (m = 0.95) were significantly more accurate in making inferences to old stimuli than were subjects in the Classification-first condition (m = 0.88), $t(36) = 2.89$, $p < 0.01$. More importantly, consistent with our prediction, subjects in the Inference-first condition made Exception-feature inferences in accordance with prototype-feature values more often (m = 0.86) than did subjects in the Classification-first condition (m = 0.54), $t(36) = 2.96$, $p < 0.01$, implying that category representations formed in the Inference-first condition and those in the Classification-first condition differed significantly in the degree that family resemblance information was incorporated into their representations.

Further support for this interpretation was found in a subject-analysis in which 17 out of 20 subjects in the Inference-first condition made responses in accordance with the prototype-feature values more than 75% of the time, and only 3 out of 20 subjects made responses in accordance with exception-feature values more than 75% of the time. In contrast, 8 out of 18 subjects in the Classification-first condition chose feature values in accordance with prototypes more than 75% of the time, and 6 out of 18 subjects made responses in accordance with exception-feature values more than 75% of the time. Thus, the results of the Exception-feature inferences reveal that subjects in the two learning orders differed in the proportion of prototype-accordance responses.

To examine further whether subjects in the Classification-first condition learned the categories by augmenting single feature information to cope with inference questions, we conducted a post hoc analysis for the inference transfer of old stimuli. In particular, we measured the standard deviations of correct responses of four feature dimensions, taking the data from individual subjects as random variables. The rationale for this analysis is that subjects selectively attending to a small number of features will show high variability of correct responses over the four feature dimensions, exhibiting accurate performance for one feature dimension but not for the others. In contrast, subjects attending to four feature-dimensions equally would display the same level of performance in all the four dimensions, resulting in less variability between feature dimensions. Subjects exhibited higher variability in the Classification-first condition ($m = 0.14$) than in the Inference-First condition ($m = 0.08$); $t(36) = 2.69$, $p < 0.05$. The results of this analysis provide additional support for our hypothesis that classification makes use of a small number of diagnostic features.

Taken together, these results of Experiment 2 are consistent with the view that subjects form different category representations in the two learning orders, supporting the hypothesis that Classification Learning and Inference Learning give rise to the acquisition of distinct category representations. In Classification Learning, subjects seem to obtain a category representation congruent with a small number of diagnostic features and exceptions, while in Inference Learning, subjects tend to obtain a category representation congruent with family resemblance information. Because the information about a diagnostic feature and exceptions is not sufficient to answer all inference questions, and because the information about family resemblance between exemplars is suitable to deal with both inference and classification questions, subjects required more trials to reach the learning criterion in the two learning tasks when categories were learned by classification first followed by inference than by the reverse order.

Some may argue that the effect observed in Inference Learning and in Classification Learning is a result of a link between encoding and retrieval tasks. In Inference Learning, subjects answered inference questions of all the four feature dimensions whereas in Classification Learning subjects answered questions about the category labels only. On this basis, the observed distinction between Inference Learning and Classification Learning may be an artifact of this experimental setting because what is learned in the two learning procedures can be no more than a link between the encoding task and the retrieval task (Estes, 1976, 1986; Medin & Schaffer, 1978; Tulving, 1983; Tulving & Thomson, 1973; see also Roediger, 1989). This explanation, though plausible, cannot account for why subjects in the two learning orders exhibited drastically different transfer performance when the order of the two learning procedures was changed. In this study, all the subjects in the two conditions learned the same categories by the same tasks—Inference Learning and Classification Learning—with the identical stimuli. Nonetheless, subjects acquired different categories if the two learning procedures were given in different orders. It is difficult to see how a simple link between encoding and retrieval procedures caused the disparity between the two learning orders. By the same token, the results of Experiment 2 cannot be explained by subjects' familiarity with the corresponding learning task. We argue that the specific mechanisms tied to inference and classification were one of the main determinants of the observed results in the two learning orders.

FITTING CATEGORIZATION MODELS

In the two experiments, we have examined the distinction between inference and classification solely by contrasting the effect of the three learning procedures. The results of the two experiments are consistent with the idea that Inference Learning and Classification Learning lead to the formation of distinct categories, supporting the view that subjects employ different strategies to make judgments related to inference or classification. This conclusion can be tested further by comparing subjects' performance within each learning condition. That is, if our argument is sound, performance for inference transfer tasks and performance for classification transfer tasks should differ considerably even within a single learning condition. We tested this idea by fitting existing models of classification to the data obtained from inference transfer tasks as well as from classification transfer tasks. If inference is carried out by the same process as by the one employed in classification, existing models of classification should be able to account for the data from classification transfer and from inference transfer equally well.

We fit Medin and Schaffer's context model (1978) and Anderson's rational model to the data (Anderson, 1990, 1991; Nosofsky, 1986). To fit the context model, we employed Nosofsky's Generalized Context Model (GCM) because of its generality and clarity (Nosofsky, 1986). We chose the rational model because of its proposal that the primary impetus for categorization is to maximize people's ability to make inferences about features of category members. The rational model treats the category label as another feature and assumes that it can be predicted in the same manner that the features of objects are predicted. The purpose of this model fitting is to test the central assumption that subjects employ qualitatively different strategies to carry out the inference and classification tasks. In this sense, this model fitting is not intended to compare the relative efficacy of the two models. The comparison between the two models is beyond the scope of the present paper because these models were developed to account for data from classification tasks not from inference tasks.

The results of the model fitting for Experiment 1 are summarized in Tables 3 and 4, and are described in the following subsections. The results of the model fitting for Experiment 2 are shown in the Appendix. For both models, parameters were found by a random-start hill-climbing algorithm that looked for the minimum values of the sum of squared deviation between predicted and observed values (SSE). The data from classification transfer and inference transfer were examined separately in each learning condition. There were 10 data points in the classification transfer of Experiments 1 and 2, and there were 32 data points in the inference transfer of Experiments 1 and 2.

Fitting the Context Model and the Rational Model

The GCM has eight parameters—*c, r, b1, w1, w2, w3, w4,* and *w5*. According to the model, the probability that subjects classify the stimulus *Si* into the category C_1, $P(C_1/S_i)$, is obtained by calculating the overall similarity between the stimulus *Si* and the category members in C_1 divided by the sum of overall similarity between the stimulus *Si* and all the members of categories available to subjects (see Nosofsky, 1986, p. 42 for the modification of the GCM to relate the model to the context model):

$$P(C_1/S_i) = \frac{b_1 \sum\limits_{j\in C_1} n_{ij}}{b_1 \sum\limits_{j\in C_1} n_{ij} + (1 - b_1) \sum\limits_{l\in C2} n_{il}},$$

where

$$n_{ij} = e^{-(C[\sum_{k=1}^{N} w_k|x_{ik}-x_{jk}|^r]^{1/r})^r} = e^{-C^r \sum_{k=1}^{N} w_k|x_{ik}-x_{jk}|^r}.$$

The similarity between the probe stimulus S_i and an exemplar stimulus S_j is denoted by n_{ij}, which decreases exponentially as a function of the discrepancy between feature values $|x_{ik} - x_{jk}|$. The parameters, w_k, represent the selective attention given to each feature dimension. r is the parameter associated with the psychological distance between feature values. c is the scale parameter representing overall discriminability of stimuli in a psychological space. In order to reduce the number of free parameters, we fixed two parameters b_1 and r($b_1 = 0.5$ and $r = 1$) and combined c and w_k. This modification yields five free parameters cw_k($0 < cw_k < \infty$) and makes the GCM identical to

Medin and Schaffer's context model (1978) where Medin and Schaffer's similarity parameters, f', s', c', and p', correspond to the attention parameters cw_k in logarithmic functions (in this setting, $cw_1 = -\ln f'$, $cw_2 = -\ln s'$, $cw_3 = -\ln c'$, $cw_4 = -\ln p'$).

To fit inference data, we slightly extended the model by treating category labels as another feature and calculated feature inferences in the same manner that classification performance was calculated. For example, when predicting classification performance, the similarity distance between a probe item (1, 1, 1, 0, 1) = (form, size, color, position, category-label) and a category item (0, 1, 1, 1, 1) was obtained by estimating the feature match excluding the category labels; in this case

$$cw_1|x_{i1} - x_{j1}| + cw_2|x_{i2} - x_{j2}| + cw_3|x_{i3} - x_{j3}| + cw_4|x_{i4} - x_{j4}|.$$

Similarly, when predicting the feature value of form, for example, the similarity distance between the two items was obtained by feature match excluding form but including category labels; in this case

$$cw_2|x_{i2} - x_{j2}| + cw_3|x_{i3} - x_{j3}| + cw_4|x_{i4} - x_{j4}| + cw_5|x_{i5} - x_{j5}|,$$

where cw_5 is an attention parameter given to category labels. To obtain psychological distance associated with each feature value (i.e., x_{ij}), the GCM requires stimulus identification data and a confusion matrix. These data were not available to us so that we used the arbitrary feature values (1 and 0) in each feature dimension.

Anderson's rational model is based on the assumption that people form categories to maximize the predictability to features of objects (Anderson, 1991). A main vehicle of the rational model lies in the internal partitions that are formed through experience with exemplars. Anderson (1990) describes two important qualities of the internal partitions. First, the clusters emerge because objects in the world have inherent qualities for clustering due to the inability to crossbreed. Second, the clusters reveal information about their members, showing, for example, the probability that a particular object exhibits a certain feature value. Anderson suggests that it is this second quality of partitions that enables people to draw predictions about objects and to classify a new object into a group (Anderson, 1990, p. 97).

To examine classification and inference performance, we calculated the probability estimated by

$$P(ij|F) = \sum_k P(k|F)P(ij|k),$$

where $P(ij|F)$ is the probability that a stimulus has a feature value j on the dimension i given the feature structure F, $P(k|F)$ is the probability that the stimulus is grouped in the partition k given its feature structure F, and $P(ij|k)$ is the probability that the stimulus has a feature value j on a feature dimension i given a partition k. The rational model has a coupling parameter c that affects the prior probability that an item comes from a particular partition. We further introduced five parameters associated with four feature dimensions and category labels to accommodate attention salience given to each feature dimension (we treated category labels as another feature dimension). These parameters yield a new equation to estimate $P(ij|k)$—the probability of displaying a feature value in a given partition k (see Anderson, 1990, p. 116 for this modification)

$$P(ij|F) = \frac{n_{ij} + g_i}{n_k + \sum g_i}.$$

To give the model extra flexibility, we also introduced a response parameter r (the same modification can be found in Nosofsky, Gluck, Palmeri, Mckinley and Glauthier, 1994, p. 359). Thus, the probability P_{iA} that

TABLE 3

The Results of Model Fitting for Experiment 1 (MCM)

	Classification transfer							
Learning	cw_1	cw_2	cw_3	cw_4		SSE	Accountability	Correlation
Inference	1.58	1.06	1.64	0.19		0.021	0.83	0.92
Mixed	1.78	1.55	1.70	0.87		0.040	0.20	0.46
Classification	1.15	3.43	2.16	1.08		0.031	0.41	0.64
	Inference transfer							
Learning	cw_1	cw_2	cw_3	cw_4	cw_5	SSE	Accountability	Correlation
Inference	0.01	0.01	0.01	0.01	2.41	0.10	−0.01	−0.61
Mixed	0.01	0.01	0.01	0.01	2.21	0.38	−0.01	−0.77
Classification	0.05	0.01	0.01	0.39	1.12	0.48	0.04	0.20

Note. cw_1–cw_4 are the parameters associated with feature dimensions (form, size, color, position). cw_5 is the parameter given to category labels. SSE is the sum of squared difference between predicted and observed values. SST is the sum of squared difference between the mean of observed values and the observed values. Accountability = 1 − (SSE/SST); MCM, modified context model.

subjects predict the feature value a in the ith dimension is estimated by

$$P_{iA} = \frac{P_{ia}^r}{(P_{ia}^r + P_{ib}^r)},$$

where p_{ia} and p_{ib} are the probability estimated by the rational model to predict the feature values a and b in the ith dimension, respectively. Because the order of stimulus presentation also affects the accountability of the rational model, 40 sequences of stimulus presentation were generated randomly, and optimal parameter values were determined separately for each presentation. The results reported below and the accountability score shown in Table 4 are based on the best fitting parameter values obtained from one of forty sequences of stimulus presentation.

Results of Model Fitting (Experiment 1)

Tables 3 and 4 summarize the results of the model fitting of the modified context model and of the rational model, respectively. Overall, the two models provide reasonably good fits to the data obtained from classification transfer performance of Inference Learning and of Classification Learning, but not of Mixed Learning (see Tables 3 and 4 for details). A similar trend appeared in the model fitting of Experiment 2 (see Appendix).[9] In contrast, the two models appear inappropriate to account for inference transfer data regardless of the learning conditions.

In accounting for inference transfer performance, for example, the modified context model produced attention parameters close to the minimum value for almost all the four feature dimensions except the category labels ($cw_i = 0.01$, $0 < cw_k < \infty$). The results indicate that the model relied on the category labels almost exclusively to derive inference judgments. Furthermore, within the range of the attention parameters the model's account-

[9] The poor performance observed in the two models in accounting for classification transfer performance in Experiment 2 might have derived primarily from their performance for the prototype stimuli. Taking out the predictions made to the prototype stimuli, the average correlations between the observed values and the predicted values reach 0.75 in the two learning orders.

TABLE 4

The Results of Model Fitting for Experiment 1 (The Rational Model)

	Classification transfer										
Learning	g_1	g_2	g_3	g_4	g_5	c	r	SSE	Accountability	Correlation	Partitions
Inference	0.648	1.656	0.508	7.526	0.002	0.26	1.08	0.017	0.86	0.93	2
Mixed	0.128	0.216	0.128	1.248	0.152	0.01	1.41	0.039	0.23	0.48	8
Classification	0.340	0.018	0.028	0.020	0.036	0.08	1.10	0.016	0.69	0.84	3
	Inference transfer										
Learning	g_1	g_2	g_3	g_4	g_5	c	r	SSE	Accountability	Correlation	Partitions
Inference	0.126	0.038	0.076	0.126	0.148	0.05	1.60	0.058	0.43	0.65	8
Mixed	0.240	0.164	0.222	0.294	0.012	0.01	2.85	0.152	0.60	0.78	8
Classification	0.180	0.090	0.032	0.392	0.064	0.01	1.08	0.223	0.56	0.76	8

Note. g_1–g_4 are the parameters given to feature dimensions (form, size, color, position). g_5 is the parameter for category labels. c is the coupling parameter and r is the parameter for the response function. SSE is the sum of squared difference between predicted and observed values. Partions are the number of internal partitions. SST is the sum of squared difference between the mean of observed values and the observed values. Accountability = 1 − (SSE/SST). The values introduced here represent the best fitting values among 40 different random sequences of stimulus presentation.

ability measures (i.e., accountability = 1 − (SSE/SST)) barely exceeded the estimates made by the overall averages of the data. Clearly, given the modification suggested in this section, the algorithm employed in the modified context model does not seem appropriate to account for inference transfer performance.

Similarly, the algorithm used in the rational model appears implausible in accounting for the data from the inference transfer task. It produced the best fitting values in all three learning conditions by creating singleton partitions for every learning stimulus (i.e., each internal partition contained only one exemplar). Among the 40 different patterns of randomly selected stimulus-presentation sequences, the rational model produced the best parameter values by using singleton partitions in 17 out of 40 cases in Classification Learning, 30 out of 40 cases in Inference Learning, and 34 out of 40 cases in Mixed Learning. These results suggest that the model obtained the best predictions by examining each exemplar separately rather than by forming clusters, indicating that the accountability of the model actually decreases as the model forms internal clusters. This phenomenon contradicts the basic assumption of the rational model that internal clusters formed by people provide a basis for feature predictions (Anderson, 1990, p. 97). Given the family resemblance category-structure employed in the two experiments, there is no reason to believe that the stimulus structure used in the two experiments deters the formation of partitions. To account for the inference transfer data, the two models may need to introduce a major modification. While we believe that fitting models of categorization can provide an insight into the distinction between inference and classification, we think that it is premature for us to develop a model of classification and inference in this stage. More empirical studies and theoretical investigation should be made to constrain the nature of the inference process before a model of the inference task can be developed.

In summary, the disparity between inference and classification is clearly present in the results of the model fitting: the two models provided reasonably good fit to the data from the classification transfer tasks, but failed to account for the data from the inference transfer tasks, at least given the modification suggested above. It may be the case that the assumption of treating category labels as equivalent to category features may not be warranted. Although the two models have been successful in accounting for a variety of classification performance (Anderson, 1990, 1991; Medin & Schaffer, 1978; Nosofsky, 1986), these models seem to require a major modification to account for inference transfer data.

GENERAL DISCUSSION

In an effort to investigate the relationship between category learning and category formation, we have contrasted inference-based learning with classification-based learning. The results of the two experiments and the model fitting show that inference and classification require different strategies to carry them out, and, because of these strategies, distinct category representations arise if people learn categories by inference or by classification. In particular, inference, which requires a focus on exemplar information within a category, helps subjects to extract family resemblance information within a category. As a result, categories formed by inference contain information consistent with the prototypical values of the category members. In contrast, classification, which tends to promote a focus on a small number of diagnostic features, guides subjects to form categories consistent with rules and exceptions or concrete exemplars.

In Experiment 1, we found that subjects' transfer performance was the best when the transfer task matched the learning task. Subjects who learned the categories by classification were the best in classification transfer. Subjects who learned the categories by inference were the best in inference transfer. Given Exception-feature inferences, subjects in the three learning conditions predominantly responded with prototype stimuli, though this tendency was reduced significantly in subjects who learned categories by classification. In Experiment 2, we found that the order in which subjects received the two learning procedures had a significant impact on subjects' performance. Subjects found these categories much easier to learn when they learned the categories by inference first and followed by classification than when they learn the same categories in the reverse order. Finally, in the model fitting, we demonstrated that the distinction between inference and classification is present within each learning procedure.

The results suggest that the nature of category formation can be specified by the task applied during learning, indicating that the acquisition of categories is inseparable from the function of categories (see also Kintsch, 1980; Schank, 1982; Wittgenstein, 1953). For this reason, we argue for the need to go beyond the study of classification as a mode of category learning to look at the impact of other uses of categories on what is learned (see also Whittlesea et al., 1994, for a similar argument). Other studies have also begun to address this issue. Ross (1996) had subjects classify algebra equations. He found that the classes formed by subjects were different if the subjects had previously manipulated the equations by solving for a variable than if they had not. Some work has also addressed the role of communication in category formation. Markman and Makin (in press) (Markman, Yamauchi, & Makin, 1997) report studies in which pairs of subjects were asked to build LEGO models collaboratively. One subject was given pictorial instructions for constructing a model, and the other subject was given the pieces needed to build the model (as well as distractor pieces). Subject pairs had to settle on a common set of labels for the pieces in order to carry out this task. After building the model, subjects were asked to sort the pieces into groups. Analysis of the sorting data revealed a higher level of agreement between subjects who communicated together than between subjects who did not communicate with each other and built different models.

The present results are a preliminary step toward understanding the relationship between category learning and category formation. Further research must focus on what is learned about categories through other uses of categories such as communication, inference, comparison, and memory (Markman et al., 1997). Research must also examine the distinction between inference and classification, while addressing precisely what is learned from these tasks. Further, we examined only one category structure with the same stimulus set. Studies must explore the impact of these variables. Additional work must also focus on how inference and classification (and other modes of category functions) are integrated to form coherent category representations.

In conclusion, although inference and classification are closely related, the two functions require different strategies to be incorporated. The present experiments suggest that these different strategies, which are related to the two functions of categories, give rise to the formation of distinct category representations.

APPENDIX

The first two tables are the results of model fitting obtained by the modified context model and the rational model for the data from Experiment 2. The last table is the predicted and observed values of the transfer performance for Experiment 1: MCM, modified context model; RM, rational model; (F, S, C, P), (form, size, color, position); cw_1–cw_4 are the parameters associated with feature dimensions (form, size, color, position); cw_5 is the parameter for category labels; g_1–g_4 are the parameters given to feature dimensions (form, size, color, position); g_5 is the parameter for category labels; c is the coupling parameter and r is the parameter for the response function; SSE, sum of squared difference between predicted and observed values; SST, sum of squared difference between the mean of observed values and the observed values; accountability = 1-(SSE/SST).

TABLE A1

Experiment 2 (Modified Context Model)

Learning	Classification transfer							
	cw_1	cw_2	cw_3	cw_4		SSE	Accountability	Correlation
Inference-first	1.96	3.16	1.44	1.54		0.033	−0.30	0.05
Classification-first	0.80	1.78	2.60	1.44		0.104	−0.12	0.16
	Inference transfer							
Learning	cw_1	cw_2	cw_3	cw_4	cw_5	SSE	Accountability	Correlation
Inference-first	0.01	0.01	0.01	0.25	2.47	0.14	0.01	0.09
Classification-first	0.01	0.01	0.01	0.01	1.35	0.88	−0.01	−0.88

TABLE A2

Experiment 2 (Rational Model)

Learning	Classification transfer										
	g_1	g_2	g_3	g_4	g_5	c	r	SSE	Accountability	Correlation	Partitions
Inference-first	0.006	0.006	0.014	0.024	0.090	0.1	1.01	0.024	0.06	0.52	5
Classification-first	0.010	0.008	0.004	0.038	0.096	0.09	0.74	0.065	0.30	0.55	5
	Inference transfer										
Learning	g_1	g_2	g_3	g_4	g_5	c	r	SSE	Accountability	Correlation	Partitions
Inference-first	0.102	0.058	0.032	0.210	0.064	0.03	1.72	0.081	0.42	0.65	8
Classification-first	0.422	0.604	0.324	0.378	0.016	0.01	2.99	0.184	0.79	0.89	8

TABLE A3

Classification transfer (Experiment 1)

	Inference			Mixed			Classification		
Stimulus	Observed	MCM	RM	Observed	MCM	RM	Observed	MCM	RM
A1	0.864	0.895	0.895	0.850	0.920	0.921	1.000	0.958	0.988
A2	0.636	0.701	0.686	0.950	0.853	0.850	1.000	0.893	0.989
A3	0.818	0.789	0.801	0.850	0.866	0.875	0.826	0.848	0.800
A4	0.773	0.709	0.708	0.900	0.846	0.850	1.000	0.955	0.993
B1	0.909	0.895	0.895	1.000	0.920	0.921	0.913	0.958	0.886
B2	0.727	0.701	0.686	0.750	0.853	0.850	0.783	0.893	0.887
B3	0.773	0.789	0.801	0.900	0.866	0.875	0.870	0.848	0.856
B4	0.636	0.709	0.708	0.800	0.846	0.850	0.957	0.955	0.913
A0	0.955	0.907	0.931	0.900	0.951	0.927	0.957	0.978	0.994
B0	0.955	0.907	0.931	0.950	0.951	0.927	0.957	0.978	0.975

Inference transfer (Experiment 1)

		Inference			Mixed			Classification		
Stimulus	Questions	Observed	MCM	RM	Observed	MCM	RM	Observed	MCM	RM
A1	F	0.909	0.918	0.911	0.800	0.902	0.935	0.652	0.718	0.721
A2	F	0.909	0.918	0.930	1.000	0.902	0.950	0.652	0.788	0.848
A3	F	0.955	0.918	0.944	0.850	0.902	0.962	0.826	0.788	0.819
A4	F	0.818	0.919	0.859	0.700	0.902	0.757	0.522	0.790	0.671
B1	F	0.909	0.918	0.911	0.900	0.902	0.935	0.652	0.718	0.721
B2	F	0.818	0.918	0.930	0.900	0.902	0.950	1.000	0.788	0.848
B3	F	0.955	0.918	0.944	1.000	0.902	0.962	0.739	0.788	0.819
B4	F	0.773	0.919	0.859	0.550	0.902	0.757	0.478	0.790	0.671
A1	S	0.955	0.918	0.945	0.900	0.902	0.954	0.696	0.723	0.755
A2	S	0.909	0.918	0.961	0.950	0.902	0.967	0.957	0.791	0.894
A3	S	0.909	0.919	0.857	0.800	0.902	0.747	0.609	0.793	0.667
A4	S	0.909	0.918	0.945	1.000	0.902	0.964	0.870	0.785	0.822
B1	S	0.909	0.918	0.945	0.900	0.902	0.954	0.696	0.723	0.755
B2	S	0.909	0.918	0.961	0.950	0.902	0.967	0.913	0.791	0.894
B3	S	0.864	0.919	0.857	0.700	0.902	0.747	0.696	0.793	0.667
B4	S	1.000	0.918	0.945	1.000	0.902	0.964	0.870	0.785	0.822
A1	C	1.000	0.918	0.930	0.900	0.902	0.940	0.783	0.723	0.782
A2	C	0.909	0.919	0.858	0.850	0.902	0.755	0.696	0.793	0.662
A3	C	1.000	0.918	0.961	1.000	0.902	0.965	0.957	0.791	0.897
A4	C	1.000	0.918	0.930	0.950	0.902	0.951	0.870	0.785	0.854
B1	C	0.909	0.918	0.930	1.000	0.902	0.940	0.783	0.723	0.782
B2	C	0.864	0.919	0.858	0.750	0.902	0.755	0.739	0.793	0.662
B3	C	1.000	0.918	0.961	1.000	0.902	0.965	0.913	0.791	0.897
B4	C	0.955	0.918	0.930	0.900	0.902	0.951	0.826	0.785	0.854
A1	P	0.864	0.919	0.859	0.900	0.902	0.762	0.739	0.760	0.675
A2	P	0.955	0.918	0.930	0.950	0.902	0.937	0.957	0.759	0.773
A3	P	0.955	0.918	0.944	1.000	0.902	0.951	0.739	0.759	0.748
A4	P	0.955	0.918	0.911	1.000	0.902	0.933	0.696	0.751	0.719
B1	P	0.864	0.919	0.859	0.800	0.902	0.762	0.652	0.760	0.675
B2	P	0.909	0.918	0.930	0.950	0.902	0.937	0.826	0.759	0.773
B3	P	0.955	0.918	0.944	1.000	0.902	0.951	0.739	0.759	0.748
B4	P	0.864	0.918	0.911	1.000	0.902	0.933	0.826	0.751	0.719

REFERENCES

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

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

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

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review,* **94,** 115–147.

Estes, W. K. (1976). Structural aspects of associative models for memory. In C. N. Cofer (Ed.), *The structure of human memory* (pp. 31–53). New York: Freeman.

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

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

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

Gelman, S. (1988). The development of induction within natural kind and artificial categories. *Cognitive Psychology,* **20,** 65–95.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical Reasoning* (pp. 199–241). New York: Cambridge University Press.

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review,* **97,** 3–18.

Harnad, S. (1987). Introduction: Psychological and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical perception* (pp. 1–28). New York: Cambridge University Press.

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

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

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps.* Cambridge, MA: MIT Press.

Kintsch, W. (1980). Semantic Memory: A Tutorial. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 595–620). Hillsdale, NJ: Erlbaum.

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

Malt, B. C. (1989). An on-line investigation of prototype and exemplar strategies in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **15,** 539–555.

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. (in press). Referential communication and category acquisition. *Journal of Experimental Psychology: General.*

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. 174–208). Washington, DC: American Psychological Association.

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

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology,* **19,** 242–279.

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

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **15,** 282–384.

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.

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

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

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

Roediger, I., H. L., , Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. R. III & C. F. I. M. (Eds.), *Variety of memory and consciousness: Essays in honor of Endel Tulving* (pp. 3–41). Hillsdale, NJ: Erlbaum.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology,* **8,** 382–439.

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., & Murphy, G. L. (1996). Category-based predictions: Influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **22**(3), 763–753.

Schank, R. C., Collins, G. C., & Hunter, L. E. (1986). Transcending inductive category formation in learning. *Behavioral and Brain Sciences,* **9,** 639–686.

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, E. E. (1994). Concepts and categorization. In E. E. Smith & D. N. Osherson (Eds.), *An invitation to cognitive science* (pp. 3–33). Cambridge, MA: MIT Press.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Tulving, E. (1983). *Elements of episodic memory.* New York: Oxford University Press.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review,* **80,** 352–373.

Tversky, A. (1977). Features of similarity. *Psychological Review,* **84,** 327–352.

Whittlesea, B. W. A., Brooks, L. R., & Westcott, C. (1994). After the learning is over: Factors controlling the selective application of general and particular knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **20,** 259–274.

Wittgenstein, L. (1953). *Philosophical investigation.* New York: Macmillan.

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

Yamauchi, T., & Markman, A. B. (in preparation). Processes underlying inference and classification using categories.

(Received November 6, 1997)
(Revision received January 9, 1998)