

Learning categories composed of varying instances: The effect of classification, inference, and structural alignment

TAKASHI YAMAUCHI and ARTHUR B. MARKMAN
Columbia University, New York, New York

The members of a natural category are not usually identical in their appearance, although at some level they can be described as having features in common. For example, birds have wings, but the actual appearance of their wings varies from one bird to another. To examine the effect of this feature variation on category acquisition, subjects in three experiments were asked to learn categories in which individual features were depicted with several different instances. The results of the experiments indicated that subjects had significant difficulty learning these categories when they were given a standard classification learning task. In contrast, subjects were able to acquire the same categories when they were given an inference learning task, in which they learned the categories by predicting a missing feature of a stimulus given the category label and information about the other features. Finally, subjects who were allowed to compare stimuli during learning were able to learn the categories. These results suggest that a common description of different instances emerges in the process of aligning stimuli.

The members of a category have features in common, but the appearance of these features is not always identical. All birds have wings, for example, but the actual instances of this feature vary from one item to another. Extracting a common relationship from different instances is crucial for categorization (Murphy & Medin, 1985; Wisniewski & Medin, 1991). However, it is not clear how the system for categorization finds commonalities and differences of feature values, distinguishing, for example, the wings of birds from the wings of airplanes or flying fish (Schyns, Goldstone, & Thibaut, 1998).

When learning categories whose features vary in appearance, people must discover the relevant similarities that go beyond the surface manifestation of the instances. For example, to recognize that different birds, such as robins, pelicans, and penguins, all have beaks, it is necessary to find some underlying commonality of these properties amid the surface differences. In the present study, we investigated the role of *structural alignment* (Markman & Gentner, 1997), a form of comparison, in category learning. We examined how subjects learn categories composed of varying feature instances in three different methods of category learning: classification

learning, inference learning, and mixed learning. In this paper, we begin by outlining the three learning procedures and then discussing the role of structural alignment in category learning (Gentner, 1983, 1989; Markman & Gentner, 1993). Finally, we describe three experiments in which subjects learned categories consisting of multiple feature instances in one of the three learning procedures.

Throughout this paper, we use the term *category label* to refer to a symbol that denotes a particular group of exemplars, and we use the term *category feature* to mean a symbol that represents a characteristic of an exemplar. *Classification* is defined as a practice in which an exemplar is placed into one of two groups (i.e., its category label is predicted) when the attributes of the exemplar are known. *Inference* is characterized as a practice in which an attribute of an exemplar (i.e., a category feature) is predicted when the group to which the exemplar belongs (i.e., the category label) and other attributes of the exemplar are known (for similar descriptions of inference, see Estes, 1994, Murphy & Ross, 1994, and Yamauchi & Markman, 1995, 1998). 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) on the basis of his/her attributes (e.g., supports affirmative action and favors reducing defense spending). In contrast, inference as we defined it is related to the situation in which people predict an attribute of a person (e.g., supports affirmative action) using a category label to which the person belongs and his/her other attributes (e.g., is a Democrat and favors reducing defense spending).

Overview of the Experiments

In all the experiments presented here, the stimuli were imaginary bugs (Figure 1a) having four features with bi-

This research was supported by NSF CAREER Award SBR-95-10924 to A.B.M. The authors thank Dedre Gentner, Dave Krantz, Douglas Medin, Page Moreau, Tomislav Pavlicic, Brian Ross, Yung-Cheng Shen, Thomas Ward, and Shi Zhang for their helpful suggestions and comments while this research was being carried out. Correspondence should be addressed to T. Yamauchi, who is now at the Learning Research and Development Center, University of Pittsburgh, Rm. 815, 3939 O'Hara St., Pittsburgh, PA 15260 (e-mail: takashi+@pitt.edu), or to A. B. Markman, who is now at the Department of Psychology, University of Texas, Austin, TX 78712 (e-mail: markman@psy.utexas.edu).

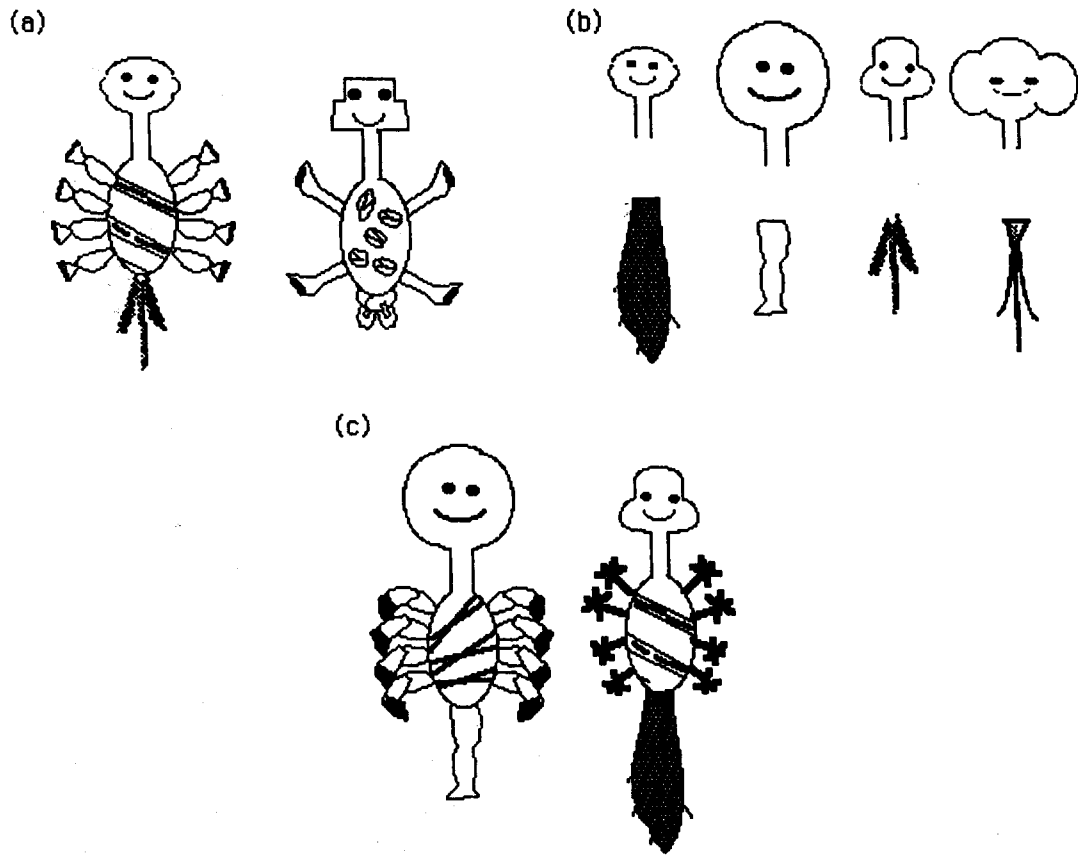


Figure 1. (a) Samples of two exemplars [e.g., A1 (1, 1, 1, 0) and B1 (0, 0, 0, 1)]. (b) Four different instances depicted each feature value in the multiple manifestation condition (e.g., round head and long tail). (c) The same exemplar [e.g., A1 (1, 1, 1, 0)] appeared different because these two items were shown by distinct feature instances in the multiple-feature-manifestation condition.

nary values: head (round, angular), body (striped, dotted), legs (eight legs, four legs), and tail (long, short) (see Medin, Wattenmaker, & Hampson, 1987, for a similar stimulus set). These stimuli were divided into two groups such that every exemplar shared three feature values with its corresponding prototype (A0 or B0) and one feature value with the prototype of the other category (these are *exception features*, which are shown in italics in Table 1).

In Table 1, the value of 1 or 0 for a particular feature dimension shows the feature value (e.g., round head vs. angular head). In the single-feature-manifestation condition, the two categories consisted of eight unique exemplars, and each feature value had only one instance (see Figure 1a). In the multiple-feature-manifestation condition, the two categories were also depicted by eight exemplars, but the feature values of each exemplar were shown by four different instances (Figure 1b). For example, Figure 1b shows the four feature values for "round head" and "long tail." In this condition, a single exemplar [e.g., A1(1, 1, 1, 0)] has $4 \times 4 \times 4 \times 4 = 256$ possible instantiations, so that the exemplars had the same feature structure as in the single-feature-manifestation condition, but they differed in appearance (Figure 1c).

All the experiments consisted of three phases: an initial learning phase, a filler phase, and a transfer phase. In the learning phase, subjects carried out one of the three learning tasks: inference learning, classification

Table 1
Stimulus Structure Used in the Experiments

Stimulus	Feature			
	Head	Body	Legs	Tail
Category Monek				
A1	1	1	1	0
A2	1	1	0	1
A3	1	0	1	1
A4	0	1	1	1
A0 (prototype)	1	1	1	1
Category Plaple				
B1	0	0	0	1
B2	0	0	1	0
B3	0	1	0	0
B4	1	0	0	0
B0 (prototype)	0	0	0	0

Note—Exception features, which have the feature value of the prototype of the opposite category, are shown in italics. The numbers 1 and 0 stand for the binary values of the features.

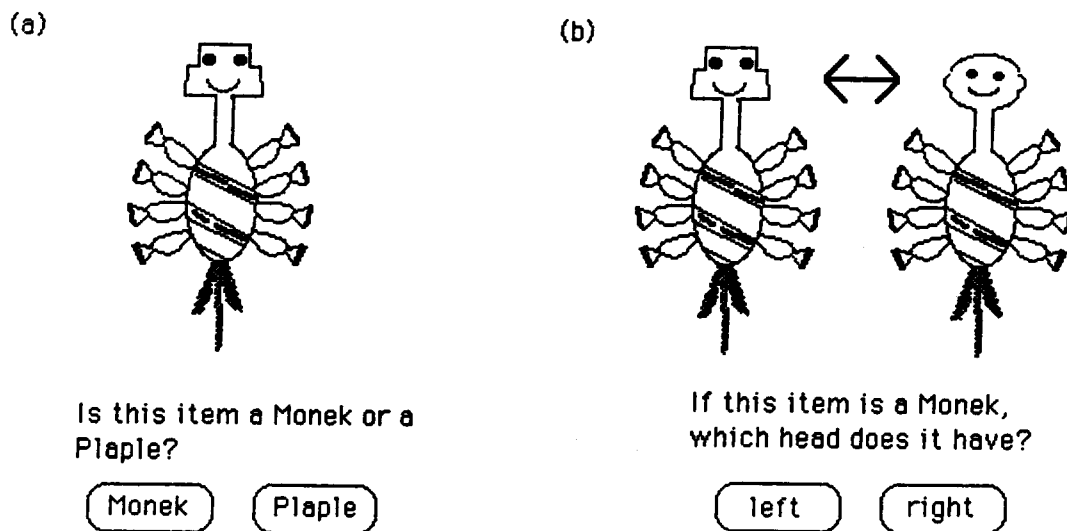


Figure 2. Sample stimulus frames used on (a) a classification trial of Experiment 1A and (b) an inference trial of Experiment 1A.

learning, or mixed learning. Classification learning was a variant of the standard classification task (e.g., Medin & Schaffer, 1978; Posner & Keele, 1968). Subjects saw a complete exemplar and predicted the label of the stimulus by clicking one of the buttons (Figure 2a). In inference learning, subjects were given the category label and some of the features of the exemplar, and they were asked to infer the value of the unknown feature (Figure 2b). Mixed learning was a mixture of the two learning conditions, in which half of the blocks were classification and half were inference. Subjects learned the two categories incrementally on the basis of feedback they received at the end of each trial (see Yamauchi & Markman, 1998). Some subjects learned categories in the single-feature-manifestation condition, and the other subjects learned categories in the multiple-feature-manifestation condition.

Inference, Classification, and Structural Alignment

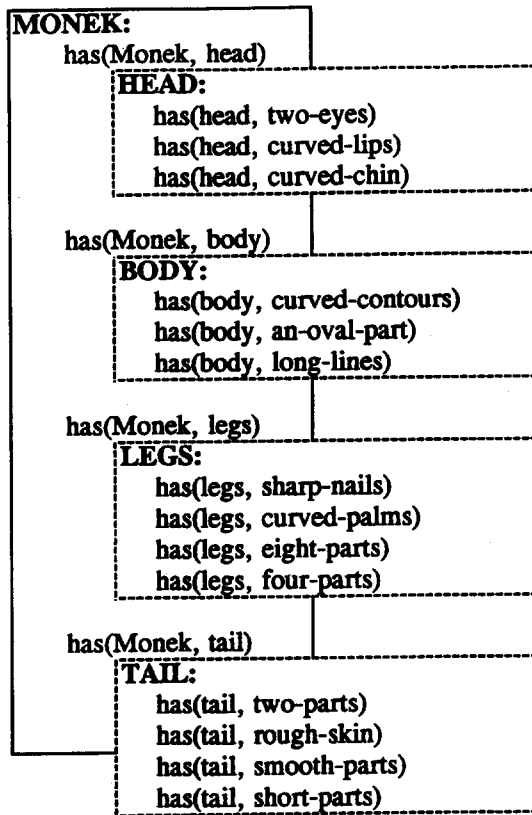
We suggest that structural alignment plays a critical role in feature extraction as well as inference using categories. In this section, we will discuss this idea in detail by describing the role of structural alignment in inference and classification. The structural alignment view of similarity was developed from Gentner's (1983, 1989) structure mapping theory of analogy. The theory has been successful in accounting for a variety of cognitive phenomena, including analogical reasoning, inductive inference, decision making, and similarity assessment (Gentner, 1989; Gentner & Markman, 1997; Lassaline, 1996; Markman & Gentner, 1993, 1997; Markman & Medin, 1995; Medin, Goldstone, & Gentner, 1993; Zhang & Markman, 1998).

Structural alignment is a mechanism by which the structured representation of a target concept is aligned with the structured representation of a base concept according to the relational match between them (Gentner, 1983, 1989; Markman & Gentner, 1997). For example, if we assume that a probe item and a category are represented by frames describing body parts (head, body, legs, tail) and other frames describing the characteristics of each part (Figure 3), inference and classification may be carried out by aligning the representation of a probe item with the representations of the known categories.

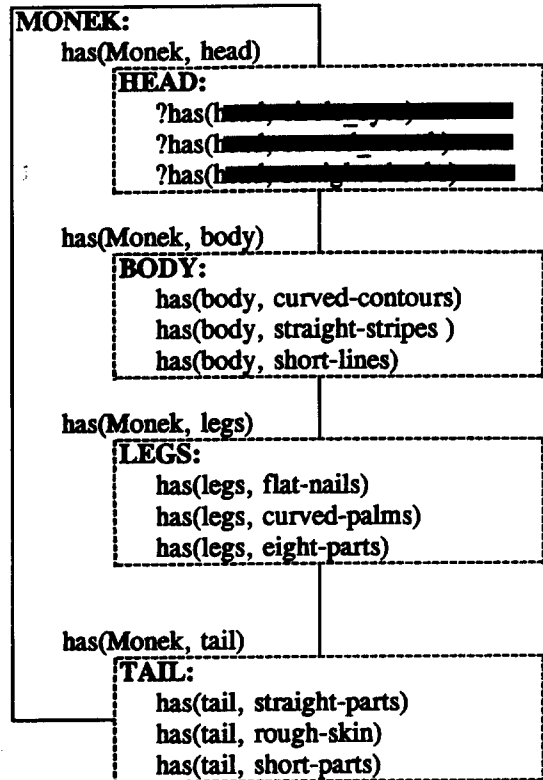
Several models of classification assume that similarity assessment is crucial for category processing (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986; Reed, 1972; Sloman, 1993). In this spirit, structural alignment is a means of assessing the similarity of pairs of concepts, but it differs from these models because the correspondences between concepts are determined by two basic constraints: *structural consistency* and *systematicity* (Clement & Gentner, 1991; Gentner, 1983, 1989; Markman, 1997; Markman & Gentner, 1993).

Structural consistency consists of two rules: *one-to-one mapping* and *parallel connectivity*. One-to-one mapping requires that each element of a target concept map to at most one element of a base concept. Parallel connectivity requires that if a correspondence is found between predicates in two concepts, then the arguments of those predicates must be also placed in correspondence. For example, given a representation of a probe item, if subjects align the probe item with the representation of a category (e.g., Figure 3a), all the corresponding predicates and arguments of the base domain will be placed in correspondence. Systematicity further requires matches based on systems of relations to be favored over

(a) A category representation



(b) Inference



(c) Classification

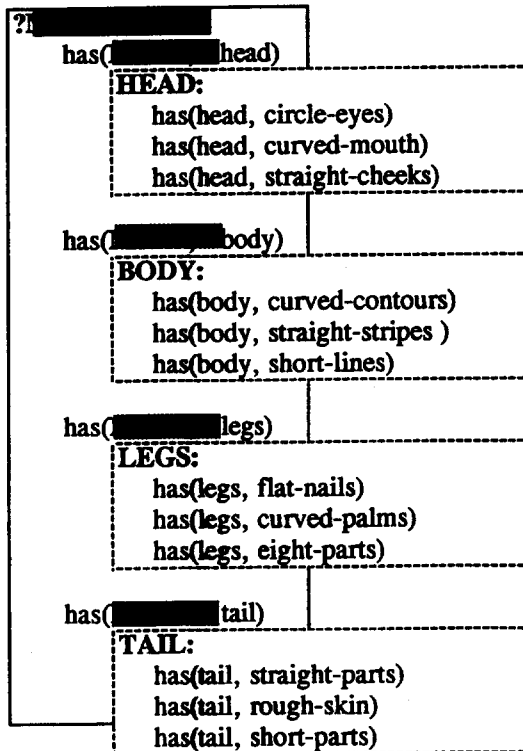


Figure 3. (a) A structured representation of Category Monek. (b) The structured representation of a probe item presented on an inference trial. (c) The structured representation of a probe item presented on a classification trial.

matches based on isolated local elements. For example, the match between two heads that have slightly different characteristics — HEAD:have (head, *straight parts*) and HEAD:have (tail, *straight lines*) — is preferred to the match between a head and a tail that have the same characteristic — HEAD:have (head, *straight parts*) and TAIL:have (tail, *straight parts*) — because the former match has a more systematic correspondence than the latter match.

These two rules are important in our discussion because they highlight the way commonalities of category members are extracted. In particular, we suggest that these two constraints are more likely to influence inference than classification. As a result, categories composed of multiple instances will be learned more easily in inference learning than in classification learning. On inference trials in our setting, a category label is shown in a stimulus frame, so that subjects are able to use that category as the base domain. In contrast, on classification trials, category labels are not present; so subjects need to compare the probe item to the representations of both categories. In this case, it is difficult to construct a single mapping that is structurally consistent. Using structural alignment in this case would require forming more than one mapping, which is extremely effortful.

In the multiple-feature-manifestation condition, the exemplars of a category are different in appearance, but they share a particular aspect of a feature value (e.g., have four legs). If the representation of a category reflects the collection of exemplars shown during the feedback, characteristics that exemplars have in common will be accumulated in the representation of a category more quickly than characteristics that differ across exemplars (see Hintzman, 1986). In inference learning, the value of the unknown feature of the probe is determined using the characteristics that the exemplars of the base category have in common. This proposal is consistent with the use of structural alignment to make candidate inferences, which is done by carrying features of the base category over to fill in the missing feature of the current exemplar (Clement & Gentner, 1991). Lassaline (1996) provides evidence that this process is at work in category-based induction. Repeated comparisons will allow the common aspects of the feature manifestations to be detected. The common elements are those aspects of the feature instances that are the same across stimuli.

In contrast, because classification does not permit the identification of a single base domain, subjects in classification learning will have difficulty using structural alignment. As a result, they will tend to employ other means to learn the categories (Markman, 1997). Many other strategies that have been able to account for classification in other laboratory settings will not work well when the features vary in their manifestations. For example, if subjects try to compare the new probe to specific exemplars seen before, it is unlikely that they will find exact

matches between the surface forms of a feature of a probe item and of that feature of other category members (Medin & Schaffer, 1978). If they try to find rules and exceptions that divide two categories, then they must first discover rules that involve the right aspects of the feature (Nosofsky, Palmeri, & McKinley, 1994). These strategies may be useful in the single-feature-manifestation condition when the number of potential feature matches is limited. These strategies will quickly become unwieldy in the multiple-feature-manifestation condition, because when there are many possible instantiations of each feature, there are a large number of potential feature matches and mismatches that will arise from each comparison.

The hypothesis that structural alignment is involved in inductive judgment and category acquisition is consistent with several other studies (Markman & Medin, 1995; Zhang & Markman, 1998). In one set of experiments, Zhang and Markman (1998) had subjects learn about a sequence of brands of new consumer products from a single product category (e.g., microwave popcorn). One brand was learned a week before two other brands were presented. In this case, brands presented late were assumed to be compared back to the first brand using structural alignment. Consistent with this prediction, subjects were better able to recall alignable differences (i.e., corresponding properties of the brands that differ) between the first brand and later brands than to recall the non-alignable differences (i.e., noncorresponding properties). A core prediction of structural alignment is that alignable differences will be favored over nonalignable differences in cognitive processing.

As discussed above, inference is more likely to involve structural alignment than classification, because only inference permits alignment between the new exemplar and the stored representation for a single category. Classification requires comparing the new exemplar back to more than one category. For this reason, subjects in the classification condition should use other strategies for learning the categories, such as learning rules and storing exceptions separately or attempting to memorize individual exemplars. Thus, subjects given classification learning should have substantial difficulty learning categories with multiple feature instances. In contrast, subjects given inference learning (or mixed learning where they perform inference learning on some trials) should not find the multiple-feature-instance condition to be very difficult. In the experiments reported below, we tested these predictions. In Experiment 1A, subjects first learned the two categories with one of three learning procedures—inference learning, classification learning, or mixed learning—with categories that had fixed feature instances. In Experiment 1B, the same three category learning procedures were tested with categories that had features with multiple instances. In Experiments 2 and 3, we examined the impact of multiple feature manifestations by manipulating the degree to which

structural alignment could be used in the three learning procedures.

EXPERIMENTS 1A AND 1B

Method

Subjects. Seventy-four subjects participated in Experiment 1A and 75 subjects participated in Experiment 1B. Five subjects were removed from the data analysis because they failed to complete the experiment. In total, 137 subjects were recruited from the Columbia University community and were paid \$6 for their participation; 12 subjects participated for course credit. In all, the data from 72 subjects in each experiment (24 per condition) were analyzed. Experiments 1A and 1B were conducted separately, and, hence, they were analyzed separately.

Materials. The stimuli used for Experiment 1A were schematic drawings of bugs with four binary feature dimensions: head (round, angular), body (striped, dotted), legs (eight legs, four legs), and tail (long, short). These stimuli, shown in Figure 1a, were like those used by Medin et al. (1987).

The structure of the two categories is illustrated in Table 1 (see Medin et al., 1987). Each category was derived from a prototype (A0 or B0) by arbitrarily assigning the value 0 or 1 in the stimulus design (Table 1). For example, by assigning 0 or 1 to each feature value, a stimulus consisting of a round head, a striped body, eight legs, and a long tail was described as (1, 1, 1, 0), whereas a stimulus containing an angular head, a dotted body, four legs, and a short tail was (0, 0, 0, 1). Different subjects saw different combinations.

The stimulus materials used for Experiment 1B were identical to those used in Experiment 1A except for one modification: In Experiment 1B, each presentation of a stimulus was depicted by a combination of individual feature instances, which were randomly taken from each feature dimension (Figure 1b). Because the computer program randomly combined the feature instances on each trial, the subjects almost never saw an exact exemplar twice even though the category structure remained constant.

Procedure. The basic procedure of both experiments involved three phases: an initial learning phase, a filler phase, and a final transfer phase.

In the initial learning phase, the subjects were arbitrarily assigned to one of three experimental conditions: classification, inference, or mixed. In the classification condition, the 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 (Figure 2a). Following each response, the correct stimulus and feedback were provided on the screen for 3 sec after their response. The subjects in inference learning made inferences of one of four features, and the category label of the exemplar and the remaining three features were depicted in the stimulus frame. For example, in Figure 2b, the subjects were given a stimulus frame containing two bugs. The bugs had three identical features (i.e., legs, bodies, and tails), but the value of one feature was different (i.e., round or angular head). The subjects were then asked to select the value of the feature given the category label of the stimulus. On other trials, other dimensions (body, legs, and tail) were inferred. The subjects responded by clicking one of two labeled buttons with the mouse. The left-right position of the correct response was determined randomly. Following each response, the correct stimulus and feedback were shown on the screen for 3 sec. In the mixed condition, half of the blocks were classification, and half were inference. The order of the blocks was determined randomly for each subject. For each condition, the subjects continued in the learning phase until they performed three consecutive blocks with a combined accuracy of greater than 90% or until they completed 30 blocks. Each stimulus appeared once

in each block. The order of stimulus presentation was determined randomly.

In classification learning, the subjects classified the four exemplars of each category but not the prototypes. Similarly, in inference learning, the subjects inferred all the feature values of the stimuli except for the "exception features." The exception features, which appear in italics in Table 1, were the feature values of a category that were consistent with the prototype of the other category. For example, the values of all the features in Set A were 1, except the values of the exception features which were 0. We did not include exception-feature inferences in the learning phase to keep inference learning and classification learning as equivalent as possible. For example, on each classification question, the subjects predicted the category label of an exemplar given the values of all four feature dimensions (e.g., Stimulus A1 in Table 1). This question has a schematic structure (1, 1, 1, 0, ?) = (head, body, legs, tail, category label) in Exemplar A1, assuming that the category label was just another feature (see Anderson, 1990). Analogously, on each inference question (e.g., a question about the head of Stimulus A1), the subjects predicted the value of an unknown feature (e.g., the value of head), and information about the other three features and the category label was given (e.g., the values of body, legs, tail, and the category label). This question had a schematic structure (? , 1, 1, 0, 1) = (head, body, legs, tail, category label) and was formally equivalent to the classification question [e.g., (1, 1, 1, 0, ?)], provided that the prediction of category labels and category features were in principle compatible. In contrast, the exception-feature trial of Stimulus A1 had a structure (1, 1, 1, ? , 1), which was analogous to the classification of a prototype (1, 1, 1, 1, ?). Because prototype stimuli were not presented in classification learning, it was necessary to exclude exception-feature inferences from inference learning to keep the two learning conditions equivalent.

The rest of the experiment was identical for all subjects. Following the learning phase, the subjects first participated in a filler phase in which they judged the pronounceability of nonsense words. This phase lasted about 10 min. After the filler phase, the subjects carried out classification and inference transfer tasks. The classification transfer task consisted of the classification of old stimuli and of prototypes as well as new prototypes, which were equivalent to old prototypes in their structure but had new manifestations of the features. In the inference transfer task, the subjects made feature inferences for old exemplars, old exception features, new exemplars, and new exception features. "Old" inference trials consisted of feature values that appeared during the learning phase. "New" inference trials were made of feature values that did not appear during the learning phase.

On exception-feature inference trials, an exemplar stimulus and the category prototype were contrasted so that the subjects chose the feature value consistent with either an exemplar seen during learning or the category prototype that was not seen during learning. For example, for the exception-feature inference of Stimulus A1, the subjects saw a figure having a head, body, and legs with a value of 1 and had to infer the feature value of the tail—either 1 or 0 [i.e., (1, 1, 1, ?, 1) = (head, body, legs, tail, category label)]. If the subjects chose the value 1, then it would be consistent with the prototype of Category Monek (1, 1, 1, 1). If the subjects chose the value 0, it would be consistent with Exemplar A1 (1, 1, 1, 0). Thus, the response to exception-feature inferences could provide insight into the degree to which the subjects attend to category instances.

The instructions specifically asked the subjects to make their decisions based on the categories learned during the initial phase. All subjects were given the transfer tasks in the same order. No feedback was given during the transfer task. The entire experiment took 30–40 min.

Design. There were three between-subjects learning conditions: inference, classification, and mixed. The main dependent measures

for learning were the number of subjects who reached the learning criterion and the number of learning blocks required to reach the criterion. The dependent measures in the transfer tasks were the proportion of correct classifications of old exemplars, old prototypes, and new prototypes, the proportion of correct inferences to old exemplars and new exemplars, and the proportion of inferences to exception features produced by old features and new features.

Results

Experiment 1A. For the analyses of learning speed and classification transfer of old exemplars, one-way analyses of variance (ANOVAs) were used. The remaining analyses involved mixed 3×2 ANOVAs with one between-subjects factor (learning condition: inference, mixed, and classification) and one within-subjects factor (stimulus type: new vs. old). For all analyses, we included the data only from those subjects who reached the 90% accuracy criterion before the 30 block maximum.¹ All 24 subjects given inference learning reached the criterion, as did 21 subjects given mixed learning and 17 subjects given classification learning. The three conditions differed significantly in the speed with which the criterion was reached [$F(2,59) = 11.29$, $MS_e = 31.11$, $p < .01$]. Planned comparisons revealed that the subjects required significantly fewer blocks in inference learning ($M = 4.8$) than did the subjects in classification learning ($M = 13.2$) [$t(39) = 5.10$, $p < .01$] and than did the subjects in mixed learning ($M = 8.4$) [$t(43) = 2.61$, $p < .05$].² The difference between classification learning and mixed learning was marginally significant [$t(36) = 2.26$, $p < .09$]. A similar pattern of results was found in previous studies using different stimulus materials (Yamauchi & Markman, 1998).

Table 2
Proportions of Correct Responses in Experiment 1A

	Learning Conditions			Average
	Inference	Mixed	Classification	
Classification Transfer				
Old Exemplars	.86	.88	.90	.88
Prototypes				
Old	1.00	.88	.97	.95
New	.81	.86	.65	.77
Average	.91	.87	.81	
Inference Transfer				
Exemplars				
Old	.98	.92	.85	.92
New	.92	.87	.73	.84
Average	.95	.90	.79	
Exception Features				
Old	.99	.86	.78	.88
New	.92	.83	.72	.82
Average	.96	.85	.75	

Note—Old exemplars (prototypes) stand for exemplars (prototypes) composed of features that were shown during the learning phase. New exemplars (prototypes) stand for exemplars (prototypes) composed of features that were not shown during the learning phase. The numbers shown in the rows of exception features represent the feature inferences made to prototype stimuli. These numbers are the proportions for subjects' selections of the prototype values of each category (e.g., Feature Value 1 in Monek and Feature Value 0 in Plaple).

The data from the transfer phase are summarized in Table 2. For classification transfer, the subjects in the three learning conditions performed equally accurately in the classification of old exemplars [$F(2,59) < 1.0$, $MS_e = 0.02$, $p > .10$] and old prototypes [$F(2,59) = 1.02$, $MS_e = 0.09$, $p > .10$]. As predicted, subjects generally classified old prototypes more accurately than new prototypes [$F(1,59) = 18.63$, $MS_e = 0.05$, $p < .01$]. This effect was more pronounced in classification learning and in inference learning than in mixed learning, as reflected in a significant interaction between learning condition and stimulus type [$F(2,59) = 4.13$, $MS_e = 0.05$, $p < .05$]. The subjects in classification learning performed more accurately for the classification transfer of the old prototypes ($M = 0.97$) than of the new prototypes ($M = 0.65$) [$t(16) = 3.0$, $p < .05$]; similarly, the subjects in inference learning were more accurate for the classification transfer of the old prototypes ($M = 1.0$) than of new prototypes ($M = 0.81$) [$t(23) = 3.2$, $p < .05$]. The subjects in mixed learning were equally accurate in the classification transfer of old prototypes ($M = 0.88$) and new prototypes ($M = 0.86$) [$t(20) = 0.44$, $p > .10$].

For the inference transfer of exemplars, the subjects in the three conditions differed significantly in their ability to make inferences of exemplars (inference learning, $M = 0.95$; mixed learning, $M = 0.90$; classification learning, $M = 0.79$) [$F(2,59) = 7.64$, $MS_e = 0.03$, $p < .01$]. The subjects in inference learning performed better than did the subjects in classification learning [$t(39) = 5.66$, $p < .01$]. As in classification transfer, the subjects were generally more accurate in the inference transfer of old exemplars ($M = 0.92$) than in the transfer of new exemplars ($M = 0.84$) [$F(1,59) = 19.30$, $MS_e = 0.01$, $p < .01$]. There was no interaction between learning conditions and stimulus types [$F(2,59) = 1.45$, $MS_e = 0.01$, $p > .10$].

For the exception-feature inferences, we measured the proportion of subjects who responded with category prototypes (i.e., prototype-accordance responses, such as the Feature Value 1 in the case of Monek and 0 in the case of Plaple). For all subjects, inferences of the exception features were made in accordance with the category prototypes more often than with exemplars, though the three learning conditions differed in the levels of prototype-accordance inferences (inference learning, $M = 0.96$; mixed learning, $M = 0.85$; classification learning, $M = 0.75$) [$F(2,59) = 6.97$, $MS_e = 0.06$, $p < .01$]. The difference between inference learning and classification learning was significant [$t(39) = 5.96$, $p < .01$]. Overall, prototype-accordance responses were preferred in old stimuli ($M = 0.88$) more often than in new stimuli ($M = 0.82$) [$F(1,59) = 5.64$, $MS_e = 0.02$, $p < .05$]. But the interaction between learning condition and stimulus type did not reach significance [$F(2,59) = 1.45$, $MS_e = 0.01$, $p > .10$].

Experiment 1B. In this experiment, we expected learning to be quite difficult for the subjects in classification learning because, as we hypothesized earlier,

The Number of Blocks in Learning

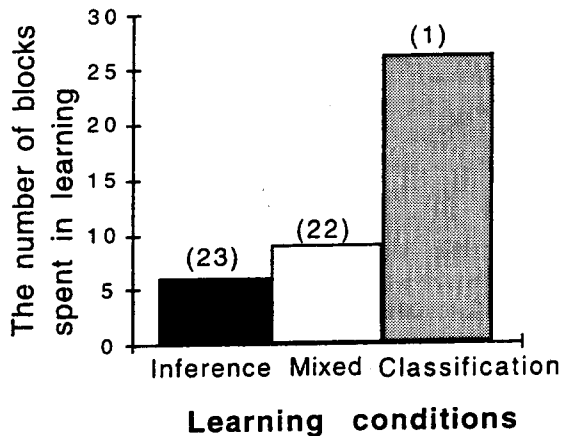


Figure 4. The average number of learning blocks spent by the subjects who reached the learning criterion in Experiment 1B. The numbers enclosed by the parentheses represent the number of subjects who reached the learning criterion.

structural alignment is involved in inference but not in classification. Consistent with this prediction, 23 subjects in inference learning, 22 subjects in mixed learning, and only 1 subject in classification learning reached the learning criterion. The average number of blocks spent by the subjects who reached the learning criterion was 6.1 in inference learning, 9.1 in mixed learning, and 26.0 in classification learning (Figure 4). The difference between inference learning and mixed learning was marginally significant [$t(46) = 1.84, p < .08$]. Because only 1 subject in classification learning reached the criterion, this condition was excluded from the remaining analyses. For these analyses, we examined the data of only the subjects who reached the 90% accuracy criterion before the 30 block maximum. In the transfer phase, the subjects in the two learning conditions (inference learning and mixed learning) performed accurately for both classification transfer and inference transfer. On average, the subjects in the two conditions performed above 85% accuracy for every dependent measure.

For the classification transfer of old stimuli, the subjects in mixed learning ($M = 0.92$) were more accurate than were the subjects in inference learning ($M = 0.86$). This difference was marginally significant [$t(43) = 1.73, .05 < p < .10$]. For the classification transfer of prototype stimuli, neither the main effect nor the interaction was significant. The subjects in the two learning conditions were equally accurate both in old prototypes ($M = 0.95$) and in new prototypes ($M = 0.95$).

For the inference transfer of exemplar stimuli, the subjects in inference learning ($M = 0.97$) were more accurate than the subjects in mixed learning ($M = 0.93$). This difference was marginally significant [$t(44) = 1.81, .05 < p < .10$]. There was no main effect of stimulus type (old vs. new) nor was there an interaction between

stimulus type and learning condition [stimulus type, $F(1,43) = 1.16, p > .10$; stimulus type \times learning condition, $F(1,43) = 2.34, MS_e < 0.01, p > .10$]. For the inference transfer of exception feature stimuli, the subjects in inference learning responded with prototype features ($M = 0.98$) more often than did the subjects in mixed learning ($M = 0.93$) [$t(43) = 1.69, .05 < p < .10$]. This tendency was pronounced in old exception features ($M = 0.96$) more than in new exception features ($M = 0.95$) [$t(44) = 1.73, .05 < p < .10$].

Discussion

The main results of Experiments 1A and 1B can be summarized as follows. First, in Experiment 1A, consistent with previous studies, inference learning was generally easier than classification learning (Yamauchi & Markman, 1998). Second, the three learning procedures were sensitive to the change in exemplar information. The subjects' performance in the three learning conditions was generally better for old stimuli than for new stimuli in Experiment 1A, though this tendency was greater in classification learning and inference learning than in mixed learning. Third, the subjects performed better when the learning task and the transfer task matched. The subjects given classification learning were accurate in classification transfer, and the subjects given inference learning were accurate in inference transfer. The subjects given mixed learning were good in both inference transfer and classification transfer. The high level of performance of the subjects given mixed learning suggests that the combination of the two learning tasks produces a category representation robust enough to allow people to process exemplars that have feature values not seen during learning.

In Experiment 1B, the distinction between inference learning and classification learning is striking. As predicted, the impact of multiple feature manifestations was apparent in classification learning. Only 1 subject given classification learning reached the learning criterion within 30 blocks (240 trials) in the multiple-feature-manifestation condition in Experiment 1B, whereas 17 subjects were able to learn the same categories with an average of 13.2 blocks (105.6 trials) in the single-feature-manifestation condition in Experiment 1A.

The multiple-feature-manifestation condition did not impede inference learning and mixed learning as much as it did classification learning, though this manipulation clearly affected the subjects' performance in the two conditions. The interpretation of the difference between inference learning and classification learning is, however, difficult because the two conditions differed in the way stimuli were presented. The subjects in inference learning were given two stimuli together on each trial, whereas the subjects in classification learning were presented with a single stimulus on each trial (Figure 2). Thus, inference learning might have been easier than classification learning, because the presentation of stimulus pairs allowed the subjects in inference learning to

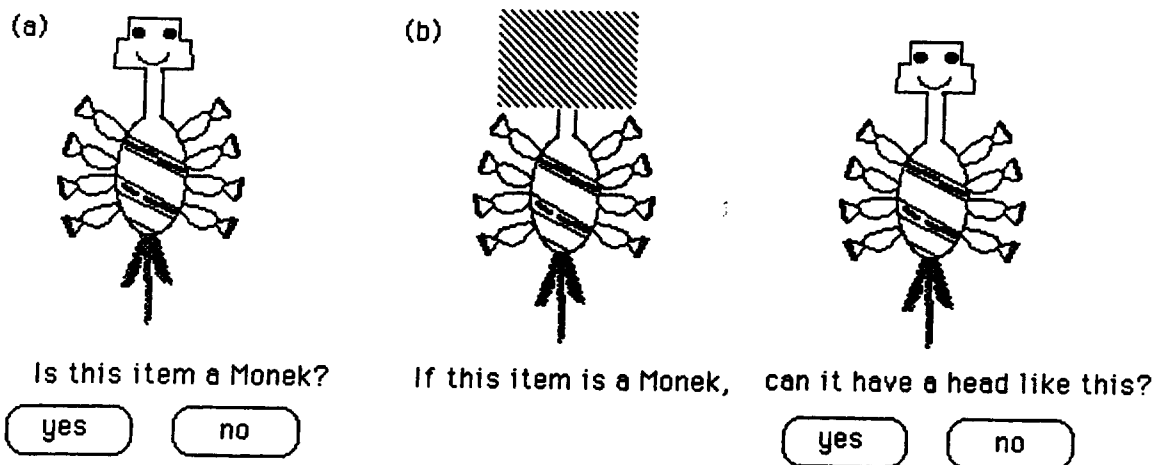


Figure 5. Sample stimuli used for (a) a classification question in Experiment 2 and (b) an inference question in Experiment 2.

compare pairs of stimuli to each other rather than comparing the new exemplar back to a stored category representation, as we hypothesized.

In Experiment 2, we tested this possibility by changing the inference trials so that the subjects in inference learning saw only a single stimulus on each trial and received the same feedback as did the subjects in classification learning (Figure 5b). If the subjects given inference learning and mixed learning in Experiment 1 were only comparing the objects in the stimulus pairs together, these conditions should be just as difficult as the classification condition. In contrast, if the difference in performance between the subjects given inference learning and those given classification learning arose from the comparisons between the new exemplar and stored category representations in the process of inference, then we should again find that inference learning and mixed learning have an advantage over classification learning in the multiple-feature-manifestation condition.

EXPERIMENT 2

Method

Subjects. The subjects were 76 members of the Columbia University community. One subject failed to complete the experiment, and the data from 3 subjects were lost due to an experimenter error. In all, the data from 72 subjects (24 per condition) were analyzed. Among them, 69 were paid \$6 for their participation and 3 were given course credit.

Design. There were three between-subjects learning conditions: inference, classification, and mixed. All subjects were tested with stimuli having multiple feature manifestations. The dependent variables used in this experiment were identical to those used in Experiment 1B except that new exemplars (exemplars that were composed of the feature instances that did not appear during learning) were included on classification transfer trials. These new exemplars were analogous in their structure to old stimuli shown during learning (Table 1).

Material. The materials used for this experiment were identical to those used in the multiple-feature-manifestation condition of Experiment 1B.

Procedure. The procedure in this experiment was identical to that employed in Experiment 1B, except for one key modification. In this experiment, classification trials involved a binary yes/no question. On one trial, for example, the computer screen showed a single stimulus and asked, "Is this item a Monek?" The subjects responded by clicking a button labeled either "yes" or "no." On another trial, the screen showed a single stimulus and asked, "Is this item a Plaple?" (Figure 5a).

The stimuli used in inference learning were identical to those shown in classification learning. A single stimulus appeared on the screen whose target feature was covered with a mask (Figure 5b). The phrase "If this item is a Monek/Plaple," accompanied the stimulus. When the subjects pressed a key, the mask disappeared from the target feature, and the rest of the question appeared ("can it have a head/a body/a tail/legs like this?"). The stimulus presented at this stage was identical to the one shown in classification learning. The subjects then responded by clicking a button labeled either "yes" or "no." The subjects in inference learning and the subjects in classification learning received the same feedback, in which the correct figure was shown on the screen for 3 sec.

The stimuli that required positive responses were exemplars of the two categories that are shown in Table 1. The stimuli that required negative responses were neutral stimuli having a structure such that they had two features consistent with the prototype of one category and two features consistent with the prototype of the other category [e.g., (1, 1, 0, 0, ?) for classification questions and (? , 0, 0, 1, 1) for inference questions]. For example, to produce a negative stimulus for Stimulus Type A1 (1, 1, 1, 0, ?), one feature was randomly selected among those features that had Value 1 (head, body, and legs), and the value of that dimension was changed to Value 0.

In the transfer phase, the subjects classified 8 old exemplars, 8 new exemplars, 2 old prototypes, and 2 new prototypes. Following classification transfer, the subjects inferred the feature values of 8 old exemplars, 8 new exemplars, 2 old prototypes, and 2 new prototypes. The subjects predicted the values of all the four features of every stimulus. The order of classification transfer trials and inference transfer trials was determined randomly. The rest of the experiment was identical to Experiment 1B.

Results and Discussion

In Experiment 2, the subjects in the three learning conditions were presented with a single stimulus on each trial and responded to binary yes/no questions. We expected that inference learning and mixed learning would be still easier than classification learning because the subjects could compare the current exemplar back to a single base category when making inferences. The results of the experiment were consistent with this prediction. In all, 17 subjects in inference learning, 9 subjects in mixed learning, and 3 subjects in classification learning reached the learning criterion. The average number of blocks spent by the subjects who reached the criterion was 8.1 in inference learning, 14.3 in mixed learning, and 15.7 in classification learning.³ The results clearly indicate that the stimulus presentation in Experiment 1 was not the factor that made classification learning difficult in the multiple-feature-manifestation condition. The results support the idea that structural alignment between exemplars and stored category representations is instrumental in learning categories consisting of varying instances and that this process is easier to carry out in inference than in classification.

Because only 3 subjects reached the learning criterion in classification learning, the results from the transfer phase, which are summarized in Table 3, will not be discussed further.

Table 3
Proportions of Correct Responses in Experiment 2

	Learning Conditions		Average
	Inference	Mixed	
Classification Transfer			
Exemplars			
Old	.66	.79	.73
New	.64	.78	.71
Average	.65	.79	
Prototypes			
Old	.74	.56	.65
New	.68	.50	.59
Average	.71	.53	
Inference Transfer			
Exemplars			
Old	.94	.87	.91
New	.88	.89	.89
Average	.91	.88	
Exception Features			
Old	.90	.83	.87
New	.85	.78	.82
Average	.88	.81	

Note—The data from classification learning were not included in this table because only 3 subjects in classification learning reached the learning criterion. Old exemplars (prototypes) stand for exemplars (prototypes) composed of features that were shown during the learning phase. New exemplars (prototypes) stand for exemplars (prototypes) composed of features that were not shown during the learning phase. The numbers shown in the rows of exception features represent the feature inferences made to prototype stimuli. These numbers are the proportions for subjects' selections of the prototype values of each category (e.g., Feature Value 1 in Monek and Feature Value 0 in Plaple).

In Experiments 1 and 2, we found that inference learning and mixed learning were easier than classification learning in the multiple-feature-manifestation condition. On this basis, we argued that structural alignment is pivotal for extracting the common core across multiple instances of features. Alternatively, the advantage of the two learning conditions might have arisen from the fact that an inference trial directs subjects to attend to a specific feature. For example, on an inference trial, the subjects were asked if the bug has a particular feature value in the stimulus frame (e.g., "If this item is a Monek, can it have a head like this?"). For this reason, the advantage of inference learning and mixed learning over classification learning might have derived solely from the attention given to specific features rather than the nature of comparison associated with the two tasks.

To test this hypothesis in Experiment 3, the subjects in the three learning conditions were given two stimuli together on each trial as in inference learning of Experiment 1B and participated in classification learning, inference learning, or mixed learning (Figure 6). Previous research suggests that presenting pairs of objects facilitates structural comparisons of the objects, even in tasks that do not appear to require these comparisons (e.g., Markman & Medin, 1995). Assuming this kind of stimulus presentation promotes alignment, classification learning should be as easy as inference learning and mixed learning in this experiment. In contrast, if the advantage of inference learning and mixed learning over classification learning was due to the focus on a specific feature made on an inference trial, then the subjects in classification learning will still have trouble learning categories in the multiple-feature-manifestation condition.

EXPERIMENT 3

Method

Subjects. The subjects were 150 members of the Columbia University community who were paid \$6 for participation. The data from 6 subjects were removed from the analyses because they failed to complete the experiment. In total, the data from 144 subjects (24 per condition) were analyzed.

Materials. The materials used in this experiment were stimulus pairs like those shown in Figure 6. One of the two figures of a pair was associated with a positive response, and the other was associated with a negative response. The two types of figures, which were shown on the screen side by side, correspond to positive and negative stimuli shown in Experiment 2. They differed only in their values on one feature dimension.

As in Experiment 1, the stimuli shown in the single-feature-manifestation condition had a fixed manifestation. The stimuli shown in the multiple-feature-manifestation condition were produced by combinations of one of four feature instances obtained from the four feature dimensions. All stimuli complied with the stimulus structure depicted in Table 1.

Procedure. The procedure for inference learning was the same as the one used for inference learning in Experiment 1 except that the arrow indicating the target feature was removed from stimulus presentation (Figure 6). This modification was introduced to make inference learning more comparable to the classification learning task used here. As in inference learning of Experiment 1, the sub-

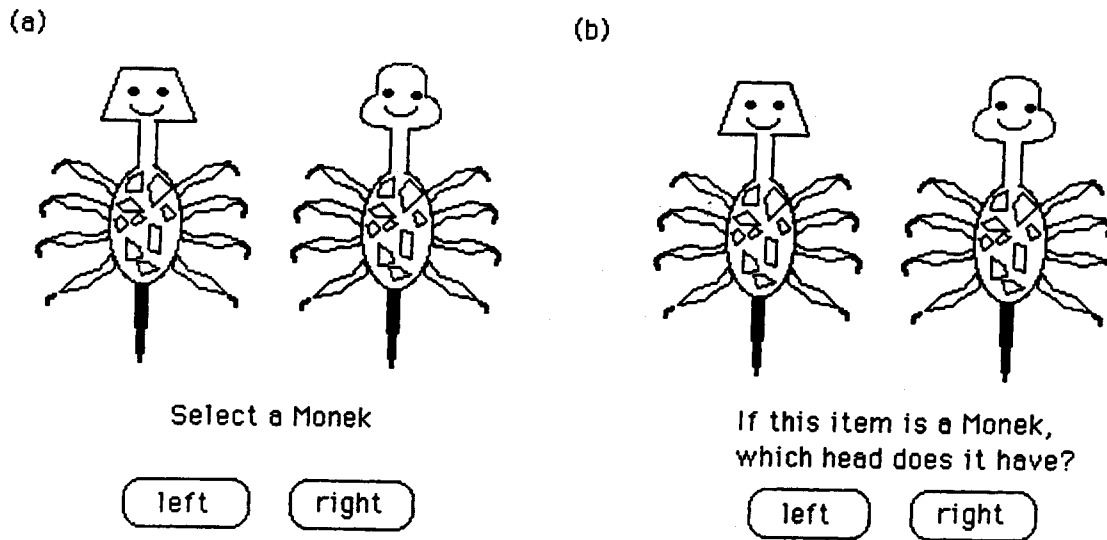


Figure 6. Sample stimuli used for (a) a classification question in Experiment 3 and (b) an inference question in Experiment 3.

jects were presented with a pair of figures and were asked to choose the relevant feature value given the category label. In classification learning, the subjects were presented with the same stimulus pairs as in inference learning and were asked to “select a Monek” on some trials or to “select a Plaple” on other trials. The subjects in inference learning and the subjects in classification learning received the same feedback, in which the correct figure was shown on the screen for 3 sec. The procedure for the mixed condition was a mixture of inference learning trials and classification learning trials.

As in Experiment 2, the transfer phase consisted of classification transfer and inference transfer. The subjects classified 8 old exemplars, 8 new exemplars, 2 old prototypes, and 2 new prototypes. Following classification transfer, the subjects inferred the 24

feature values of 8 old exemplars, the 24 feature values of 8 new exemplars, the 8 feature values of 2 old exception-feature stimuli, and the 8 feature values of 2 new exception-feature stimuli. The order of classification transfer trials and inference transfer trials was determined randomly. The rest of the experiment was identical to Experiment 2.

Design. The experimental design was a 3 (learning type: inference, classification, and mixed) × 2 (stimulus presentation: single manifestation and multiple manifestations) × 2 (stimulus type: new and old) mixed factorial. The first two factors—learning type and stimulus presentation—were between-subjects factors, and the last factor—stimulus type—was a within-subjects factor. The subjects were assigned randomly to one of six conditions.

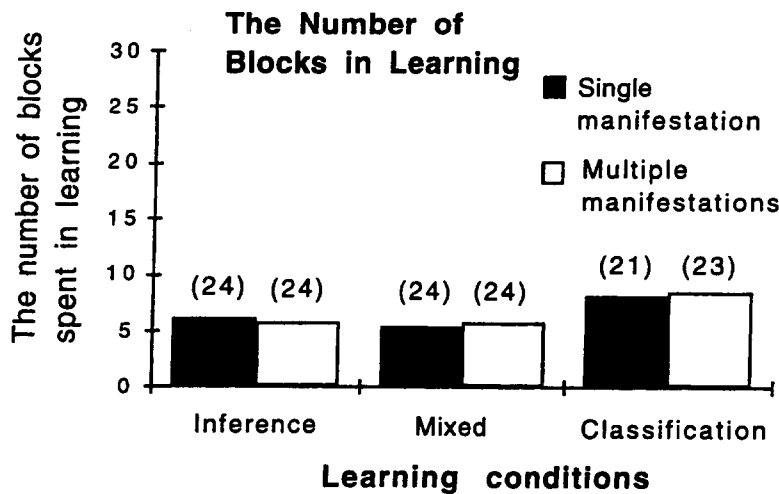


Figure 7. The average number of learning blocks spent by the subjects who reached the learning criterion in Experiment 3. The numbers enclosed by the parentheses represent the number of learning subjects who reached the learning criterion.

Table 4
**Proportions of Correct Responses in the Single-Feature-
 Manifestation Condition of Experiment 3**

	Learning Conditions			Average
	Inference	Mixed	Classification	
Classification Transfer				
Exemplars				
Old	.92	.96	.96	.95
New	.81	.89	.89	.86
Average	.87	.93	.93	
Prototypes				
Old	.92	.96	.95	.94
New	.83	.88	.86	.86
Average	.88	.92	.91	
Inference Transfer				
Exemplars				
Old	.93	.95	.95	.94
New	.88	.91	.91	.90
Average	.91	.93	.93	
Exception Features				
Old	.95	.95	.93	.94
New	.88	.91	.92	.90
Average	.92	.93	.93	

Note—Old exemplars (prototypes) stand for exemplars (prototypes) composed of features that were shown during the learning phase. New exemplars (prototypes) stand for exemplars (prototypes) composed of features that were not shown during the learning phase. The numbers shown in the rows of exception features represent the feature inferences made to prototype stimuli. These numbers are the proportions for subjects' selections of the prototype values of each category (e.g., Feature Value 1 in Monek and Feature Value 0 in Plaple).

Results

For the analysis of learning speed, a 2 (stimulus presentation) × 3 (learning condition) ANOVA was used. For the analyses of transfer performance, 2 (stimulus presentation) × 3 (learning condition) × 2 (stimulus type) ANOVAs were employed. In Experiment 3, we predicted that the subjects in classification learning should be able to learn categories in the multiple-feature-manifestation condition as well as in the single-feature-manifestation condition.

Consistent with this prediction, classification learning was much easier in this experiment than it was in Experiment 2. In the single-feature-manifestation condition, all 24 subjects in inference learning, all 24 subjects in mixed learning, and 21 subjects in classification learning reached the learning criterion. Similarly, in the multiple-feature-manifestation condition, all 24 subjects in inference learning, all 24 subjects in mixed learning, and 23 subjects in classification learning reached the criterion (Figure 7). An ANOVA indicated that there was a main effect of learning condition [$F(2,134) = 5.47, MS_e = 18.18, p < .01$]. Among the three learning conditions, the subjects in classification learning spent significantly more learning blocks ($M = 8.3$) than did the subjects in inference learning ($M = 5.9$) and the subjects in mixed learning ($M = 5.6$) [$t(90) = 2.45, p < .05$, and $t(90) = 2.78, p < .05$, respectively]. The difference between inference learning

and mixed learning was not significant [$t(94) = 0.55, p > .10$]. There was no interaction between learning condition and stimulus presentation [$F(2,134) < 1.0$].

The main results of the transfer phase are summarized in Tables 4 and 5. The subjects in the three learning conditions were very accurate in all dependent measures. Both in classification transfer and in inference transfer, they performed above 80% on both the old stimuli and the new stimuli. There were neither main effects nor interactions involving learning condition on any of the two transfer measures. The results indicate that the effect of learning condition was virtually eliminated for all dependent measures in this experiment.

The subjects were more accurate for old stimuli than for new stimuli in the single-feature-manifestation condition but not in the multiple-feature-manifestation condition. For the classification transfer of old stimuli, there was a main effect of stimulus type (old stimuli, $M = 0.94$; new stimuli, $M = 0.89$) [$F(1,139) = 14.56, MS_e = 0.01, p < .01$]; yet an interaction between stimulus type and stimulus presentation [$F(1,134) = 4.55, MS_e = 0.01, p < .05$] indicated that the effect of stimulus type was enhanced in the single-feature-manifestation condition (old stimuli, $M = 0.94$; new stimuli, $M = 0.86$) [$t(68) = 4.0, p < .01$]; but not in the multiple-feature-manifestation condition (old stimuli, $M = 0.93$; new stimuli, $M = 0.91$) [$t(70) = 1.29, p > .10$]. A similar pattern of results was obtained for the classification transfer of proto-

Table 5
**Proportions of Correct Responses in the Multiple-Feature-
 Manifestation Condition of Experiment 3**

	Learning Conditions			Average
	Inference	Mixed	Classification	
Classification Transfer				
Exemplars				
Old	.94	.92	.94	.93
New	.93	.89	.91	.91
Average	.94	.91	.93	
Prototypes				
Old	.94	.92	.87	.91
New	.94	.90	.98	.94
Average	.94	.91	.93	
Inference Transfer				
Exemplars				
Old	.94	.91	.92	.92
New	.94	.90	.94	.93
Average	.94	.91	.93	
Exception Features				
Old	.94	.90	.93	.92
New	.92	.91	.96	.93
Average	.93	.91	.95	

Note—Old exemplars (prototypes) stand for exemplars (prototypes) composed of features that were shown during the learning phase. New exemplars (prototypes) stand for exemplars (prototypes) composed of features that were not shown during the learning phase. The numbers shown in the rows of exception features represent the feature inferences made to prototype stimuli. These numbers are the proportions for subjects' selections of the prototype values of each category (e.g., Feature Value 1 in Monek and Feature Value 0 in Plaple).

type stimuli. There was not the main effect of stimulus type (old stimuli, $M = 0.93$; new stimuli, $M = 0.90$) [$F(1,139) = 1.71$, $MS_e = 0.03$, $p > .10$], but there was an interaction between stimulus type and stimulus presentation [$F(1,134) = 6.91$, $MS_e = 0.03$, $p < .01$]. As in the classification transfer of exemplar stimuli, the subjects were more accurate on the old stimuli ($M = 0.94$) than on the new stimuli ($M = 0.86$) in the single-feature-manifestation condition [$t(68) = 2.67$, $.05 < p < .10$], but not in the multiple-feature-manifestation condition (old stimuli, $M = 0.91$; new stimuli, $M = 0.94$) [$t(70) = 0.94$, $p > .10$].

The results from the inference transfer tasks were equivalent to those from the classification transfer tasks. The subjects were more accurate for old stimuli ($M = 0.94$) than for new stimuli ($M = 0.91$) [$F(1,139) = 8.9$, $MS_e = 0.004$, $p < .01$]. This tendency was pronounced in the single-feature-manifestation condition but not in the multiple-feature-manifestation condition, leading to an interaction between stimulus type and stimulus presentation [$F(1,134) = 7.15$, $MS_e = 0.004$, $p < .01$] (see Tables 4 and 5). In the inference transfer of exception features, there was a main effect of stimulus type [$F(1,139) = 4.19$, $MS_e = 0.005$, $p < .05$] and an interaction between stimulus type and presentation [$F(1,134) = 8.0$, $MS_e = 0.005$, $p < .01$]. As in the other analyses, the subjects were more accurate on the old stimuli than on the new stimuli in the single-feature-manifestation condition [$t(68) = 3.14$, $p < .05$], but not in the multiple-feature-manifestation condition [$t(70) = 0.66$, $p > .10$] (see Tables 4 and 5).

Discussion

The results of Experiment 3 reveal that alignment is a strong force in learning categories composed of multiple feature instances. Unlike the previous experiments, the number of blocks spent in each learning procedure was almost identical in the single-feature-manifestation condition and in the multiple-feature-manifestation condition. There was a tendency for the subjects to require more blocks in classification than in inference, but this difference has been also observed in studies using different materials (Yamauchi & Markman, 1998). For example, using simple geometric figures having a single feature manifestation, Yamauchi and Markman showed that subjects who learned categories in classification learning required significantly more trials to reach a learning criterion than did subjects in inference learning when the categories were linearly separable.

In contrast to Experiments 1 and 2, there was no effect of learning procedure on the transfer tasks in Experiment 3. Indeed, transfer performance was excellent in all conditions. The average performance of the subjects in the three learning conditions exceeded 80% for every transfer measure in the single-feature-manifestation condition and 85% in the multiple-feature-manifestation condition. The subjects given classification learning performed accu-

rately in classification transfer as well as in inference transfer; the subjects given inference learning performed accurately in classification transfer as well as in inference transfer. The subjects' transfer performance was generally better for old stimuli than for new stimuli. This effect was present only in the single-feature-manifestation condition but not in the multiple-feature-manifestation condition. These results imply that flexible descriptions of features are likely to emerge in the process of comparing a variety of different instances. Taken together, these results are consistent with the view that comparison has a powerful influence on subjects' ability to learn categories and that structural alignment plays a critical role in category acquisition.

The purpose of Experiment 3 was to determine whether subjects given classification learning can learn categories with features that have multiple manifestations when they are permitted to make comparisons between stimuli. To this end, subjects in all conditions were given pairs of stimuli and were allowed to make comparisons of them during learning. This type of comparison differs from the type of comparison involved in inference in Experiment 2, in which the subjects could compare only the newly presented exemplar back to some stored category representation. We assume that structural alignment was employed in both cases, because other research indicates that structural alignment plays a central role in similarity judgment and decision making both when a stimulus is compared to a representation stored in memory (Zhang & Markman, 1998) and when two stimuli are compared in the same stimulus frame (Markman & Gentner, 1993; Markman & Medin, 1995).

In Experiment 3, the subjects in classification learning were likely to employ structural alignment for two reasons. First, the subjects in this setting were allowed to compare two stimuli on the same stimulus frame; so they were likely to align stimuli and to focus on commonalities of different feature manifestations. Second, because the subjects in the classification task were given the category label on each trial (e.g., "Select a Monek"), they could attend to a single base category, just like the subjects in the inference task. As a consequence, the subjects given classification learning were able to acquire the categories in the multiple-feature-manifestation condition.

GENERAL DISCUSSION

How does the category learning system discover commonalities among the members of a category when the appearance of the members differs widely? The experiments presented here suggest that comparison, particularly structural alignment, plays a key role. In classification learning, in which structural alignment is difficult, the subjects had substantial difficulty learning categories composed of multiple feature manifestations, although they were able to learn categories with the same structure when each feature had a single manifestation. The subjects in inference learning could learn categories both in

the single-feature-manifestation condition and in the multiple-feature-manifestation condition. Likewise, the subjects in classification learning were also able to learn categories both in the single-feature-manifestation condition and in the multiple-feature-manifestation condition when alignment between stimuli was encouraged by the stimulus presentation. The results of the experiments were consistent with the view that inference and classification differ in the way they are carried out and that structural alignment, which may be incorporated in inference more easily than in classification, is pivotal in learning categories composed of multiple feature instances.

How does structural alignment facilitate category learning? The category label given in a stimulus frame in inference learning may play an important role. The label provides a symbol that eases the alignment of the probe item with the stored category representation. Previous research suggests that alignment between stimuli facilitates attention to relational similarities between items (Markman, 1996; Markman & Gentner, 1993). These relational commonalities may be particularly important in the multiple-feature-manifestation condition, in which different stimuli share few surface similarities.

In a categorization setting, several studies have suggested that comparisons help people learn the set of features relevant for categorization. As an example, Schyns and Rodet (1997) asked subjects to learn a pair of categories defined by perceptual features. One group first learned a category defined by the presence of a perceptual feature X, and then they learned another category defined by the presence of a single perceptual element that was the conjunction of feature X and a second feature Y ($X \rightarrow XY$). In later transfer tasks, this group acted as if they had learned the features X and Y separately, presumably because comparing the conjunction of X and Y back to the known feature X allowed these subjects to segment the conjunction into parts. A second group learned the two categories in the reverse order ($XY \rightarrow X$). In a later transfer task, this group acted as if they learned the feature X and a second feature consisting of the conjunction of X and Y. For this group, comparing X to the learned conjunction XY did not lead to a segmentation of the initial feature. Notice, in this case, that subjects were learning one category at a time. The present results suggest that this aspect of the experimental design eases the comparison of new exemplars to stored category representations.

Structural alignment has been also shown to benefit memory for features that are aligned (Markman & Gentner, 1997; Zhang & Markman, 1998). In one study, subjects were asked to compare pairs of scenes (Markman & Gentner, 1997). After a 30-min delay, they were shown parts of the scenes that were either alignable differences or nonalignable differences of the scenes and were asked to remember as much as they could about the scenes. Subjects were able to recall significantly more about the pictures when the recall cues were objects that had been alignable differences of the original pair than when they had been nonalignable differences.

These properties of structural alignment are likely to help people to form a category that preserves commonalities across the feature manifestations of different exemplars. When the members of a category have a feature structure based on correlated attributes, structural alignment facilitates the acquisition of a category representation that reflects the overall resemblance of category members. More empirical studies are needed to determine exactly what aspect of alignment contributes to the extraction of feature commonalities; however, in the present experiments, structural alignment was likely to focus attention on the common elements across instances (as well as on the alignable differences related to those commonalities). This focus is helpful for finding the common aspects across different manifestations of the same feature value and eventually gives rise to a coherent representation of a category.

CLOSING THOUGHTS

A number of formal models of classification have suggested that performance on laboratory classification tasks is based on the storage and retrieval of exemplar information (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986). The results of the present experiments provide additional support for these models, since the subjects in classification learning were particularly sensitive to the changes in the specific exemplar information. However, the subjects' difficulty in learning the categories with multiple feature manifestations by classification raises a question about the degree to which classification can be taken as the primary route to category acquisition. Instead, we suggest that category acquisition involves a number of different procedures, including classification, inductive inference, comparison, abductive reasoning, generalization, and communication, acting in concert (see Markman, Yamauchi, & Makin, 1997; Ross, 1996, 1997).

In conclusion, the results of the present experiments are consistent with the view that the extraction of an appropriate description of a feature depends on the mechanism of comparison adopted during category learning. In particular, the alignment process applied during learning plays a key role in determining the commonalities and differences across different instances of a category. The use of alignment during category learning by inference allows categories to be learned, even when the feature manifestations vary on each trial. In contrast, the process underlying performance in classification makes it difficult for people to learn categories when the feature manifestations vary.

REFERENCES

- ANDERSON, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- CLEMENT, C. A., & GENTNER, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science*, 15, 89-132.
- ESTES, W. K. (1994). *Classification and cognition*. New York: Oxford University Press.

- GENTNER, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.
- 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.
- GENTNER, D., & MARKMAN, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, 52, 45-56.
- HINTZMAN, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411-428.
- 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, & Cognition*, 22, 754-770.
- MARKMAN, A. B. (1996). Structural alignment in similarity and difference judgments. *Psychonomic Bulletin & Review*, 3, 227-230.
- MARKMAN, A. B. (1997). Constraints on analogical inference. *Cognitive Science*, 21, 373-418.
- MARKMAN, A. B., & GENTNER, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, 25, 431-467.
- MARKMAN, A. B., & GENTNER, D. (1997). The effects of alignability on memory. *Psychological Science*, 9, 363-367.
- MARKMAN, A. B., & MEDIN, D. L. (1995). Similarity and alignment in choice. *Organizational Behavior & Human Decision Processes*, 63, 117-130.
- 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.
- MEDIN, D. L., GOLDSTONE, R. L., & GENTNER, D. (1993). Respects for similarity. *Psychological Review*, 100, 254-278.
- 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., & MEDIN, D. L. (1985). The roles of theories in conceptual coherence. *Psychological Review*, 92, 289-316.
- 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.
- POSNER, M. I., & KEELE, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353-363.
- REED, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, 3, 382-407.
- ROSS, B. H. (1996). Classification and the effects of interacting with instances. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22, 1249-1265.
- ROSS, B. H. (1997). The use of categories affects classification. *Journal of Memory & Language*, 37, 240-267.
- SCHYNS, P. G., GOLDSTONE, R. L., & THIBAUT, J. P. (1998). The development of features in object concepts. *Behavioral & Brain Sciences*, 21, 1-54.
- SCHYNS, P. G., & RODET, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 23, 681-696.
- SLOMAN, S. A. (1993). Feature-based induction. *Cognitive Psychology*, 25, 231-280.
- WISNIEWSKI, E. J., & MEDIN, D. L. (1991). The fiction and nonfiction of features. In R. S. Michalski & G. T. Tecuci (Eds.), *Machine learning: A multistrategy approach* (pp. 63-84). San Francisco: Morgan Kaufmann.
- 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 Seventeenth Annual Meeting of the Cognitive Science Society* (pp. 786-790). Hillsdale, NJ: Erlbaum.
- YAMAUCHI, T., & MARKMAN, A. B. (1998). Category-learning by inference and classification. *Journal of Memory & Language*, 39, 124-148.
- ZHANG, S., & MARKMAN, A. B. (1998). Overcoming the early entrant advantage via differentiation: The role of alignable and nonalignable differences. *Journal of Market Research*, 35, 413-426.

NOTES

1. Because analyses of the data from all the subjects (including the 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 all the experiments reported in this article.
2. The α level for all contrasts in this paper is corrected using the Bonferroni adjustment.
3. The average number of blocks spent in classification learning ($n = 3$) is considerably lower than that observed in Experiment 1B. This is due to 1 subject who finished the learning phase with only 4 blocks. Without this subject, the number of blocks spent in classification learning is on average 21.5.

(Manuscript received May 7, 1997;
revision accepted for publication December 18, 1998.)