Journal of Experimental Psychology:
Learning, Memory, and Cognition
2000, Vol. 26, No. 3, 776–795

0278-7393/00/$5.00 DOI: 10.1037//0278-7393.26.3.776

Inference Using Categories

Takashi Yamauchi and Arthur B. Markman
Columbia University

How do people use category membership and similarity for making inductive inferences? The authors addressed this question by examining the impact of category labels and category features on inference and classification tasks that were designed to be comparable. In the inference task, participants predicted the value of a missing feature of an item given its category label and other feature values. In the classification task, participants predicted the category label of an item given its feature values. The results from 4 experiments suggest that category membership influences inference even when similarity information contradicts the category label. This tendency was stronger when the category label conveyed class inclusion information than when the label reflected a feature of the category. These findings suggest that category membership affects inference beyond similarity and that category labels and category features are 2 different things.

Inductive inference is a fundamental use of categories. Current research identifies at least three crucial factors that govern inductive judgments using categories: (a) feature inheritance, (b) correlation of features across category members, and (c) focus on a single target category. Feature inheritance is facilitated by the hierarchical structure of many taxonomic categories in which specific categories are linked to more abstract categories with class inclusion relations. One reason that this structure is powerful is that it permits properties of subordinate categories to be inferred from superordinate categories (e.g., Quillian, 1968). For example, we may predict that a dolphin bears live young rather than laying eggs, because a dolphin is a mammal and mammals bear live young.

Category-based induction is also influenced by the correlations among features of category members. Members of a category generally have features in common. Individual birds differ in appearance, but they share many attributes such as having wings, beaks, and feathers. People are sensitive to the correlation among attributes (Anderson & Fincham, 1996; Medin, Altom, Edelson, & Freko, 1982), and these correlations can be used to infer properties of new items (Thomas, 1998). Gelman and Markman (1986) showed that young children predict that objects grouped by a common label share features even when they differ in appearance.

Finally, categories limit the information considered during inference. When an object is likely to belong to a category, people tend to derive inductive judgments about the object based primarily on that category, while paying less attention to features of other categories (Malt, Ross, & Murphy, 1995; Murphy & Ross, 1994; Ross & Murphy, 1996; Yamauchi & A. B. Markman, 1995, 1998, 2000). In predicting characteristics of a person, for example, we often emphasize social group membership, while dismissing other relevant categories. In an experimental setting, Murphy and Ross (1994) showed that participants primarily use feature information about a single category when making predictions rather than examining feature information about all possible categories, as a normative Bayesian model would predict (Anderson, 1990).

These examples suggest that category membership plays a crucial role in inductive judgments. One difficulty with this research, however, is that it is difficult to separate the influence of feature similarity from the influence of category membership in inference. Research addressing this topic has yielded mixed results (e.g., Gelman & E. Markman, 1986; Sloman, 1998). In one set of studies, Sloman (1998) held category membership constant and found that similarities among items modulated the way people made inductive inferences. In contrast, Gelman and E. Markman (1986) found that children's inferences were guided by category membership rather than overall similarities (e.g., the number of features shared between stimuli in question).

In this article, we explore the relative influence of similarity and category membership in predictive inference. To this end, we use an inference task and a classification task and manipulate both similarity and category membership. In the inference task, participants predict the value of one of five features of an item given four dimensions of feature values along with the category label of the test stimulus. In

Takashi Yamauchi and Arthur B. Markman, Department of Psychology, Columbia University.

This research was supported by National Science Foundation Grant SBR-99-05013. We thank Jim Corter, Dedre Gentner, Julia Kalmanson, Dave Krantz, Bob Krauss, Douglas Medin, Gregory Murphy, Page Moreau, Gail Musen, Robert Nosofsky, Tomislav Pavlicic, Brian Ross, Yung-Cheng Shen, Shi Zhang, and two anonymous reviewers for their helpful suggestions and comments while this research was being carried out.

Takashi Yamauchi is now at the Learning Research and Development Center, University of Pittsburgh. Arthur B. Markman is now at the Department of Psychology, University of Texas.

Correspondence concerning this article should be addressed to Takashi Yamauchi, the Learning Research and Development Center, Room 815, 3939 O'Hara Street, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, or to Arthur B. Markman, Department of Psychology, University of Texas, Austin, Texas 78712. Electronic mail may be sent to takashi+@pitt.edu or to markman@psy.utexas.edu.

the classification task, participants predict the category label of an item given five dimensions of feature values (see A. B. Markman, Yamauchi, & Makin, 1997; Yamauchi & A. B. Markman, 1995, 1998, 2000, for the inference and classification tasks). Thus, the two tasks are equivalent if category labels are like other features. That is, in the two tasks, participants predict the value of an unknown feature given five dimensions of feature values. In this setting, we examine the influence of similarity by varying the number of matching features between test items and members known to be in the category. We also assess the effect of category membership by modifying the meaning denoted by the label. If category labels are like other features, the performance on the inference task and that on the classification task would be equivalent. In contrast, if the presence of a shared category label (and thus membership in a common category) is different from the presence of a shared feature, then performance on the inference task and the classification task should diverge.

In the following sections, we begin by introducing operational definitions of inference and classification and describing how similarity and category membership are involved in classification and inference. Then, we provide a brief overview of the experiments. Finally, we present four experiments that examine the relationship between category membership and inductive judgment.

The Use of Similarity and Category Membership in Inference and Classification

We operationally define classification as a practice in which an exemplar is placed into one of two groups. We operationally define inference as a practice in which an attribute of an exemplar is predicted given the category label of the item as well as information about its other attributes. For example, classification as we define it is akin to the prediction of a category to which a person belongs (e.g., Democrat), having observed their attributes (e.g., supports affirmative action and favors reduced defense spending). Inference as we define it is akin to predicting an attribute of a person (e.g., supports affirmative action) given a category to which they belong, and other known attributes (e.g., is a Democrat and favors reduced defense spending).

We further define the term *category label* as a symbol that represents category membership by denoting a particular group of exemplars, and the term *category feature* as a symbol that denotes a characteristic of an exemplar.[1] Classification requires the prediction of the category label based on the features of the item; inference requires the prediction of a category feature based on the information about other features and the category label. Classification and inference are therefore formally equivalent if a category label is treated as other features (Anderson, 1990, 1991; but see Yamauchi & A. B. Markman, 1998, 2000, for further discussion).

We compare the mechanisms involved in classification and inference in order to probe the effects of similarity and category membership on inductive judgment. In many experimental settings, classification performance has been shown to be based on the number of overlapping features between a test stimulus and the memory trace of the stimuli that had been classified (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986; but see Malt, Sloman, Gennari, Shi, & Wang, 1999; Rips, 1989; Smith, 1989). Thus, if people make inductive judgments based solely on the similarity (i.e., feature overlap) between a test item and the category representation, then performance on the inference task should be the same as performance on the classification task. In contrast, if category membership is not treated as just another feature, then performance on an inference task should differ from that on a classification task.

Although some research suggests that a category label is just another feature of an item (Anderson, 1990, 1991), we believe that there are good reasons to regard category labels as distinct from other features of categories. One important difference between them is that labels and features are each linked to members of a category in a different way. Category labels are connected to an object by using class inclusion relations (e.g., This object *is a* dog). Category features are connected to an object using partonomic relations (e.g., This object *has* four legs). Not only is the relation different but the scope of the property is different as well. Category labels refer to a whole object, whereas category features refer only to parts of objects (Miller & Johnson-Laird, 1976; Tversky & Hemenway, 1984; but see Murphy, 1991). In language, category labels are often denoted by count nouns, whereas category features are often denoted as adjectives. Research has shown that infants as young as 12 to 21 months of age treat count nouns as if they name objects and adjectives as if they name properties of a basic level category (Waxman & Markow, 1995, 1998; Waxman, Shipley, & Shepperson, 1991).

Yamauchi and A. B. Markman (1998, 2000) explored the distinction between inference (in which participants predicted a feature value given information about other features and the category label) and classification (in which participants predicted the category label of a stimulus given information about other features) in the context of category learning. In the classification learning task, participants were given stimuli individually and learned to classify them into one of two categories as in a standard classification learning task (Medin & Schaffer, 1978; Posner & Keele, 1968). In the inference learning task, participants learned to infer unknown features of stimuli given a category label for the exemplar and information about its other features. The data from one set of studies (Yamauchi & A. B. Markman, 1998) were modeled with Nosofsky's generalized context model (GCM; Nosofsky, 1986) and Anderson's rational model (Anderson, 1990, 1991). The two models accounted nicely for the data obtained in classification transfer tasks, but they considerably underfit the data from inference transfer tasks. Perhaps more important than the overall fit of the model, however, when the GCM was fit to the data, the derived feature weights suggested that people were attending primarily to the category label. Although the focus on the category

[1] As we discuss later, category labels and category features also differ in the relation that connects the label and the category.

label was not sufficient to allow the GCM to account for the entire pattern of data obtained in the studies, it does suggest that category labels are particularly important for inference relative to other features.

In short, we think that there is a good reason to believe that category labels and category features are different entities. For this reason we argue that category labels, along with other features, exert a discrete influence on inductive inferences using categories. The present studies aim to provide evidence for this supposition and to explore why labels and features differ and how people make inferences.

Overview of the Experiments

The present studies place category labels and category features in opposition. In each experiment, participants are given a sample sheet depicting 10 members of two categories (5 for each category; see Figure 1) and they answer 30 classification and/or 30 inference questions shown in a booklet based on the group defined in the sample sheet (see Murphy & Ross, 1994, for a similar procedure). Each category consists of schematic figures of imaginary bugs that possess five features with binary values (*horn* = long or short, *head* = round or angular, *body* = dotted or striped, *legs* = eight legs or four legs, *tail* = short or long) and a category label ("monek" or "plaple").

As shown in Table 1, the two categories have a family-resemblance structure, which was derived from two prototypes (M0 and P0) by modifying the values of one of five features (see Medin, Wattenmaker, & Hampson, 1987). For example, to produce the stimulus *M1*, the value of the tail was changed from the prototype M0 so that it had four features consistent with the prototype M0 and one feature consistent with the prototype P0 (Table 1). On classification questions, participants predicted the category label of a stimulus given five dimensions of feature values (Figure 2a). On inference questions, participants predicted the value of one of the five features, given four dimensions of feature values along with the category label of the test stimulus (Figure 2b). Thus, the stimuli used for the classification questions were formally equivalent to the stimuli used for the inference questions if category labels and category features are the same.

The key variable of interest was the degree of similarity between the test stimulus and the exemplars on the sample sheet, which was defined by the number of matching features of the test stimulus to the prototype of the corresponding sample category. We devised three levels of similarity (i.e., ***high, medium, or low level of feature match***) with 10 classification questions and 10 inference questions at each level (see Table 2). At the high level of feature match, each test stimulus had four features in common with the prototype of the corresponding category and one feature in common with the prototype of the other category, and participants were asked to predict the value of the unspecified characteristic of the stimulus (Figure 2). Similarly, each test stimulus at the medium level of feature match had three features in common with the prototype of the corresponding category and two features in common with the prototype of the other category. Each test stimulus at the low level of feature match had two features in common with the prototype of the corresponding category and three features in common with the prototype of the other category. We examined participants' response patterns as a function of these three levels of feature match (i.e., high, medium, and low).

The number of responses in accordance with the category from which the exemplar was derived (which we call *category-accordance responses* in Table 2) should mirror the three levels of feature match, if participants assess matching features to make classification judgments (see Medin & Schaffer, 1978). On a classification question, for example, the stimulus *M11* has four features in common with the prototype of "monek" and one feature in common with the prototype of "plaple"; therefore, participants should prefer to respond "monek" rather than "plaple."

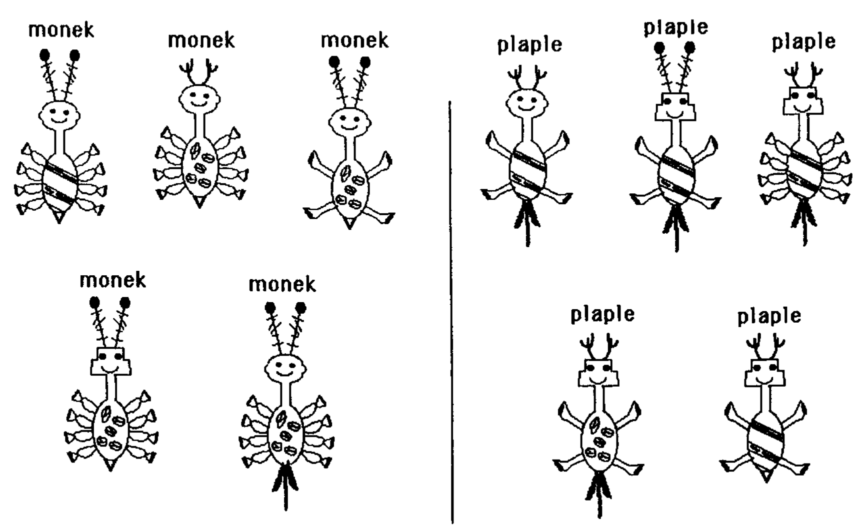

Figure 1. The sample sheet used for Experiments 1–4.

Table 1
Category Structure Used in Experiments 1–4

Monek							Plaple						
Stimuli	Horns	Head	Body	Legs	Tail	Label	Stimuli	Horns	Head	Body	Legs	Tail	Label
M1	1	1	1	1	***0***	1	*P1*	0	0	0	0	***1***	0
M2	1	1	1	***0***	1	1	*P2*	0	0	0	***1***	0	0
M3	1	1	***0***	1	1	1	*P3*	0	0	***1***	0	0	0
M4	1	***0***	1	1	1	1	*P4*	0	***1***	0	0	0	0
M5	***0***	1	1	1	1	1	*P5*	***1***	0	0	0	0	0
M0 (prototype)	1	1	1	1	1	1	P0 (prototype)	0	0	0	0	0	0

Note. The value 1 = long horns, round head, dotted body, eight legs, and short tail. The value 0 = short horns, angular head, striped body, four legs, and long tail. M = Monek; P = Plaple. Bold italics = feature values consistent with the prototype of the opposite category.

This tendency should diminish to some extent at the medium level of feature match (e.g., *M16*). At the low level of feature match, this tendency should fall sharply. Because the stimuli at this level have two features consistent with the corresponding category and three features consistent with the other category, participants should classify this stimulus into the opposite category (e.g., *M21*). As a consequence, the proportion of category-accordance responses is expected to be lowest at the low level of feature match.

In contrast to classification questions, inference questions could yield three possible response patterns, depending on the way participants use category labels for their judgments. One possibility is that participants will treat the category label as another feature. In this case, the performance on the inference questions will be indistinguishable from that for the classification questions. That is, the number of category-accordance responses in each task will mirror the three levels of feature match.

A second possibility is that participants will ignore similarity altogether and use category membership exclusively to make inferences. In this case, the three levels of feature match will not affect inference performance, and so we should have about the same number of category-accordance responses at each level of feature match in the inference task.

A third possibility, which we favor, suggests that category membership and similarity both influence inference. Studies, like the one by Sloman (1998) described previously, demonstrate that similarity affects predictive inferences when category membership is held constant. Nonetheless, similarity does not completely determine category membership (Malt et al., 1999; Rips, 1989; Smith, 1989). In this sense, it is likely that category labels, along with other features, separately affect inferences. In our experiments, this view predicts that the difference between inference and classification will emerge primarily on the questions at the low level of feature match. At the high level of feature match, there are three features and one category label consistent with the prototype of the corresponding category and one feature consistent with the prototype of the opposite category. Similarly, given the medium level of feature match, there are two features and one category label consistent with the prototype of the corresponding category and two features consistent with the prototype of the opposite category (Table 2). In these cases, both the matching features and the category label favor a response in accordance with the corresponding category, though there will be more category-accordance response at the high level of feature match than at the medium level of feature match.

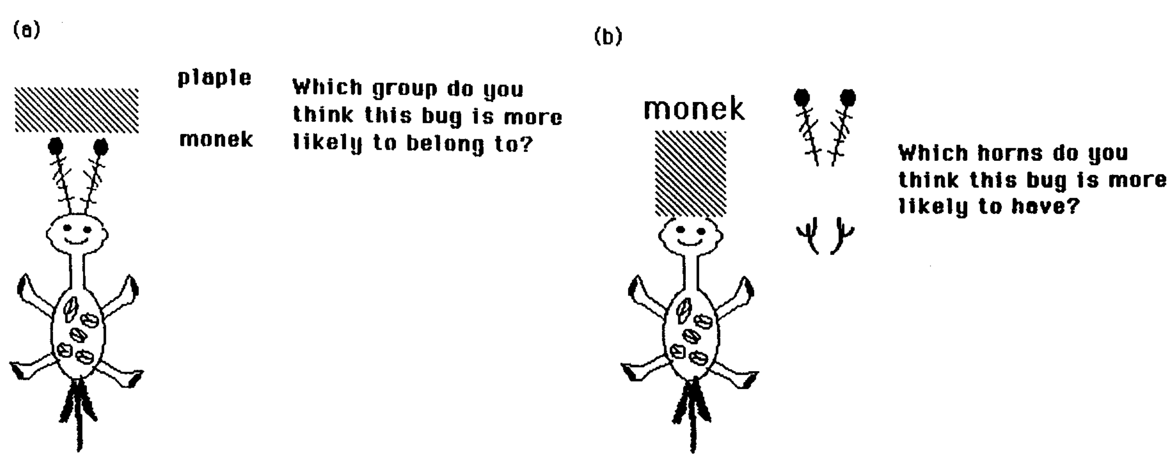

Figure 2. (a) A sample stimulus used for a classification question. (b) A sample stimulus used for an inference question.

Table 2
Stimulus Structure Used for Classification and Inference Questions in Experiments 1, 2, and 4

Monek							Plaple						
Stimuli	Horns	Head	Body	Legs	Tail	Target labels	Stimuli	Horns	Head	Body	Legs	Tail	Target labels
High													
M11	***1***	1	1	1	0	1	*P11*	***0***	0	0	0	1	0
M12	1	1	1	0	***1***	1	*P12*	0	0	0	1	***0***	0
M13	1	1	0	***1***	1	1	*P13*	0	0	1	***0***	0	0
M14	1	0	***1***	1	1	1	*P14*	0	1	***0***	0	0	0
M15	0	***1***	1	1	1	1	*P15*	1	***0***	0	0	0	0
Medium													
M16	***1***	1	1	0	0	1	*P16*	***0***	0	0	1	1	0
M17	1	1	0	0	***1***	1	*P17*	0	0	1	1	***0***	0
M18	1	0	0	***1***	1	1	*P18*	0	1	1	***0***	0	0
M19	0	0	***1***	1	1	1	*P19*	1	1	***0***	0	0	0
M20	0	***1***	1	1	0	1	*P20*	1	***0***	0	0	1	0
Low													
M21	***1***	0	1	0	0	1	*P21*	***0***	1	0	1	1	0
M22	0	1	0	0	***1***	1	*P22*	1	0	1	1	***0***	0
M23	1	0	0	***1***	0	1	*P23*	0	1	1	***0***	1	0
M24	0	0	***1***	0	1	1	*P24*	1	1	***0***	1	0	0
M25	0	***1***	0	1	0	1	*P25*	1	***0***	1	0	1	0

Note. High, Medium, and Low = three levels of feature match (high, medium, and low levels of feature match, respectively). Bold italics = target features for inference questions; M = Monek; P = Plaple. Category-accordance responses (which were the dependent measure of Experiments 1–4) were the responses consistent with the values indicated in the target features and the target labels. The value 1 = long horns, round head, dotted body, eight legs, and short tail. The value 0 = short horns, angular head, striped body, four legs, and long tail.

At the low level of feature match, feature-matching information and category labels are opposed. The test stimuli at the low level of feature match have one feature and one category label consistent with the corresponding category, but they have three features that are typical of the opposite category (Table 2). Given the stimulus *M21,* for example, the feature body and the category label have the value 1; the features head, legs, and tail have the value 0; and participants are asked to predict the value of horns. Thus, if a category label along with other features uniquely influences inductive judgments, then participants should still favor feature values consistent with the *target* category rather than the *opposite* category even though feature-match information predicts otherwise (for the use of a single category for inductive judgment, see Malt et al., 1995; Murphy & Ross, 1994; Ross & Murphy, 1996). This pattern of responses would result in an interaction between the degree of feature match and the question type at the low level of feature match.

Throughout the four experiments, we assume that the number of category-accordance responses reflects the extent to which category membership was used for inductive judgment. On the basis of this assumption, we first compare participants' response patterns on classification questions and on inference questions in Experiment 1. In order to test directly the impact of category labels on the inference task, we manipulate the meaning denoted by category labels in three inference conditions in Experiments 2 and 3. Finally, in Experiment 4, we probe further a specific element that distinguishes category labels from other features.

Experiment 1

The goal of Experiment 1 was to demonstrate the distinction between inference and classification by contrasting the impact of category labels and category features at the low level of feature match. In Experiment 1, the participants responded to 30 classification questions as well as to 30 inference questions, which are subdivided into high, medium, and low levels of feature match. As discussed earlier, we expected that participants would make category-accordance responses to inference questions more often than to classification questions. However, given the hypothesis that similarity as well as category labels affect inductive inference, we predicted that participants would make category-accordance responses to inference questions more often than to classification questions at the low level of feature match.

Method

Participants. Twenty-five members of the Columbia University community participated in the experiment for a cash payment.

Materials. The materials used in Experiment 1 were shown in a booklet that consisted of a 1-page sample sheet (Figure 1), 1 page of instructions, 2 pages of examples, and 60 pages of questions (Figure 2). The sample sheet showed schematic drawings of bugs that were arranged into two categories given the novel names "monek" and "plaple." The two categories had a family-resemblance structure (see Table 1 and Medin et al., 1987) and were derived from a prototype (M0 or P0). All participants saw the same stimulus set. To separate the two categories, the bugs belonging to "monek" were placed on the left side of the sheet, and the bugs belonging to "plaple" were placed on the right side of the sheet.

Sixty question sheets were produced on the basis of the structure shown in Table 2. Each question sheet contained a drawing of a bug, a question about the bug, and a 7-point confidence scale. There were 30 classification questions and 30 inference questions. For the classification questions, the category name was covered by a mask as shown in Figure 2a, and for the inference questions, one of five features of a drawing was covered by a mask.

The 30 inference and 30 classification questions were subdivided into three levels of feature match (high, medium, and low) based on the number of features shared with the corresponding prototype (Table 2). For example, assuming that a category label is just another feature, stimuli with the high level of feature match had four features in common with the prototype of their own category and one feature in common with the prototype of the other category. The stimuli with the medium level of feature match had three features in common with the prototype of their own category and two features in common with the prototype of the other category. The stimuli with the low level of feature match had two features in common with the prototype of their own category and three features in common with the prototype of the other category. In this manner, 60 questions were grouped into 6 sets of 10 questions.

The sample sheet was not attached to the rest of the booklet, so that participants were able to look at the samples during the experiment. All participants received an identical booklet containing the sample sheet, instructions, examples, and question sheets. The order of the question sheets was determined randomly for each participant.

Procedure. The basic procedure of the experiment involved answering 60 questions (30 classification and 30 inference questions) shown in a stimulus booklet. The instructions indicated that the labels "monek" and "plaple" represented "two types of imaginary bugs" and that participants were asked to answer each question on the basis of the sample sheet (see the Appendix for the exact instructions).

Thirty classification questions and 30 inference questions were produced from the stimulus structure shown in Table 2. Classification questions and inference questions differed in the type of features being predicted. For the classification questions, participants predicted the category label of a stimulus given information about the other features (Figure 2a). For the inference questions, participants predicted the value of one of five features given the other four features and the category label (Figure 2b). The classification question was phrased as "Which group do you think this bug is more likely to belong to?" The inference question was phrased as "Which legs do you think this bug is more likely to have?" Participants were tested individually and proceeded through the booklets at their own pace. The entire experiment took about 25 min.

Design. The design of the experiment was a 2 (question type: classification questions and inference questions) × 3 (feature match: high, medium, and low) within-subjects factorial. The dependent measure was the proportion of responses made in accordance with the prototype of the target category (i.e., category-accordance responses; see Table 2 for the definition). For example, given the stimulus *M11*, we measured the proportion of 1 responses; given the stimulus *P11*, we measured the proportion of 0 responses.

Results

The main results of Experiment 1 are shown in Figure 3. The data were analyzed with a 2 (label type: inference and classification) × 3 (feature match: high, medium, and low) analysis of variance (ANOVA). In this experiment, we made the following two predictions. First, because similarity is expected to affect inferences, the number of category-accordance responses should decrease as feature match decreases for both classification questions and inference questions. Second (and more importantly) at the low level of feature match, participants should make category-accordance responses to the inference questions more often than to the classification questions because category labels are expected to have a strong impact on inference.

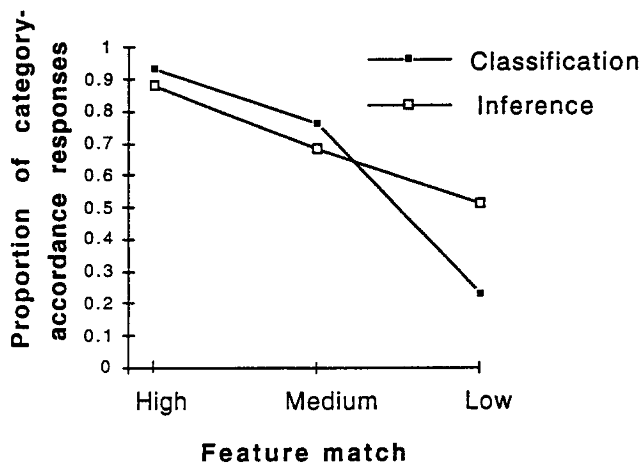

Figure 3. Performance for classification and inference questions in Experiment 1.

Consistent with this prediction, there was a main effect of feature match, $F(2, 48) = 77.26$, $MSE = 0.05$, $p < 0.01$, as well as an interaction between question type and feature match, $F(2, 48) = 16.62$, $MSE = 0.03$, $p < 0.01$. Participants given inference questions at the low level of feature match made category-accordance responses more often than did participants given classification questions, $t(24) = 3.82$, $p < 0.03$ (Bonferroni adjustment). Participants did not differ, however, at the high level of feature match (inference questions and classification questions), $t(24) = 2.14$, $p > 0.10$; or at the medium level of feature match (inference questions and classification questions), $t(24) = 1.54$, $p > 0.10$. Overall, there was no difference in the two question types, $F(1, 24) = 2.46$, $MSE = 0.04$, $p > 0.10$.

Because the ANOVA is based on aggregate data, it may hide differences among individual participants. In order to describe the extent to which participants made category-accordance responses in classification and inference, we counted the number of individual participants who made responses that were either based primarily on the corresponding categories or primarily on the opposite category at the low level of feature match. For inference questions, 7 out of 25 participants made responses consistent with the corresponding category more than 75% of the time, and 8 out of 25 participants made responses in accordance with the opposite category more than 75% of the time at the low level of feature match. In contrast, for classification questions at the low level of feature match, only 1 out of 25 participants made responses consistent with the corresponding category more than 75% of the time, but 18 out of 25 participants made responses in accordance with the opposite category more than 75% of the time. Thus, most of the participants used feature-match information on classification questions, but this tendency was substantially diminished for the inference questions. These results suggest that participants adopted different strategies to deal with inference and classification questions.

Discussion

The goal of Experiment 1 was to contrast participants' performance on inference and classification questions in order to examine the relative contribution of category membership and category features in inductive judgment. As discussed earlier, the two conditions are equivalent if category labels are processed the same way as other features. In contrast, if labels are not treated as features, we predicted that participants would make category-accordance responses on inference questions more often than on classification questions at the low level of feature match. The data supported this prediction. Despite the fact that the stimuli at the low level of feature match have more features in common with the members of the other category than with those of the corresponding category, participants were likely to focus on category membership and to discount feature-matching information. These results are in accordance with the view that category membership influences inductive judgment beyond similarity information. We must now address why category labels and category features are different. In the following three studies, we explore ways that labels and features differ and examine the impact of these differences on participants' performance in the inference and classification tasks.

Experiment 2

In Experiment 1, we hypothesized that the category membership information provided by category labels would lead feature predictions to be consistent with the label. On this view, if a category had a label that did not strongly imply category membership, then the number of category-accordance responses should be lower than when the label does imply category membership (particularly at the low level of feature match). To test this hypothesis, we contrasted performance on the two conditions run in Experiment 1 with two other inference conditions: one that reduces the degree to which the label implies category membership and one that replaces the label with a picture.

Thus, there are three different inference tasks in Experiment 2. The inference questions given in Experiment 1 are called the *inference-type condition,* because the category labels ("monek" and "plaple") are described as denoting *types* of bugs (see the Appendix). In a second inference task, called the *inference-shape condition,* the labels are described as denoting *shapes* of bugs' wings. Specifically, the instructions explained that the bugs labeled with "monek" *have* "monek-shaped wings," and the bugs labeled with "plaple" *have* "plaple-shaped wings." Finally, in a third task, called the *inference-picture condition,* the labels are replaced on the sample sheet and in the inference questions with drawings of wings (as shown in Figure 4).

In short, the inference-shape condition was identical to the inference-type condition except for the way that the two labels were described in the instructions. The inference-picture condition was equivalent to the inference-shape condition except that all of the labels were removed from the stimulus set and the sample sheet, and actual drawings of wings were attached to the bodies of the bugs. This design

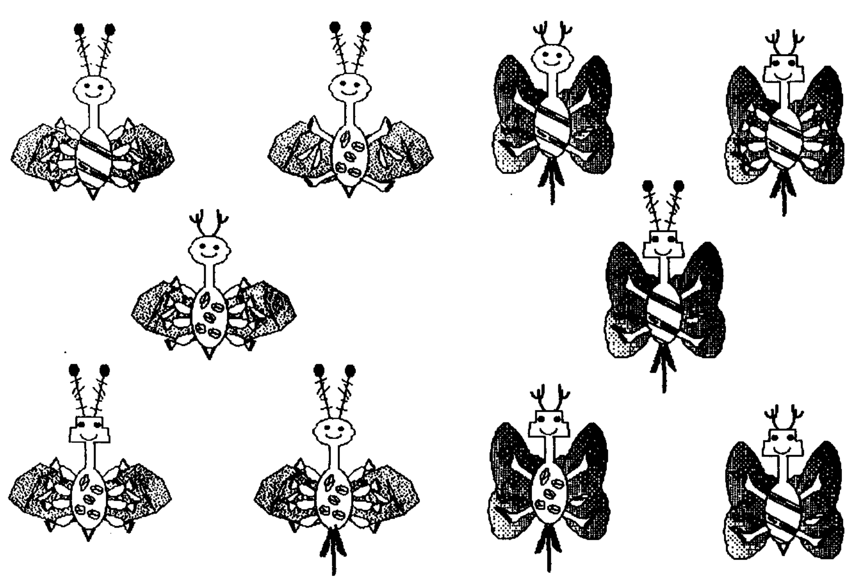

Figure 4. The sample sheet used for the inference-picture condition in Experiment 2.

allowed us to directly examine the impact of category membership on the inference task.

The classification and inference-type conditions are expected to replicate the results of Experiment 1, with the classification condition yielding fewer category-accordance responses at the low level of feature match than the inference-type condition. If the class inclusion information contained in the label drives the observed difference between inference and classification, then fewer category-accordance responses at the low level of feature match should be made in the inference-shape and inference-picture conditions than in the inference-type condition.

Because the label is completely absent in the inference-picture condition, performance in this condition should be indistinguishable from that in the classification condition. If the category membership information implied by the label is the only thing that differentiates inference from classification, then the inference-shape condition should also be indistinguishable from the classification condition. If there are other influences than the presence of a category label (e.g., procedural differences associated with the two tasks), then we would observe performance in the inference-shape condition that is between that in the inference-type and classification conditions.

Method

Participants. Participants were 100 members of the Columbia University community, 90 of whom received a payment of $4 and 10 of whom received course credit.

Materials. The stimuli and sample sheets used in the classification condition were identical to the stimuli used in the classification questions of Experiment 1, except that the line separating the categories was removed from the sample sheet. The stimuli and sample sheet used in the inference-type condition and in the inference-shape condition were identical to the stimuli used in the inference questions in Experiment 1, except that the line separating the categories was removed from this sample sheet as well. In the inference-picture condition, the category labels were removed from the sample sheet, and the stimulus set and drawings of wings were attached to each sample stimulus (as shown in Figure 5; see the Appendix for a minor change in the instructions between the inference-shape condition and the inference-picture condition). In the classification and inference-type conditions, the category labels "monek" and "plaple" were described as denoting "two types of bugs," and in the inference-shape condition, the labels were described as denoting "two shapes of wings." The Appendix contains the text of the instructions.

Procedure. The basic procedure of the experiment involved answering 30 questions provided in a booklet (either 30 classification questions or 30 inference questions depending on conditions). Prior to the experiment, participants were randomly assigned to one of four conditions: classification, inference-type, inference-shape, and inference-picture. In the classification condition, participants predicted the category label of each stimulus as in Experiment 1. In the other inference conditions, participants predicted one of five features of each stimulus.

Design. The design of this experiment was a 4 (label–type: classification, inference-type, inference-shape, and inference-picture) $\times$ 3 (feature–match: high, medium, and low) mixed factorial. Label–type was a between-subjects factor and feature match was a within-subjects factor.

Results

The major results of Experiment 2 are shown in Figure 6. Three different ANOVAs were used to examine three different questions. First, a 2 (label–type: classification vs. inference-type) $\times$ 3 (feature match: high, medium, and low) ANOVA was used to investigate the distinction between inference and classification. Second, the data were analyzed with a 3 (label–type: inference-type, inference-shape, and inference-picture) $\times$ 3 (feature match: high, medium, and low) ANOVA to probe the impact of the modification of the category labels. Finally, another 2 (label–type: classification

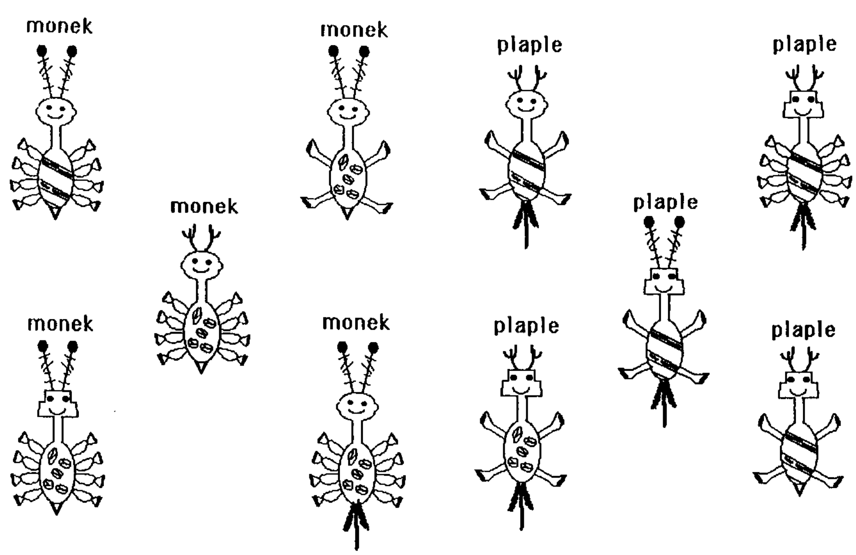

Figure 5. The sample sheet used for the classification, inference-type, and inference-shape conditions in Experiment 2.

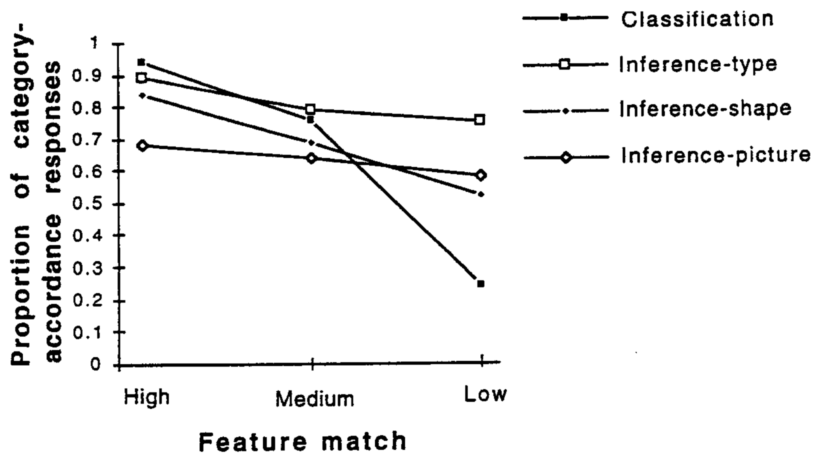

Figure 6. Performance in the classification, inference-type, inference-shape, and inference-picture conditions in Experiment 2.

vs. inference-picture) × 3 (feature match: high, medium, and low) ANOVA was used to test the idea that inference and classification would be identical if category labels were removed from the question and sample sheets.

First, the comparison between the inference-type and classification conditions indicates that participants in the inference-type condition ($M = 0.81$) made on average significantly more category-accordance responses than did participants in the classification condition ($M = 0.65$), $F(1, 48) = 15.29$, $MSE = 0.07$, $p < 0.01$. The interaction between label–type and feature match was also significant, $F(2, 96) = 36.12$, $MSE = 0.03$, $p < 0.01$. Consistent with the results from Experiment 1, planned *t* tests at the low level of feature match show that participants in the inference-type condition, which is formally equivalent to the classification condition, made significantly more category-accordance responses than participants given the classification condition, $t(48) = 6.8$, $p < 0.01$ (Bonferroni). The main effect of feature–match was also significant, $F(2, 96) = 72.7$, $MSE = 0.03$, $p < 0.01$.

For individual participants, 16 out of 25 participants in the inference-type condition made category-accordance responses at the low level of feature match more than 75% of the time, and 3 out of 25 participants responded consistently with the opposite category at the low level of feature match more than 75% of the time. In contrast, no participant in the classification condition made category-accordance responses at the low level of feature match more than 75% of the time, and 14 out of 25 participants responded with the opposite category more than 75% of the time. The results indicate that the responses made in the inference-type and classification conditions, though formally identical, were qualitatively different. These results are consistent with the findings in Experiment 1.

As predicted, the three inference conditions (inference-type, inference-shape, and inference-picture) yielded distinct patterns of performance (Figure 6). There was a main effect of feature match, $F(2, 144) = 21.44$, $MSE = 0.03$, $p < 0.01$, which simply reflects that the number of category-accordance responses decreased with the level of feature match. There was also a main effect of label–type, $F(2, 72) = 6.84$, $MSE = 0.09$, $p < .01$. Further analysis revealed that participants in the inference-type condition, in which the two labels were described as representing "two types of bugs," made more category-accordance responses than participants in the inference-picture condition, $t(48) = 3.95$, $p < .01$. More category-accordance responses were made in the inference-type condition than in the inference-shape condition (where the two labels were described as representing "two shapes of wings"), but this difference was only marginally significant, $t(48) = 2.23$, $.05 < p < .10$.

Interpretation of these main effects must be made in light of a significant interaction between label–type and feature match, $F(4, 144) = 2.67$, $MSE = 0.03$, $p < .05$. To identify the source of the interaction, we conducted planned comparisons with three *t* tests at each level of feature match because the difference among the three inference conditions is theoretically important only within a given level of feature match. First, comparisons between the inference-type condition and the inference-shape condition suggest that the two conditions differed, as expected, at the low level of feature match. The number of category-accordance responses that was elicited at the high and the medium level of feature match was roughly equivalent in the two conditions; at the high level of feature match, $t(48) = 1.06$, $p > .10$; at the medium level of feature match, $t(48) = 1.60$, $p > .10$. However, at the low level of feature match, participants in the inference-type condition made category-accordance responses significantly more often than did participants in the inference-shape condition, $t(48) = 2.48$, $p < .05$.

Comparisons between the inference-type condition and the inference-picture condition suggest that participants in the inference-type condition made category-accordance responses more often than did participants in the inference-picture condition at all three levels of feature match; the high level of feature match, $t(48) = 4.82$, $p < .01$; the medium level of feature match, $t(48) = 3.22$, $p < .01$; the low level of feature match, $t(48) = 2.25$, $.05 < p < .10$. Participants in the inference-shape condition also made significantly more category-accordance responses than did participants in the inference-picture condition at the high level of feature match, $t(48) = 3.07, p < .05$.

Analyses of the patterns of responses by individual participants clarify the impact of the category labels. In the inference-type condition, 16 out of 25 participants made category-accordance responses at the low level of feature match more than 75% of the time, and 3 out of 25 participants responded consistently with the opposite category at the low level of feature match more than 75% of the time. Given the inference-shape condition and the inference-picture condition, the number of participants who made category-accordance responses more than 75% of the time decreased substantially. In the inference-shape condition, 9 out of 25 participants made category-accordance responses at the low level of feature match more than 75% of the time, and 2 out of 25 participants responded consistently with the opposite category at the low level of feature match more than 75% of the time. In the inference-picture condition, 5 out of 25 participants made category-accordance responses at the low level of feature match more than 75% of the time, and 4 out of 25 participants responded consistently with the opposite category at the low level of feature match. The data support the notion that the number of category-accordance responses is higher when the category labels convey class-inclusion information.

Finally, the comparison between the classification condition and the inference-picture condition revealed some differences. Overall there was no main effect of label–type: classification condition ($M = 0.65$); inference-picture condition ($M = 0.63$), $F(1, 48) < 1.0$, $MSE = 0.02$, $p > .10$. However, there was an interaction between question type and feature match, $F(2, 96) = 41.15$, $MSE = 0.03$, $p < .01$; as well as a main effect of feature match, $F(2, 96) = 74.5$, $MSE = 0.03$, $p < .01$. The interaction indicated that the classification condition and the inference-picture condition differed at the high level of feature match and at the low level of feature match, respectively, $t(48) = 7.5$, $p < .01$; $t(48) = 5.7, p < .01$ (Bonferroni).

Discussion

Several aspects of the results of Experiment 2 merit special attention. First, the data support the claim that category membership information provided by category labels is an important determinant of the difference between inference and classification. In particular, changing the interpretation of the label from denoting category membership to just denoting the shape of a wing decreased the number of category-accordance responses. Furthermore, eliminating the label altogether but characterizing the task as a feature-prediction task also decreased the number of category-accordance responses relative to the inference-type condition.

Though not predicted, the overall difference between the inference-type condition and the other two inference conditions was also substantial. The main effect observed in the three inference conditions may suggest that different strategies were indeed used for inference when category labels were present in the stimulus frame and when category labels were not present in the stimulus frame. In particular, participants in the inference-picture condition made category-accordance responses significantly fewer times than did participants in the inference-type condition at the three levels of feature match. It is possible that without any labels participants in the inference-picture condition adopted a variety of different strategies, including feature matching or looking for identical exemplars. We address this issue further in Experiment 4.

Contrary to our prediction, participants in the inference-picture condition made category-accordance responses more often than did participants in the classification condition at the low level of feature match. We hypothesized earlier that performance in the two conditions would be the same, because inference and classification are expected to differ in the presence of category labels.

A close examination of the two tasks suggests that participants may have to use the sample sheet differently in each. In the inference-picture condition, the participants predicted the value of one feature whose value was unknown, given the values for the shape of the wings and three of the other features. The question stimuli supplied in the inference-picture condition always contained the feature wings. On the sample sheet, all of the wings of one shape appeared on the left side of the sheet, and all of the wings with the other shape appeared on the right. Thus, in the inference-picture condition, it might have been convenient for participants to direct their attention only to the side of the sample sheet containing bugs with wings that matched those on the question sheet. This procedure would lead participants in the inference-picture condition to select the feature value most prevalent on that side of the sample sheet.

By contrast, in the classification condition, participants were asked to predict the label of a stimulus, but none of the features that were presented in a given question stimulus was shown on only one side of the sample sheet. Thus, participants in the classification condition had to search both sides of the sample sheet in order to respond. This analysis suggests that there were more category-accordance responses in the inference-picture condition than in the classification condition because the two conditions differed in the way that the sample sheet was used. This procedural distinction is clear in Table 1. The participants in the classification condition predicted the category label of a stimulus given the features that have a family resemblance structure; the participants in the inference-picture condition predicted the value of an unspecified feature given other features that have a family-resemblance structure and the feature wings whose values are perfectly correlated with the division of the two groups of the bugs.

The same argument can be applied to the two other inference conditions. That is, some amount of the difference between each of the inference conditions and the classification may be due to the way that the sample sheet was used in the inference conditions rather than because of some inherent difference between classification and inference. We address this issue in Experiment 3.

Table 3
Category Structure Used in the Three Inference Conditions in Experiment 3

Monek							Plaple						
Stimuli	Label	Head	Body	Legs	Tail	Horns	Stimuli	Label	Head	Body	Legs	Tail	Horns
M1	1	1	1	1	***0***	1	*P1*	0	0	0	0	***1***	0
M2	1	1	1	***0***	1	1	*P2*	0	0	0	***1***	0	0
M3	1	1	***0***	1	1	1	*P3*	0	0	***1***	0	0	0
M4	1	***0***	1	1	1	1	*P4*	0	***1***	0	0	0	0
M5	***0***	1	1	1	1	1	*P5*	***1***	0	0	0	0	0
M0 (prototype)	1	1	1	1	1	1	*P0 (prototype)*	0	0	0	0	0	0

Note. The category structures used in Experiment 3 were inference-type, inference-shape, and inference-picture condition. The value 1 = long horns, round head, dotted body, eight legs, and short tail. The value 0 = short horns, angular head, striped body, four legs, and long tail. M = Monek; P = Plaple. The positions for Label and Horns are switched so that the two labels have the same structure as the other four features, and the feature Horns has a necessary-and-sufficient structure. Bold italics = feature values consistent with the prototype of the opposite group as shown in the sample sheet (see Figure 7; the category labels are assumed to be just another feature).

Experiment 3

In Experiment 3, the sample sheets for the three inference conditions were modified so that no single feature presented in the question stimuli divided the sample sheet in half. This change makes the inference conditions more like the classification conditions of Experiments 1 and 2. As shown in Table 3, in the inference-type and inference-shape conditions, the arrangement of the feature horns on the sample sheet is switched with the arrangement of the two labels. Thus, in Figure 7, all of the stimuli with one type of horns are on the left, and the stimuli with the other type of horns are on the right. The participants in the inference-type and inference-shape conditions predicted the value of the feature horns (i.e., long or short) given this sample sheet. We expected that this modification would lead participants to consider all of the items on the sample sheet when responding in the inference conditions as well as in the classification condition.

As in Experiment 2, the inference-shape condition is identical to the inference-type condition except that the two labels are described in the instructions as representing "two shapes of wings." The inference-picture condition is equivalent to the inference-type condition except that the two labels are eliminated from the sample sheet and actual drawings of wings are attached to the bodies of sample bugs (Figure 8).

In a change from Experiment 2, slightly different test stimuli are used in this experiment (Table 4). In the three inference conditions, 8 of 10 stimuli at each level of feature match have category labels consistent with the target category, and 2 of 10 stimuli at each level of feature match have category labels consistent with the opposite category. At the medium level of feature match, for example, the stimuli *M16–M19* and *P16–P19* have the feature value 1 and 0, respectively, and the stimuli *M20* and *P20* have the feature value 0 and 1, respectively. This modification is introduced to ensure that the same number of test stimuli have a category label consistent with the corresponding category at the three levels of feature match.

In this manipulation, the number of category-accordance

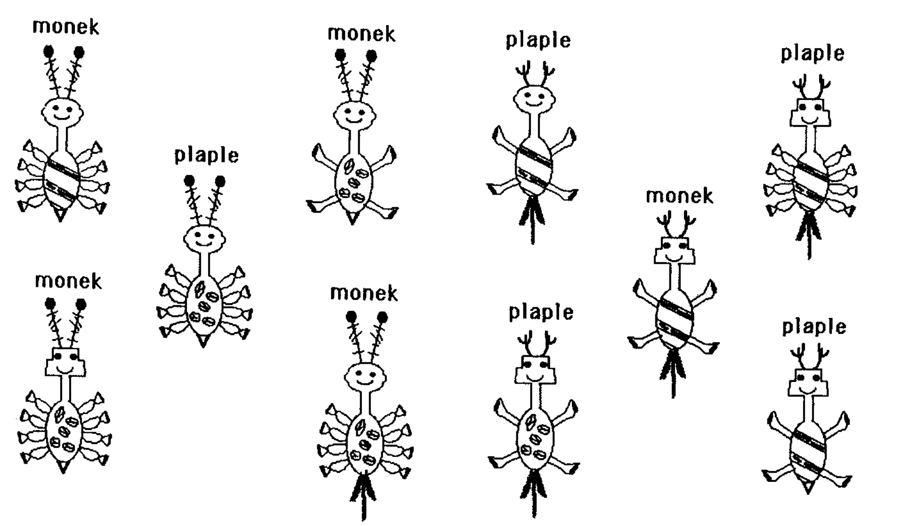

Figure 7. The sample sheet used for the inference-type and inference-shape conditions in Experiment 3.

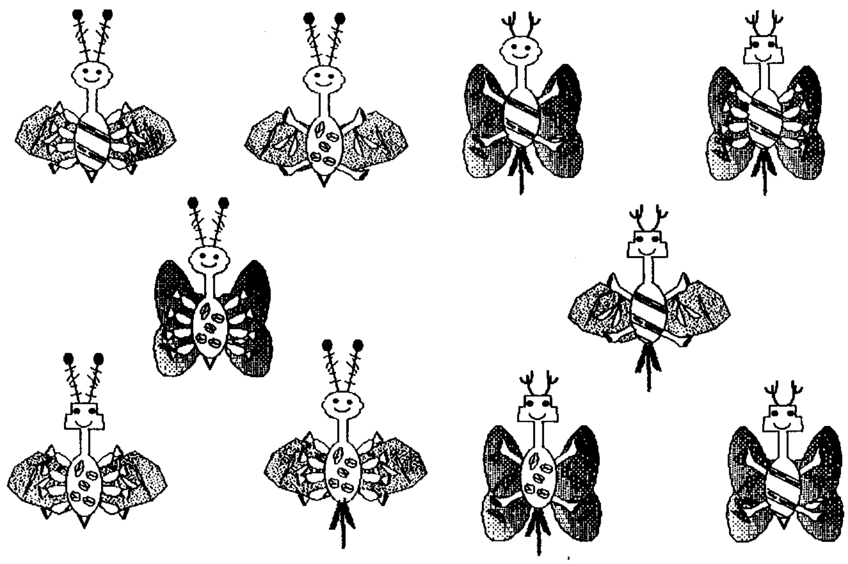

Figure 8. The sample sheet used for the inference-picture condition in Experiment 3.

responses should decrease substantially in the three inference conditions at the low level of feature match. Nonetheless, participants in the inference-type condition should make category-accordance responses more often than participants in the classification condition, if category membership and similarity both affect inductive judgment. Furthermore, participants should make more category-accordance responses at the low level of feature match when the labels denote types of bugs (in the inference-type condition) than when they denote shapes of wings (in the inference-shape condition) or when they are removed from the sample sheet and replaced with drawings of wings (i.e., inference-picture condition). In the inference-picture condition, participants' response patterns should be indistinguishable from those in

Table 4
Inference/Classification Question Stimuli Structure Used for Experiment 3

Monek							Plaple						
Stimuli	Label/Horns	Head	Body	Legs	Tail	Target Horns/Label	Stimuli	Label/Horns	Head	Body	Legs	Tail	Target Horns/Label
						High							
M11	1	1	1	1	0	1	*P11*	0	0	0	0	1	0
M12	1	1	1	0	1	1	*P12*	0	0	0	1	0	0
M13	1	1	0	1	1	1	*P13*	0	0	1	0	0	0
M14	1	0	1	1	1	1	*P14*	0	1	0	0	0	0
M15	0	1	1	1	1	1	*P15*	1	0	0	0	0	0
						Medium							
M16	1	1	1	0	0	1	*P16*	0	0	0	1	1	0
M17	1	1	0	0	1	1	*P17*	0	0	1	1	0	0
M18	1	0	0	1	1	1	*P18*	0	1	1	0	0	0
M19	1	0	1	1	0	1	*P19*	0	1	0	0	1	0
M20	0	1	1	0	1	1	*P20*	1	0	0	1	0	0
						Low							
M21	1	0	0	1	0	1	*P21*	0	1	1	0	1	0
M22	1	0	1	0	0	1	*P22*	0	1	0	1	1	0
M23	1	1	0	0	0	1	*P23*	0	0	1	1	1	0
M24	1	0	0	0	1	1	*P24*	0	1	1	1	0	0
M25	0	1	0	0	1	1	*P25*	1	0	1	1	0	0

Note. For inference questions, participants predicted the value of the feature Horns, and for classification questions, participants predicted the value of the category labels. Category-accordance responses are responses consistent with these values. High, Medium, and Low = three levels of feature match (high, medium, and low levels of feature match, respectively). The value 1 = long horns, round head, dotted body, eight legs, and short tail. The value 0 = short horns, angular head, striped body, four legs, and long tail. M = Monek; P = Plaple.

the classification condition. These predicted outcomes would result in an interaction between question type and degree of feature match at the low level of feature match.

Method

Participants. One hundred members of the Columbia University community participated in the experiment (25 per condition). Seventeen participants were given course credit, and 83 participants were paid. One participant in the inference-shape condition was excluded from the analyses because of a failure to follow the instructions.

Materials. The stimuli for Experiment 3 were identical to those used in Experiment 2, except for one key modification. In the inference-type condition and in the inference-shape condition, the horns feature and the category labels were switched (Table 3) so that the two types of horns divided the sample stimuli in the inference-type and inference-shape conditions (Figure 7). That is, on the sheet used in these conditions, the stimuli with long horns were placed on the left side of the sheet; the stimuli with short horns were placed on the right side of the sheet. Like other features (i.e., head, body, legs, tail), there were four moneks and one plaple on the left side of the sheet; there were four plaples and one monek on the right side of the sheet. In this manner, these two conditions were equivalent to the classification condition in which the two category labels—monek and plaple—divided the sample stimuli (Figure 5). The sample sheet used in the inference-picture condition was produced from the sample sheet given in the inference-type condition by removing the category label from each sample stimulus and attaching actual drawings of wings to the body of the stimulus (Figure 8). The materials used in the classification condition were identical to those used in the classification condition of Experiment 2.

Procedure. The basic procedure of the experiment involved answering 30 questions provided in a booklet (either 30 predictions of the category labels or 30 predictions of the feature values of horns, depending on the condition).

Participants were randomly assigned to one of four conditions: classification, inference-type, inference-shape, and inference-picture. In the classification condition, participants predicted the category label of each stimulus, as in the classification condition of Experiment 2. In the inference-type condition and the inference-shape condition, participants predicted the value of horns when the values of the other four features and the label were shown. In the inference-picture condition, participants predicted the value of horns when the values of the other five features including wings were presented.

Design. The design of the experiment was identical to that of Experiment 2.

Results

The major results of Experiment 3 are summarized in Figure 9. The data were analyzed in the same way as in Experiment 2. To test the difference between inference and classification, we first used a 2 (label–type: classification vs. inference-type) $\times$ 3 (feature match: high, medium, and low) ANOVA. This analysis showed that the overall difference between the classification condition ($M = 0.61$) and the inference-type condition ($M = 0.64$) was not significant, $F(1, 48) < 1.0$, $MSE = 0.04$, $p > 0.1$, though there was a main effect of feature match, $F(2, 96) = 120$, $MSE = 0.04$, $p < 0.01$. As expected, the interaction between label–type and feature match was significant, $F(2, 96) = 7.11$, $MSE = 0.04$, $p < 0.01$. Planned comparisons reveal that this interaction came from the low level of feature–match. Although participants given the classification condition saw about the same amount of feature information as the participants given the inference-type condition, participants given the inference-type condition made more category-accordance responses than did participants given the classification condition at the low level of feature match (Figure 9), $t(48) = 2.91$, $p < 0.01$. Clearly, this result suggests that the presence of category labels affects the way participants make inferences.

The analysis based on individual subjects showed that no participants in the classification condition made category-accordance responses at the low level of feature match more than 75% of the time, whereas 15 out of 25 participants in the classification condition responded consistently with the opposite category at the low level of feature match more than 75% of the time. In contrast, 3 out of 25 participants in the inference-type condition made category-accordance re-

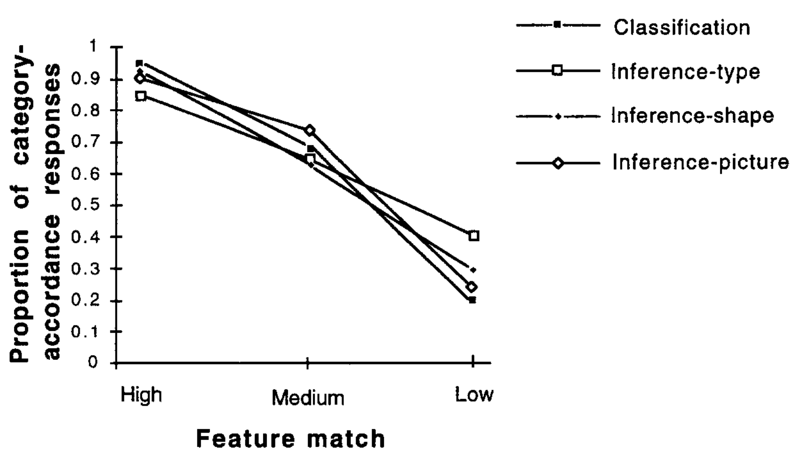

Figure 9. Performance in the classification, inference-type, inference-shape, and inference-picture conditions in Experiment 2.

sponses at the low level of feature match more than 75% of the time, and 8 out of 25 participants responded with the opposite category at the low level of feature match more than 75% of the time. These results indicate that the presence of category labels indeed elevated the level of within-category responses at the low level of feature match in the inference-type condition.

We predicted that the inference-picture condition, in which the labels were replaced with drawings of wings, would yield a similar pattern of results as the classification condition. The data support this prediction. The number of category-accordance responses was roughly equivalent in the two conditions. A 2 (label–type: inference-picture vs. classification-) × 3 (feature match: high, medium, and low) ANOVA showed that the two conditions were not statistically different. Neither the main effect of label–type, $F(1, 48) < 1.0$, $MSE = 0.02$, $p > 0.10$, nor the interaction between label–type and feature match, $F(2, 96) < 1.0$, $MSE = 0.04$, $p > 0.10$, reached significance; though there was the main effect of feature match, $F(2, 96) = 181$, $MSE = 0.03$, $p < 0.01$. Taken together, these results provide strong support for the idea that inference and classification differ because category labels affect feature inferences in a way that category features do not.

We investigated further the impact of category labels with a 3 (label–type: inference-type, inference-shape, and inference-picture) × 3 (feature match: high, medium, and low) ANOVA. In the three conditions, the two labels were described as "two types of bugs" (inference-type), "two shapes of wings" (inference-shape), or they were removed from the sample sheet and drawings of wings were added to each sample stimulus (inference-picture). This analysis revealed that the main effect of label–type was not significant, $F(2, 71) < 1.0$, $MSE = 0.04$, $p > 0.1$, though the main effect of feature match was significant, $F(2, 142) = 161$, $MSE = 0.04$, $p < 0.01$. As expected, there was an interaction between feature match and label–type, $F(4, 142) = 3.31$, $MSE = 0.04$, $p < 0.05$.

As in the previous analysis, the results from planned comparisons support our basic claim. The data indicate that feature match information affects inference depending on the extent to which the two labels convey information about category membership although the overall difference was only marginally significant, $F(2, 71) = 2.56$, $MSE = 0.06$, $0.05 < p < 0.10$. The largest difference was observed between participants given the inference-picture condition and those given the inference-type condition, a difference that was marginally significant, $t(48) = 2.34$, $0.05 < p < 0.10$ (Bonferroni). There was also a tendency for participants given the inference-type condition to make more category-accordance responses than participants given the inference-shape condition at the low level of feature match, but this difference was not significant, $t(47) = 1.32$, $p > 0.10$.

In this experiment, six of the stimuli—*M15, M20, M25, P15, P20,* and *P25* (see Table 4)—have category labels that are consistent with the opposite category. For example, the eight stimuli at the high level of feature match (*M11–M14* and *P11–P14*) have the values in accordance with the corresponding category (1 = monek and 0 = plaple, respectively), but the two stimuli *M15* and *P15* agree with the value of the opposite category (0 = plaple and 1 = monek, respectively). Because the category labels for these stimuli can lead to the response contrary to the corresponding category, the number of category-accordance responses is expected to be higher if the data are analyzed without these stimuli, and the number of category-accordance responses should be lower if the data are analyzed only for these stimuli. This tendency should be strongest in the inference-type condition, smaller in the inference-shape condition, and absent in the inference-picture condition if category labels are indeed the primary factor that influences the degree of category-accordance responding in this task.

Our data generally agree with this prediction. We calculated the mean level of category-accordance responding for the six stimuli (i.e., *M15, M20, M25, P15, P20,* and *P25*) and the mean performance for the remaining stimuli and then calculated the difference in mean performance for each participant. This difference in mean performance was largest in the inference-type condition (0.13), followed by the inference-type condition (0.5) and the inference-picture condition (0.3), though these values are not statistically different.

Discussion

In this study, the arrangement of the sample stimuli was modified in the three inference conditions (i.e., inference-type, inference-picture, and inference-shape) in order to make the inference questions comparable to the classification questions in their use of the sample sheet. In these conditions, participants in the three inference conditions predicted the feature value of horns when the sample sheet was divided by the two types of horns. Likewise, participants in the classification condition predicted the value of the category labels when the sample sheet was divided by the two labels. In this manner, to respond to inference or classification questions, participants needed to look at both sides of the sample sheet.

Unlike Experiment 2, the number of category-accordance responses decreased considerably in the three inference conditions. However, the basic patterns of the results were unchanged. The results suggest that the number of category-accordance responses depends on the degree to which the category label conveys information about category membership (e.g., "two types of bugs"). The number of category-accordance responses was higher in the inference-type condition than in the inference-picture condition at the low level of feature match. Though not statistically significant, a similar trend appeared between the inference-type condition and the inference-shape condition. Furthermore, the response patterns observed in the inference-picture condition were almost identical to those observed in the classification condition, indicating that without category labels inference and classification are indeed indistinguishable. These results clearly favor the view that category labels have a distinct status from category features, and inference is distinguished from classification because category labels are different from other features (see Yamauchi & Markman, 1998, in press).

The present results also suggest that the focus on a single target category evident in Experiments 1 and 2 was also a result of the arrangement of stimuli in the sample sheet. Compared with the results of Experiment 2, the proportion of category-accordance responses in the inference-type condition was considerably lower in this experiment than in Experiment 2. Similar trends appeared in the inference-shape conditions and in the inference-picture conditions. Clearly, the spatial division of the sample sheet promoted category-accordance responses. Taken together, the results of Experiment 3 indicate that the focus on a single target category is enhanced by (a) a common label assigned to different items, (b) the grouping information implied by the two labels, and (c) grouping information enforced in the sample sheet.

So far we have observed that participants in the inference-type condition made category-accordance responses more often than did participants in the inference-shape condition. Thus, the present studies provide strong evidence that category labels and category features are treated differently by the inference process. What makes labels and features different? In the studies so far, the labels and features have had very different semantic meanings, and so it is difficult to determine the key characteristics that make something a category label. Experiment 4 focuses on one hypothesis about the nature of category labels.

Experiment 4

One of the key distinctions between category labels and category features may derive from the relations that bind the property to the category. As discussed earlier, category membership specifies a class inclusion relation and relates the whole object to the category (e.g., a *dog* is an animal). In contrast, a category feature denotes a characteristic or function of an object and thus relates a part of an object to the category with a possession relation (e.g., a dog has a *paw*). This analysis suggests that if a label refers to a whole object, then it may be treated more like a category label than a feature. In contrast, if a label refers to a part of an object, then it may be treated more like a feature than a category label.

We tested this possibility by taking the same property and describing it either as referring to a whole object (i.e., the *whole* condition) or to a part of an object (i.e., the *part* condition). In the whole condition, the instructions state that the bugs that *are* poisonous (nonpoisonous) are named monek (plaple). The instructions in the part condition indicated that the bugs tagged with monek (plaple) *have* poisonous (nonpoisonous) needles (the italics type in this example is not in the instructions). In this manner, the same feature (i.e., poisonous) is described by a class inclusion relation (i.e., is_a) in the whole condition but a partonomic relation (i.e., has_a) in the part condition (see Appendix for the instructions).

Having a poisonous needle and being a poisonous bug mean almost the same thing. That is, when we say that a bug has a poisonous needle, we are typically trying to communicate that the bug is poisonous. Thus, this manipulation should not yield a difference between conditions if the distinction between category labels and category features arises from some aspect of the meaning associated with the labels. In contrast, if the distinction is enforced by language and is embedded in the whole–part relations involved in creating the category, then more category-accordance responses should appear for participants in the whole condition than for participants in the part condition. This tendency should be particularly noticeable at the low level of feature match.

Method

Participants. Fifty-two members of the Columbia University community participated in the experiment. Eleven participants received course credit, and 41 received a payment. Two participants were excluded from the analyses because they did not understand the instructions or did not finish the experiment. In total, the data from 50 participants (25 per condition) were analyzed.

Materials. The stimuli for Experiment 4 were identical to those used in the inference-type condition in Experiment 2.

Procedure. The procedure of the experiment involved answering 30 questions provided in a booklet. Prior to the experiment, participants were randomly assigned either to the whole condition or to the part condition. The two conditions were identical except for their instructions. The instructions in the whole condition indicated that the bugs that are poisonous are named monek and that the bugs that are not poisonous are named plaple. The instructions in the part condition stated that the bugs that are tagged with monek have poisonous needles and the bugs that are tagged with plaple have nonpoisonous needles (see the Appendix for the instructions). The rest of the instructions were identical. The procedures used in the two conditions were identical to those used in the inference-type condition of Experiment 2.

Design. The design of the experiment was a 2 (label–type: whole and part) $\times$ 3 (feature match: high, medium, and low) mixed factorial. Label–type was a between-subjects factor, and feature match was a within-subjects factor.

Results

The main results of this experiment are shown in Figure 10. These results were analyzed with a 2 (label type: whole and part) $\times$ 3 (feature match: high, medium, and low) ANOVA. We expected that participants in the whole condition would make more category-accordance responses than would participants in the part condition at the low level of feature match, if category labels and category features are distinguished by the use of holistic and partonomic expressions.

An ANOVA reveals that there were main effects of the whole–part condition, $F(1, 48) = 5.93$, $MSE = 0.11$, $p < 0.05$; and feature match, $F(2, 96) = 7.53$, $MSE = 0.03$, $p < 0.01$. Participants in the whole condition ($M = 0.84$) made significantly more category-accordance responses than did participants in the part condition ($M = 0.71$). The interaction between the whole–part conditions and feature match was not significant, $F(2, 96) = 2.0$, $MSE = 0.03$, $p > 0.1$.

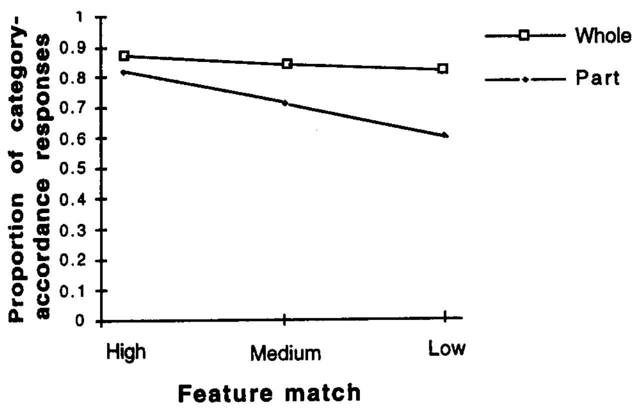

Figure 10. Performance in the whole and part conditions in Experiment 4.

Although the ANOVA does not show an overall interaction between the two factors, we conducted planned comparisons because our main concern was the impact of the whole–part condition within each of the three levels of feature match. This analysis suggests that participants in the part condition made fewer category-accordance responses than did participants in the whole condition at the low level of feature match, $t(48) = 3.0$, $p < 0.05$ (Bonferroni). The two conditions were not statistically different at the other levels of feature–match.

The difference between the whole and the part conditions was also evident in the subject analyses. In total, 17 out of 25 participants in the whole condition responded with the corresponding category more than 75% of the time at the low level of feature match, and no participants in the same condition made responses in accordance with the opposite category more than 75% of the time. In contrast, 10 out of 25 participants in the part condition made within-category responses more than 75% of the time at the low level of feature match, and 3 out of 25 participants in the same condition responded in accordance with the opposite category more than 75% of the time. Thus, participants in the whole condition were responding predominantly with respect to category labels, but participants in the part condition appear to have used both category labels and feature-match information.

To investigate further the actual strategy used during judgment, we conducted a supplementary study using 20 participants (10 participants in each condition) in the same experimental setting. In this study, we interviewed the participants about the way they made responses. In addition, we provided the participants with a short questionnaire asking how they used the two labels and the sample sheet to make judgments.

The basic response patterns observed in these participants were analogous to those reported in the *Results* section. In the whole condition, 9 out of 10 participants explicitly indicated that they used the two labels as a guide to divide the sample sheet into two groups. They suggested that, given a stimulus with one label, they examined only the single group that was consistent with the label. In the part condition, 5 out of 10 participants gave a similar explanation when asked how they used the two labels and the sample sheet. For the remaining 5 participants in the part condition, 3 participants indicated that they simply looked for similarity match between a test stimulus and all the sample stimuli because they were not aware of any division between the groups. One participant stated that his decision was totally intuitive because the sample sheet did not exhibit any meaningful information, and the remaining participant said that she used the sample sheet to find a perfect match between a test stimulus and a sample stimulus, treating the two labels as just another feature. The results of the supplemental study are consistent with the response data.

Discussion

In Experiment 4, we held the meaning of the labels constant and varied whether the relation specified by the label referred to the whole object or just a part. In this setting, we predicted that participants in the whole condition would produce more category-accordance responses than would those in the part condition. The results were consistent with this prediction. On average, participants in the whole condition made within-category responses more often than did participants in the part condition. The two conditions differed most at the low level of feature match.

As the subject analyses imply, the two types of reference did not lead to dichotomous decision strategies—one using category membership and the other based on overall feature match. Rather, modifying the description of the two labels reduced the number of people who divided the stimuli into groups, allowing participants to adopt a variety of different strategies to make inferences. The present results, however, make it clear that the notion of category membership and category feature stems in part from whether the label is associated with a whole object or a part of an object.

How does the distinction between category labels and category features actually arise? One possibility is that there is a linguistic bias to treat a label that is assigned to an object as referring to a taxonomic category. E. M. Markman (1987, 1989, 1991) argued that when children hear a novel label, they assume that it refers to a whole object. Children further assume that objects that share the label are bound by a taxonomic relationship rather than by a thematic relationship.[2] In a similar vein, Waxman and her colleagues (Waxman & Markow, 1995, 1998; Waxman, Senghas, & Benveniste, 1997; Waxman et al., 1991) have demonstrated that children as young as 12 months have tacit knowledge that count nouns name basic level categories and that adjectives name properties of basic level categories. In the context of the present studies, these findings suggest that participants may have assumed that stimuli that appeared along with a label belonged to the same category and thus were likely to share other features.

[2] There is some evidence that children extend count nouns to objects with the same shape as a given label rather than strictly to objects from the same taxonomic category (e.g., Imai, Gentner, & Uchida, 1994; Landau, Smith, & Jones, 1988).

Although this explanation is broadly consistent with our findings, it does not account for the details of the present results. All of the stimuli in the two conditions carried one of the two labels (monek or plaple). Thus the distinction observed in the two conditions was not due to the presence or absence of labels per se (E. M. Markman, 1989; Waxman & Markow, 1995, 1998; Waxman, et al. 1991). Moreover, the labels in the two conditions were not contrasted by the presence of count nouns or adjectives. Instead, the labels were associated with a whole object having a particular attribute or with part of the object having that attribute (e.g., a poisonous bug vs. a poisonous needle). Thus the difference between the whole and part conditions rests on the relation linking the label to the category rather than the syntactic category of the label.

General Discussion

Summary of the Results

Category-based inference involves a number of strategies in addition to similarity matching. For one, people make use of category labels to guide their attention. Unlike classification, which has been the main focus of categorization research, predictive inference is governed not only by feature–match information but also by class inclusion relationships specified by category labels. In contrast to ordinary features, category membership guides participants to pay attention to features within a target category.

In Experiment 1, we examined the influence of category labels and category features on participants' performance on classification and inference questions. In this setting, when similarity and category membership were placed in opposition, participants were more likely to select the feature value that was consistent with the category label given to the test item rather than the value consistent with other stimuli that shared matching features with the test item. The distinction between category labels and category features was investigated further by modifying how the category label was described in the instructions.

In Experiment 2, changing the label from a category label referring to two types of bugs in the inference-type condition to a label referring to a feature in the inference-shape condition or to an actual picture of a feature in the inference-picture condition led to a decrease in the number of category-accordance responses. This disparity was also present in Experiment 3, even when the category labels had a structure equivalent to that possessed by other features.

In Experiment 4, we explored the element that determines category membership. We associated a label with the feature "poisonous" but varied whether the label referred to a whole object or to just a part of an object. In the whole condition, the labels referred to poisonous or nonpoisonous bugs, and in the part condition, the labels referred to whether or not the bug had a poisonous needle. Participants in the whole condition were more likely to make responses consistent with a target category than were participants in the part condition. Taken together, these results provide strong support for the view that the information about category membership molds the way people infer a characteristic of an object. When the category membership of an object is known, people pay particular attention to the feature value most prevalent in the members of the corresponding category. Otherwise, they fall back on a strategy like inferring the feature value possessed by other objects with matching features. Thus, inference and classification differ precisely because category labels are treated differently from other features.

Interpretation of the Results

Several points need to be addressed in order to clarify the interpretation of the results of these studies. One potential problem with the present results comes from the assumption that the involvement of category membership in inductive judgment is revealed by the degree to which participants respond with a single target category (i.e., category-accordance responses). The problem with this assumption is the difficulty of distinguishing the impact of attention weight and category membership. In the four experiments, the category labels were denoted by the words "monek" and "plaple" on the sample sheet, and other features were depicted by specific figures. The two labels might have been more salient than other features so that participants allocated more attention to the labels than to other features. A higher proportion of category-accordance responses thus appeared at the low level of feature match given inference questions than classification questions.

This account does not explain why category labels are more salient than category features, or why the weight given to the two labels shifts systematically if they are characterized as "two types of bugs," "two shapes of wings," or if the labels are substituted by actual drawings of wings. Moreover, we observed different response patterns in the whole and part conditions of Experiment 4 even when the two labels were associated with the same attribute "poisonous" or "nonpoisonous." It is difficult to see why a description related to wholes of objects leads to greater attention weight than does a description related to local parts.

Similarly, the results of the present studies cannot be explained solely by a general judgment strategy balancing between accuracy and effort (Ross & Murphy, 1996). On this view, participants opted for focusing on the corresponding category in inference, because it provides a reasonable level of accuracy with a moderate amount of cognitive effort. Category-accordance responding was reduced when the characteristics of category labels were changed only in the instructions (Experiments 2, 3, and 4). It is not clear why the accuracy–effort balance shifts depending on whether the labels were described as class-inclusion or partonomic statements.

We argue that the focus on a single target category arises because category membership goes beyond simply specifying another feature of an object. A substantial amount of research on category coherence suggests that people's categories are marked by deep theoretical knowledge (Keil, 1987, 1989; Murphy & Medin, 1985; Wisniewski & Medin, 1994; see also Malt et al., 1999; Rips, 1989; Smith, 1989).

Thus, category labels can be viewed as reliable pointers to systematic knowledge structures that may provide a basis for making predictions about unknown features.

Finally, the present results are consistent with other research that local and global aspects of objects are treated differently. Barsalou's (1983) work on ad hoc categories has suggested that a set of objects that shares a property (such as being valuable) can be treated as a category by attaching a label to the whole object (e.g., a thing to take out of the house in the event of a fire). In addition, there is research on memory systems that suggests that different systems may mediate processing of local and global aspects of objects (Yamauchi, Cooper, Hilton, Szerlip, & Schacter, 2000). One plausible explanation of this view is that one type of memory for objects is organized around whole objects, and another type of memory for objects is organized with their features such as their functions and semantic associations (Cooper & Schacter, 1992; Cooper, Schacter, Ballesteros, & Moore, 1992; Tulving & Schacter, 1990). This distinction in memory may in turn link to the difference between category labels and category features. Although this link is more speculative, it does provide a potential avenue for relating research on categorization to research on memory more broadly.

Implications of the Results

Two major assumptions of categorization research are that categories reduce uncertainty about unknown events and objects, and that categories facilitate the prediction of unknown features of objects (Anderson, 1990, 1991; Corter & Gluck, 1992; Rosch, 1978; Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Smith, 1994). Our studies indicate that category membership is indeed a key determinant of inductive judgment and that category labels are not simply another feature on a par with other category features. These results argue against an assumption made in models developed by Anderson (1990, 1991) and Sloman (1993). In both of these models, category labels, which convey information about category membership, are assumed to be another feature to be predicted, and the processes of inference (i.e., predicting category features) and classification (i.e., predicting category labels) are identical.

The present results suggest that inference and classification are indeed equivalent, if information about category membership is absent. Both tasks are strongly influenced by feature matching. However, if category labels are present during inference, then performance on classification and inference tasks diverges. Although feature inheritance from superordinate categories may be rare (Sloman, 1998), the impact of category membership is not limited to feature inheritance. Category membership is instrumental in guiding people to focus on specific categories. In this regard, the present results are consistent with a number of studies investigating inductive judgment (Gelman & Markman, 1986; Malt et al., 1995; Murphy & Ross, 1994; Ross & Murphy, 1996). Our work extends previous findings by demonstrating that category membership is not only determined by feature overlap between stimuli (see also Malt et al., 1999; Rips, 1989; Smith, 1989) but also arises from the presence of a common label assigned to different items and labels that refer to whole objects.

References

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.

Anderson, J. R., & Fincham, J. M. (1996). Categorization and sensitivity to correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 259–277.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11,* 211–227.

Cooper, L. A., & Schacter, D. L. (1992). Dissociations between structural and episodic representations of visual objects. *Current Directions in Psychological Science, 1,* 141–146.

Cooper, L. A., Schacter, D. L., Ballesteros, S., & Moore, C. (1992). Priming and recognition of transformed three-dimensional objects: Effects of size and reflection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 43–57.

Corter, J. E., & Gluck, M. A. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin, 111,* 291–303.

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

Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development, 9,* 45–75.

Keil, F. C. (1987). Conceptual development and category structure. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 175–200). New York: Cambridge University Press.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

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

Landau, B., Smith, L., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3,* 299–321.

Malt, B. C., 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.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language, 40,* 230–262.

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

Markman, E. M. (1987). How children constrain the possible meanings of words. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 255–287). Cambridge, MA: Cambridge University Press.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction.* Cambridge, MA: MIT Press.

Markman, E. M. (1991). The whole-objects, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and* *thought: Interrelations in development* (pp. 72–106). Cambridge, MA: Cambridge University Press.

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

Medin, D. L., & 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.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception.* Cambridge, MA: Harvard University Press.

Murphy, G. L. (1991). Parts in object concepts: Experiments with artificial categories. *Memory & Cognition, 19,* 423–438.

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.

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

Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 227–270). Cambridge: MA: MIT Press.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & N. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). Cambridge, England: Cambridge University Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.

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

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25,* 231–280.

Sloman, S. A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology, 35,* 1–33.

Smith, E. E. (1989). Concepts and induction. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 501–526). Cambridge, MA: MIT Press.

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

Thomas, R. D. (1998). Learning correlations in categorization tasks using large, ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 119–143.

Tulving, E., & Schacter, D. L. (1990, January). Priming and human memory systems. *Science, 247,* 301–306.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113,* 169–191.

Waxman, S. R., & Markow, D. B. (1995). Words as invitation to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29,* 257–302.

Waxman, S. R., & Markow, D. B. (1998). Object properties and object kind: Twenty-one-month-old infants' extension of novel adjectives. *Child Development, 69,* 1313–1329.

Waxman, S. R., Senghas, A., & Benveniste, S. (1997). A cross-linguistic examination of the noun-category bias: Its existence and specificity in French and Spanish-speaking preschool-aged children. *Cognitive Psychology, 32,* 183–218.

Waxman, S. R., Shipley, E. F., & Shepperson, B. (1991). Establishing new subcategories: The role of category labels and existing knowledge. *Child Development, 62,* 127–138.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science, 18,* 221–281.

Yamauchi, T., Cooper, L. A., Hilton, H. J., Szerlip, N., & Schacter, D. L. (2000). *Global axis-based structure and local component-based structure support separable systems for the representation and retrieval of information about visual objects.* Manuscript in preparation.

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 and Language, 39,* 124–148.

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

(Appendix follows)

Appendix

Excerpts From the Instructions Used in Experiments 1, 2, and 4

An Excerpt From the Instructions Used in Experiment 1

The sheet that you have just received shows samples of two types of imaginary bugs—"moneks" and "plaples." The moneks are on the left side of the page and the plaples are on the right side of the page. Based on these samples, please answer the questions described in this booklet as accurately as you can.

An Excerpt From the Instructions Used in the Inference-Shape Condition of Experiment 2

The sheet that you have just received shows samples of imaginary bugs, which exhibit different characteristics in their appearances of six parts—horns, head, body, legs, tail and wings. Because the wings of the bugs are folded on their back, we were not able to show them, so that we specified them with two names—monek and plaple, which roughly stand for two different shapes of wings. In other words, the bugs tagged with monek right above their horns have monek-shaped wings, and the bugs tagged with plaple right above their horns have plaple-shaped wings.

Based on these samples, please answer the questions described in this booklet as accurately as you can.

An Excerpt From the Instructions Used in the Inference-Picture Condition of Experiment 2

The sheet that you have just received shows samples of imaginary bugs, which exhibit different characteristics in their appearances of six parts—horns, head, body, legs, tail and wings. Based on these samples, please answer the questions described in this booklet as accurately as you can.

An Except From the Instructions for the Whole Condition of Experiment 4

The sheet that you have just received shows samples of imaginary bugs, which exhibit different characteristics in their appearances of five parts—horns, head, body, legs, and tail. In addition, the bugs that are poisonous are named "monek," and the bugs that are not poisonous are named "plaple."

Based on these samples, please answer the questions described in this booklet as accurately as you can.

An Excerpt From the Instructions for the Part Condition of Experiment 4

The sheet that you have just received shows samples of imaginary bugs, which exhibit different characteristics in their appearances of six parts—horns, head, body, legs, tail and needle. Because the needle of a bug is hidden in its mouth, so that we specified it with one of two names—monek or plaple, which stand for a poisonous or non-poisonous needle, respectively. In other words, a bug tagged with monek right above its horns has a poisonous needle, and a bug tagged with plaple right above its horns has a non-poisonous needle.

Based on these samples, please answer the questions described in this booklet as accurately as you can.

Received January 25, 1999
Revision received October 28, 1999
Accepted October 28, 1999 ■