

Category labels versus feature labels: Category labels polarize inferential predictions

TAKASHI YAMAUCHI AND NA-YUNG YU
Texas A&M University, College Station, Texas

What makes category labels different from feature labels in predictive inference? This study suggests that category labels tend to make inductive reasoning polarized and homogeneous. In two experiments, participants were shown two schematic pictures of insects side by side and predicted the value of a hidden feature of one insect on the basis of the other insect. Arbitrary verbal labels were shown above the two pictures, and the meanings of the labels were manipulated in the instructions. In one condition, the labels represented the category membership of the insects, and in the other conditions, the same labels represented attributes of the insects. When the labels represented category membership, participants' responses became substantially polarized and homogeneous, indicating that the mere reference to category membership can modify reasoning processes.

Category labels are generally linked to the identity of things, whereas feature labels are associated with the possession of properties. Consider the following two sentences, based on the Preface to Walt Whitman's *Leaves of Grass* (1855/1985): (1) The United States are essentially the greatest *poem*; (2) The United States have the fullest *poetical* nature. These sentences roughly mean the same thing, but they evoke different kinds of implications. Gelman and Heyman (1999) suggested that categorical noun labels (e.g., "poem" in Sentence 1), unlike feature labels (e.g., "poetical" in Sentence 2), evoke a sense of immutability, endurance, and centrality linked to the identity of an object, even in 5-year-old children (see also Walton & Banaji, 2004; Yamauchi, 2005).

The distinction between categorical noun labels and feature labels in inductive generalization has been well documented in developmental research on children's acquisition of new knowledge (e.g., Brown, 1957; Gelman & Heyman, 1999; E. M. Markman & Hutchinson, 1984; Waxman & Booth, 2001; but see also Sloutsky, 2003; Sloutsky & Fisher, 2004). A number of cognitive studies have also demonstrated the importance of category labels in inductive generalization (Clapper & Bower, 2002; Waldmann & Hagmayer, 2006; Yamauchi, 2005, in press; see also Murphy & Medin, 1985; Murphy & Ross, 1994; Rosch, 1978; Ross & Murphy, 1996). However, there is no clear consensus among cognitive scientists on exactly how category labels and feature labels differ. Researchers exploring computational aspects of inductive generalization tend to assume that category labels and feature labels are basically the same thing (e.g., Anderson, 1990). In such studies, category labels may be useful for predicting certain attributes and may attract more attention, but the distinction is regarded as trivial, and the major computational models of inductive generalization make no qualitative distinction

between the two types of labels (Kruschke, 1992; Love, Medin, & Gureckis, 2004; Osherson, Smith, Wilkie, López, & Shafir, 1990; Sloman, 1993; Sloutsky, 2003; Sloutsky & Fisher, 2004; see Heit, 2000, for a review).

In this study, we contrasted category labels and feature labels and investigated how they differ in their influence on predictive inference. Yamauchi, Kohn, and Yu (2007) provided two findings illustrating the distinction between category and feature labels in an inference task. Because the present study is directly related to that of Yamauchi et al., in the next paragraph we describe in detail the results of the previous study.

In Yamauchi et al. (2007), participants were shown a sample stimulus and a test stimulus side by side and predicted the value of a hidden feature of the test stimulus on the basis of the sample stimulus (Figure 1A). Both the sample and test stimuli carried verbal labels, which were blurred, and only became legible to the participant when the cursor was moved over a particular location. The meanings of the labels were manipulated in two conditions. In one (the category condition), the labels (e.g., "monek" in Figure 1A) represented names of two different insect categories; in the other (the attribute condition), the same labels represented names of attributes of the insects (the shapes of wings hidden beneath their bodies). The results showed that when the two labels were associated with category information, they were a reliable guide for predictive inference. For example, when the sample and test stimuli had the same label (Figure 1A), the projection of attributes from sample to test stimuli increased dramatically (i.e., selecting the long antennae ["horns" in the stimuli and instructions] in Figure 1A); when the sample and test stimuli had different labels (Figure 1B), the projection of attributes from the sample to the test stimulus declined dramatically (i.e., selecting the short antennae in Figure 1B); this dramatic

T. Yamauchi, tya@psyc.tamu.edu

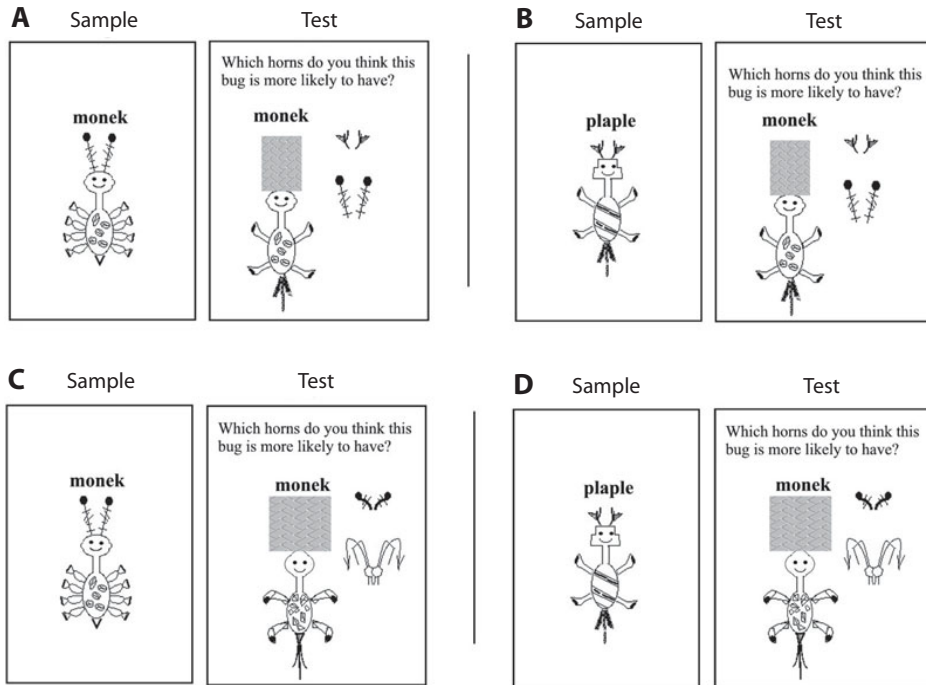


Figure 1. Sample stimulus frames used in Experiments 1 and 2. In the matched stimuli (A and C), the sample and test stimuli had the same labels. In the mismatched stimuli (B and D), the sample and test stimuli had different labels. Half of the test stimuli were produced from the same feature instances used for the sample stimuli (A and B), and the remaining half were produced from feature instances different from those used for the sample stimuli (C and D). Note that the instantiations of features in the sample and test stimuli are identical in panels A and B and different in test C and D.

fluctuation in the projection of feature values, depending on the matched or mismatched labels, is called a *polarity effect* in this article. This tendency was particularly conspicuous when the labels were associated with category membership information.

The focus of the Yamauchi et al. (2007) study was on the impact of categorical labeling on the time course of decision making. In that study, the movement of the computer cursor was tracked every 50 msec in order to examine when and how often participants viewed the hidden labels in each trial. The results indicated that participants tended to view the labels more often and earlier in each trial when category labels were given. In the present study, on the other hand, we compared the inductive potentials of category labels and of a wide variety of feature labels; specifically, the category labels were contrasted with more fundamental and diverse feature labels related to the biological, behavioral, and physical characteristics of imaginary insects. If the distinction between category labels and feature labels is real, the polarity effect that was observed in the previous study should be replicated.

In the present study, we introduced three conditions, in which two labels were characterized as the names of diseases that the imaginary insects carry, foods that they eat, or islands that they live on, in order to assess the polarity and homogeneity created by categorical labeling. Specifically, we measured “polarity scores” by subtracting the

proportion of features selected that were consistent with the sample stimulus when the sample and test stimuli had different labels (Figures 1B and 1D) from the proportion of consistent features selected when the sample and test stimuli had the same labels (Figures 1A and 1C). We also measured the extent to which participants’ responses became similar to each other (i.e., the homogeneity effect). If labels are used as abstract rules, feature information should be discounted, and participants’ response patterns should become highly uniform (the details of measuring the homogeneity effect will be described later, in the Results section of Experiment 1).

As in Yamauchi et al. (2007), participants in the present study were shown schematic pictures of insects side by side and made predictions about a test stimulus on the basis of a sample stimulus (Figure 1A). These schematic insects were produced from combinations of five feature dimensions (antennae = long/short, head = round/angular, torso = dotted/striped, legs = eight/four, tail = long/short) and two verbal labels (“monek”/“plaple”; see Table 1). The test stimulus had one feature missing, and participants were asked to make a judgment about this feature. Two choices were given for the missing feature: One was *consistent* and the other was *inconsistent* with the sample stimulus. For example, in Figure 1A, the choice of long antennae is consistent with the feature in the sample stimulus, whereas the choice of short antennae is not. We

Table 1
The Stimulus Structure in Experiments 1 and 2

	Antennae	Head	Body	Legs	Tail	Labels
Sample "Monek"	1	1	1	1	1	1
Test 1	?	1	1	0	0	1
Test 2	1	1	0	0	?	1
Test 3	1	0	0	?	1	1
Test 4	0	0	?	1	1	1
Test 5	0	?	1	1	0	1
Sample "Plaple"	0	0	0	0	0	0
Test 6	?	0	0	1	1	0
Test 7	0	0	1	1	?	0
Test 8	0	1	1	?	0	0
Test 9	1	1	?	0	0	0
Test 10	1	?	0	0	1	0

Note—"?" stands for the feature dimension that was queried. 1 and 0 represent the values of the feature dimensions: (1, 0) = (long, short antennae), (round, angular head), (dotted, striped body), (8, 4 legs), or (short, long tail). A total of 20 test stimuli were produced from two sets of stimuli, A and B (see Figure 2). The 20 stimuli were shown twice, in the matched and mismatched conditions, yielding 40 trials for each participant.

compared the proportions of making a "consistent choice" when the labels carried category membership or attribute information, to produce a consistency score.

In three attribute conditions, the two labels were characterized as the names of islands on which the insects live, foods that they eat, or diseases that they carry. In this manner, the arbitrary labels were associated with the habitats, sustenance, and biological dispositions of the insects. Because these features are likely to affect the behavior of the insects, we assumed that these attributes were central to the category (see Sloman, Love, & Ahn, 1998, for a discussion of the centrality of category features). In the category condition, the two labels were characterized as the names of categories to which the insects belonged.

To investigate the distinction between category and feature labels, we created 20 test stimuli from two sets of sample stimuli (10 from Set A and 10 from Set B in Figure 2). These sample stimuli (Sets A and B) were related to each other in their abstract feature values but were different in their exact appearances. For example, the "monek" samples in Sets A and B both had long antennae, round faces, dotted torsos, eight legs, and short tails. Similarly, the "plaple" samples in the two sets had short antennae, angular faces, striped torsos, four legs, and long tails. The specific instantiations of the individual features, though, were different. Thus, this variable would increase the variation of test stimuli; however, we think this variation would be discounted when the labels carry category membership information.

Predictions

We think that category labels play a guiding role in feature inference (E. M. Markman, 1989; Yamauchi, 2005; Yamauchi et al., 2007; Yamauchi & Markman, 2000) and promote a reasoning strategy using abstract rules (see Sloman, 1996, and Smith & Sloman, 1994, for the distinction between rule-based and similarity-based reasoning strategies). First, participants' inferential projections should depend primarily on the matching/mismatching status of the labels when the two labels are characterized as the names of categories. That is, consistency scores should rise

sharply when sample and test stimuli have the same labels (Figures 1A and 1C, matched condition) and decline when sample and test stimuli have different labels (Figures 1B and 1D, mismatched condition)—that is, a polarity effect. This tendency should increase dramatically when the two arbitrary labels carry category information.

Second, individual responses made in the category condition should be homogeneous relative to those in the attribute conditions. That is, participants should exhibit a tendency to ignore feature information (e.g., the number of matching and mismatching features) and to use labels as abstract decision rules. As a result, individual responses in the category condition should become similar to each other. To test this idea, we analyzed individual response patterns with cluster analyses (the details of this procedure are explained later, in the Experiment 1 Results section).

EXPERIMENT 1

Method

Participants. A total of 211 undergraduate students participated in this experiment for course credit. The participants were randomly

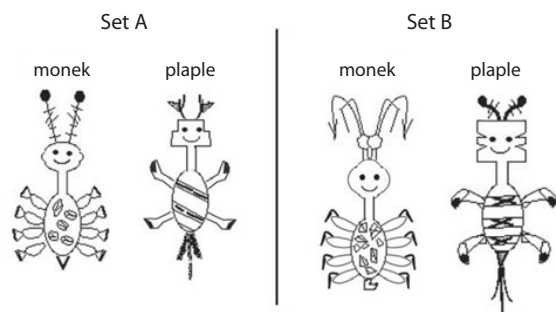


Figure 2. Two sets of sample stimuli. All test stimuli were produced from these sample stimuli. In one version, the two stimuli ("monek" and "plaple") in Set A were used as sample stimuli; in the other version, the sample stimuli were taken from Set B.

assigned to one of four conditions: category ($n = 58$), disease-attribute ($n = 49$), food-attribute ($n = 46$), and island-attribute ($n = 58$).

Materials. All test stimuli had two features consistent with the sample of one category, two features consistent with the sample of the other category, and one feature masked for an inference question. By assigning binary values of 1 or 0 to each feature dimension, the sample stimulus in the “monek” group (“monek” stimuli in Figure 2) can be expressed as (1, 1, 1, 1, 1) = (long antennae, round head, dotted body, eight legs, short tail), and the sample stimulus in the “plaple” group can be expressed as (0, 0, 0, 0, 0) = (short antennae, angular head, striped body, four legs, long tail) (see Table 1). Twenty test stimuli were produced in this manner from the two different instantiations of the sample stimuli.

In one version of the experiment, the two stimuli of Set A were shown as the sample stimuli; in the other version, the two stimuli of Set B were shown as the samples.

Procedure. Participants were shown a pair of sample and test stimuli on a computer screen and predicted one of two feature values of the test stimulus on the basis of the sample stimulus (Figure 1). One choice was consistent with a feature shown in the sample stimulus (e.g., the long antennae in Figure 1A), and the other was inconsistent with that feature. The 20 test stimuli were shown twice. In one case, a test stimulus was paired with a sample stimulus that had the same label (i.e., the *match* condition—Figures 1A and 1C). In the other case, the same test stimulus was paired with a sample stimulus that had a different label (i.e., the *mismatch* condition—Figures 1B and 1D). Each participant received a total of 40 trials; the order of the trials was determined randomly for each participant.

Design. The experiment had a 4 (label condition: category, disease-attribute, food-attribute, island-attribute; between subjects) \times 2 (match status: match, mismatch; within subjects) \times 2 (feature instantiation: same, different; within subjects) \times 2 (stimulus version: 1, 2; between subjects) factorial design. The stimulus version, which was created for counterbalancing, did not interact with the other factors; therefore, this factor was collapsed for subsequent data analyses. *Label condition* was manipulated solely in the instructions (see the Appendix). In the category condition, the two labels were characterized as representing two “types” of bugs. In the three attribute conditions (disease, food, and island), the same two labels were characterized as representing two kinds of diseases that the bugs carry, two kinds of foods that they eat regularly, or two different islands on which they live. *Match status* represents the matched/mismatched status of the labels attached to the sample and test stimuli (e.g., matched labels in Figures 1A and 1C, mismatched label in Figures 1B and 1D). *Feature instantiation* is the correspondence of the feature instances used to depict the sample and test stimuli. Ten of the test stimuli were produced from the *same* feature instances used to depict the sample stimuli (e.g., Set A in Figures 1A and 1B); the remaining 10 were produced from instances *different* from the sample stimuli (e.g., Set B in Figures 1C and 1D). This factor was introduced in order to avoid a ceiling effect while ensuring that inductive judgments made in the four label conditions would vary sufficiently.

Responses consistent with the sample stimulus were coded as “consistent responses” (e.g., selecting the long antennae in Figures 1A and 1C). The polarity effect, which reflects the influence of the matching and mismatching of labels, was measured by subtracting the proportions of consistent responses obtained with the mismatched stimuli from those with the matched stimuli (i.e., the polarity score). Thus, for the actual data analyses, 4 (label condition) \times 2 (feature instantiation) ANOVAs were applied to the polarity scores calculated for individual participants.

Analyses with both participant-based (F_1) and item-based (F_2) ANOVAs were reported in order to draw appropriate interpretations. For the item-based ANOVAs, we combined the two stimulus versions and analyzed the 40 trials averaged across participants. The two independent variables, label condition and feature instantiation, were treated as “between-subjects” (i.e., here, the individual items) factors.

Results

Polarity effect. Overall, the mean polarity score was substantially larger in the category condition than in the attribute conditions (Table 2). There was a main effect of label condition [$F_1(3,207) = 2.74$, $MS_e = 0.24$, $p < .05$, $\eta^2 = .04$; $F_2(3,72) = 5.21$, $MS_e = 0.02$, $p < .01$, $\eta^2 = .18$]. There was no interaction between label condition and feature instantiation [$F_1(3,207) = 2.09$, $MS_e = 0.05$, $p = .10$, $\eta^2 = .03$; $F_2 < 1$].

Pairwise comparisons suggest that the mean polarity score in the category condition was higher than in the food- and island-attribute conditions [category vs. food, $t_1(102) = 2.17$, $p = .03$, $d = 0.43$; $t_2(38) = 2.72$, $p = .01$, $d = 0.86$; category vs. island, $t_1(114) = 2.57$, $p = .01$, $d = 0.65$; $t_2(38) = 3.09$, $p = .004$, $d = 1.23$]. The difference between the category and disease-attribute conditions was significant only in the item-based analysis [$t_1(105) = 1.94$, $p = .05$, $d = 0.38$; $t_2(38) = 2.70$, $p = .01$, $d = 0.85$]. A comparison between the category condition and an aggregate of all three attribute conditions showed that the mean polarity score in the category condition was considerably larger than in the combined attribute conditions [$t_1(209) = 2.85$, $p = .04$, $d = 0.45$; $t_2(38) = 2.98$, $p = .005$, $d = 0.94$].

Common responses: A cluster analysis. We used a hierarchical cluster analysis to examine the extent to which individual responses obtained from each participant were similar to each other (i.e., a homogeneity effect). In this analysis, every individual response was transcribed to a vector of 40 dimensions (with each dimension representing a response score of 1 or 0 obtained from an individual trial and with N_i response vectors in each label condition, where N_i represents the number of participants in the i th label condition). We measured the proximity distances of all pairs of individual response vectors in each label condition and applied a hierarchical cluster analysis to these response vectors (see Johansen & Kruschke, 2005, for a similar analysis). The logic behind this analysis was that if individual response patterns in a given label condition were homogeneous, these response vectors should be highly similar to each other, resulting in a large cluster of response vectors.

To create a hierarchical tree, we used the unweighted average distance method, which measures the average distance between all pairs of cases in two clusters. The proximity distance was measured by the city-block metric because that method is easy to interpret. For example, a proximity distance of 12 in the city-block metric suggests that 2 participants made different responses in 12

Table 2
A Summary of the Results From Experiment 1

	Polarity Score		
	Match	Mismatch	(Match – Mismatch)
Category	.69	.31	.38
Disease-attribute	.64	.40	.24
Food-attribute	.62	.39	.23
Island-attribute	.57	.36	.21
All attribute conditions	.61	.38	.23

Note—These numbers represent the mean consistency scores obtained in each label condition.

and identical responses in 28 of 40 trials. In this manner, the value from the city-block metric corresponds to the number of trials with different responses.

Overall, the proximity distances observed in the category condition were significantly smaller than those observed in the other attribute conditions (Wilcoxon ranked sum tests: $z_s > 8.7$, $ps < .001$; see Table 3). This suggests that the individual response patterns obtained in the category condition were highly similar to each other, as compared with those in the attribute conditions. Given a proximity distance of 12, 43% of the participants in the category condition were included in the largest cluster, indicating that these participants made identical responses on an average of 28 out of the 40 trials (i.e., 70%). Given the same proximity distance, the percentages of participants in the largest cluster were only 29%, 24%, and 24% for the disease-, food-, and island-attribute conditions, respectively.

Was the homogeneity effect in the category condition linked to a rule-based strategy prompted by the category labels (see Sloman, 1996; Smith & Sloman, 1994)? To explore this idea, we counted the number of participants whose average polarity scores were .5 or above and the number whose average polarity scores were 0 or below. If a rule-based strategy was adopted, the feature information would be discounted, and very high polarity scores would result. In contrast, if a similarity-based strategy was adopted, labels would be discounted, and very low polarity scores would emerge. If the labels were ignored completely, the polarity score should be about 0, because the sample and test stimuli all had two matching features and two mismatching features.

This additional analysis showed that 43.1% of the participants in the category condition predominantly adopted a rule-like strategy (their polarity scores were at least .5), whereas 26.5%, 23.9%, and 22.4% in the disease-, food-, and island-attribute conditions, respectively, followed such a strategy (category vs. disease-attribute, $z = 1.58$, $p = .11$; category vs. food-attribute, $z = 1.84$, $p = .07$; category vs. island-attribute, $z = 2.18$, $p = .03$; category vs. all attribute conditions combined, $z = 2.52$, $p = .01$). In contrast, 20.7% of the participants in the category condition

employed a similarity-like strategy (polarity scores ≤ 0), whereas 40.8%, 32.6%, and 37.9% in the disease-, food-, and island-attribute conditions, respectively, adopted such a strategy (category vs. disease-attribute, $z = 2.05$, $p = .04$; category vs. food-attribute, $z = 1.15$, $p = .25$; category vs. island-attribute, $z = 1.84$, $p = .07$; category vs. all attribute conditions combined, $z = 2.13$, $p = .03$). These results favor the view that the category and attribute conditions differed in terms of the degree to which rule- and similarity-based strategies were adopted.

To test this idea further, we compared the skewness of the distributions of the polarity scores in each label condition. Because the polarity scores depend on the overall consistency scores (i.e., the proportion of times that feature values consistent with the sample stimuli were selected), it is important to compare the distributions of the polarity scores independently of the overall levels of consistency scores. A skewness score indicates the degree of asymmetry of a distribution. With a score of 0, the distribution is said to be symmetric. With a large positive skew score, the distribution is asymmetric with most data placed around the left side of the distribution, with relatively few extreme data points placed around the right side. If a single strategy is predominantly employed (e.g., a rule-based strategy), the distribution should be highly symmetric. Because miscellaneous factors should influence participants' performance, the polarity scores under a uniform (e.g., rule-based) strategy should be roughly normally distributed. In contrast, if the two reasoning strategies, rule- and similarity-based, were employed to different degrees, the distribution of the polarity scores would be relatively asymmetric. For example, if 70% of the participants used a similarity-based process and 30% of the participants used a rule-based process, many low polarity scores should arise from the similarity-based process and a few high polarity scores from the rule-based process, yielding a highly skewed distribution. This analysis showed that the distribution from the category condition was fairly symmetric (skewness score = 0.08, 95% confidence interval [CI] = [-0.34, 0.51]). In contrast, the distributions from the attribute conditions were skewed considerably (disease-attribute = 0.78, 95% CI = [0.28, 1.32]; food-attribute = 0.69, 95% CI = [0.22, 1.17]; island-attribute = 0.75, 95% CI = [0.31, 1.21]).¹ This extra analysis suggests that participants in the category condition were more likely to follow a single, rule-based strategy, whereas those in the attribute conditions primarily used a similarity-based strategy, with some use of a rule-based strategy.

Discussion

In the attribute conditions, the two arbitrary labels were associated with three basic characteristics of insects—with diseases, foods, and locations connected with them; in the category condition, the same labels were associated with the category membership of the insects. These labels were characterized solely in the instructions, which differed only in a few words across the four conditions. Even with these subtle manipulations, participants' responses became substantially polarized and homogeneous when the labels carried category information.

Table 3
A Summary of the Cluster Analysis in Experiment 1

	Proximity Distance		Average Proximity Distance	z Score	p Value
	8	12			
Category	36%	43%	16.60		
Disease	22%	29%	18.70	9.90	<.001
Food	17%	24%	18.75	8.75	<.001
Island	16%	24%	18.83	10.67	<.001

Note—The left columns show the percentages of participants who formed the largest cluster at a given level of proximity distance. For example, given a proximity distance of 8, 36% of the participants in the category condition were grouped together. This means that these participants made different responses on an average of 8 out of 40 trials. The proximity distance was measured by the city-block metric. "Average proximity distance" represents the average distance of all pairs of the participants in a given label condition. The z scores and p values were obtained from Wilcoxon ranked sum tests comparing the proximity distances of all pairs of participants in the category condition with the proximity distances obtained in each of the other attribute conditions.

One may argue that these results simply reflect how relevant these labels were in predicting the physical characteristics of the insects. For example, a label representing “food” might have had little to do with the body parts of the imaginary insects. As a result, the labels in the food-attribute condition were less useful as predictors than the labels in the category condition. The same argument could be applied to the labels in the category condition, which represented the names of two types of insects without specifying exactly what each “type” consisted of. However, because the contextual relevance of attributes is an important factor determining the strength of inductive judgments (Heit & Rubinstein, 1994; Medin, Coley, Storms, & Hayes, 2003; Sloman, 1993), this relevance variable should be taken into account when comparing category and feature labels. Experiment 2 addressed this issue.

EXPERIMENT 2

Are category labels different from feature labels, even when the feature labels are directly related to predictions about body parts? In Experiment 2, we introduced a new, DNA condition, in which the labels were characterized as the names of DNA components, and the instructions explicitly explained that these DNA components affect the physical characteristics of the insects (see the Appendix). In this manner, the two labels in the DNA condition were linked to the prediction of body parts of the insects.

The polarity scores in the DNA condition should be large because of the direct predictability of the labels. The polarity scores in the category condition should be no less pronounced than those in the DNA condition. We think that participants would make an assumption that the “types” of insects are related to the insects’ internal structures, and that they would use the labels to guide their inferential judgments (Gelman, 2003). However, we still expect systematic differences between the two conditions, even in this setting.

Our previous study investigating the time course of inductive judgments indicated that category labels, unlike feature labels, expedite the inferential process by providing substantial background information earlier in the time course of decision making (Yamauchi et al., 2007; see also Luhmann, Ahn, & Palmeri, 2006; Palmeri & Blalock, 2000). When labels conveyed category membership information, participants viewed them more often and earlier during the decision making process. Unlike feature labels, category labels are used as a heuristic to facilitate a reasoning process. This means that the more participants use the labels to make predictions, the more likely they are to respond quickly. In other words, there should be a positive correlation between the degree of using labels and the speed of making inferential judgments. This tendency should be pronounced in the category condition but not in the DNA condition.

We assume that high consistency scores with the matched stimuli (i.e., selection of a feature value consistent with the sample when both stimuli have the same label) would represent a high degree of using labels in the matched stimuli. Similarly, low consistency scores with the mismatched stimuli (i.e., selection of a feature value inconsistent with the sample when the sample and test

stimuli have different labels) would also represent a high degree of using labels. According to these assumptions, we predict that high consistency scores with the matched stimuli should correlate with fast responses in the category condition but not the DNA condition. Similarly, low consistency scores with the mismatched stimuli should also correlate with fast responses in the category condition but not the DNA condition.

Method

Participants. A total of 116 undergraduate students participated in the experiment for course credit. The data from 4 of the participants were removed because they either misunderstood the instructions or did not carry out the experiment as the instructions directed.² Thus, 112 participants were randomly assigned to two conditions: category ($n = 53$) or DNA ($n = 59$).

Materials and Procedure. The materials and procedure for this experiment were identical to those described in Experiment 1, except for minor modifications. In the DNA condition, the labels “monek” and “plaple” were characterized as names of two different components of DNA in these insects. The instructions further stated that the DNA components “affect the physical characteristics of these bugs” (see the Appendix). The category condition was identical to that condition in Experiment 1, in which the instructions characterized the labels as the names of two different types of insects. No reference was made about a link between these “types” and the physical characteristics of these insects.

Design. The experiment had a 2 (label condition: category, DNA) \times 2 (match status: matched, mismatched) \times 2 (feature instantiation: same, different) \times 2 (stimulus version: 1, 2) factorial design. The stimulus version factor did not interact with the others; therefore, this factor was collapsed in subsequent data analyses. As in Experiment 1, the polarity effect was measured by subtracting the proportion of consistent responses obtained with the mismatched stimuli from the one obtained with the matched stimuli (i.e., a “polarity score”), and the actual data analyses were carried out using 3 (label condition) \times 2 (feature instantiation) ANOVAs.

Results

Polarity effect. As predicted, the mean polarity score in the category condition was statistically indistinguishable from that in the DNA condition (Table 4). There was no main effect of label condition nor any interaction of label condition and feature instantiation (F_1 and $F_2 < 1$).

Correlation between response times and the use of labels. Given the matched stimuli in the category condition, there was a significant negative correlation between consistency scores and response times ($r = -.36, p < .01$), indicating that the participants who used labels more also tended to respond more quickly when the matched stimuli were given (Figure 3A). In the DNA condition, no correlation was present ($r = .05, p = .72$; Figure 3B).

With the mismatched stimuli, the correlation between consistency score and response times was not significant

Table 4
A Summary of the Results From Experiment 2

Label Condition	Match	Mismatch	Polarity Score (Match – Mismatch)
Category	.68	.34	.34
DNA	.66	.37	.29

Note—These numbers represent the mean consistency scores obtained in the two label conditions.

in the category condition ($r = .07, p = .64$). However, this lack of correlation was due to five outliers (Figure 3C; note that the dots for two nearly identical scores overlap). Without these outliers, there was a significant positive correlation between response times and consistency scores in the category condition ($r = .37, p < .05$). This supports the view that participants who used labels more also responded faster. There was a significant negative correlation in the DNA condition. However, the direction of the correlation was opposite to that in the category condition: The participants who used the labels more were in fact slower in their responses ($r = -.29, p < .05$; Figure 3D).

To investigate further the relationship between polarity scores and response times, we divided the participants in each label condition into two groups—those who made fast responses and those who did not (median split). We then calculated the polarity scores among fast responders and slow responders. Among the fast responders, the mean polarity score in the category condition ($M = .40$) was marginally higher than that in the DNA condition ($M = .23$) [$t(64) = 1.73, p = .09$]. Without the five outliers in the category condition (see the previous paragraph and Figure 3C), the difference between the category ($M = .50$) and DNA ($M = .23$) conditions was significant [$t(59) = 2.65, p = .01$].

Common responses: A cluster analysis. The results from the cluster analysis, which measured the degree of response commonality among the participants in each label condition, show that the proximity distances observed in the category condition were significantly

smaller than those observed in the DNA condition ($z = 5.34, p < .001$; Table 5). This suggests that participants in the category condition were similar to each other in their response patterns. Note that the mean proximity distances were not very different between (category condition, $M = 17.28$; DNA condition, $M = 17.95$), but the distributions of proximity scores in the two conditions were substantially different. Specifically, the proximity score distributions were different in their medians (category, 18; DNA, 19), standard deviations (category, 4.93; DNA, 5.39), and skewness (category, -0.56 ; DNA, -0.87); the p s for all measures were $< .05$.³

Given the proximity distance of 12, 40% of the participants in the category condition were included in the largest cluster. Given the same proximity distance, 37% of the participants in the DNA condition formed the largest cluster.

Did the category and DNA conditions differ in the participants' use of rule- and similarity-based strategies? An additional analysis showed that 40.0% of the participants in the category condition and 33.9% of those in the DNA condition adopted a rule-based strategy (i.e., their polarity scores were at least .5; $z = 0.43, p = .67$). In contrast, 20.8% in the category condition and 32.2% in the DNA condition employed a similarity-based strategy (i.e., their polarity scores were ≤ 0 —see the Results section of Experiment 1 for the operational definitions of rule- and similarity-based reasoning strategies; $z = 1.15, p = .25$). An additional analysis of the distributions of polarity scores was also consistent with the results from Experiment 1: The polarity distribution in the DNA condition

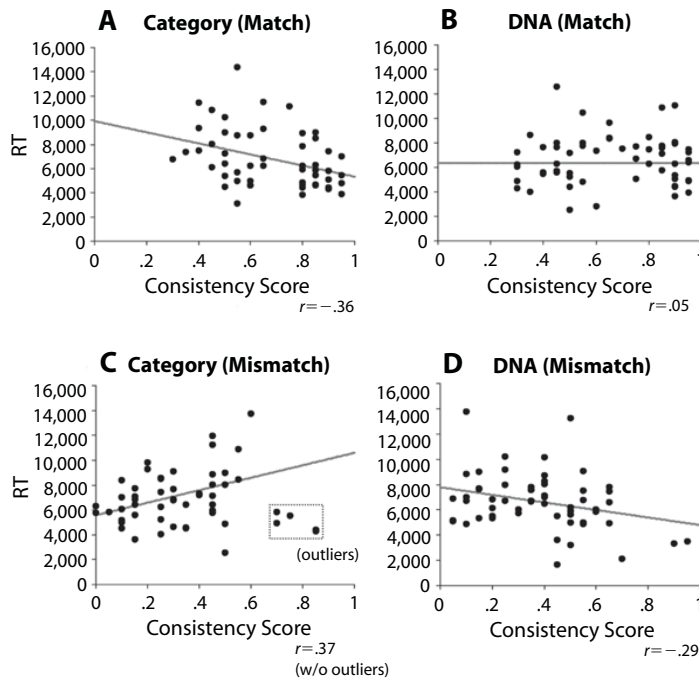


Figure 3. Correlations between consistency scores and response times. The straight lines included in these figures are linear regressions.

Table 5
A Summary of the Cluster Analysis in Experiment 2

	Proximity Distance		Average Proximity Distance	z Score	p Value
	8	12			
Category	23%	40%	17.28		
DNA	29%	37%	17.95	5.34	<.001

Note—The left columns show the percentages of participants who formed the largest cluster at a given level of proximity distance. Proximity distance was measured by the city-block metric. “Average proximity distance” represents the average distance of all pairs of the participants in a given label condition. The *z* scores and *p* values represent results from a Wilcoxon ranked sum test comparing the proximity distances in the category condition with those in the DNA condition.

was skewed relative to that in the category condition (DNA skewness = 0.22, 95% CI = [−0.22, 0.61]; category skewness = 0.09, 95% CI = [−0.33, 0.49]).

Taken together, the results from Experiment 2 suggest that the category and DNA conditions were roughly equivalent in their overall mean polarity scores but were still substantially different, in that the labels were used as an inferential guide to expedite decision processes in the category condition, whereas they were used as a salient feature in the DNA condition.

Discussion

Experiment 2 addressed the issue of relevance in predictive inference. Clearly, relevance is an important factor in constraining inference: The mean polarity score in the DNA condition was statistically equivalent to the one in the category condition, but the two conditions remained distinct in terms of the ways the labels interacted with decision processes. There was a significant correlation between response times and the degree of using labels in the category but not in the DNA condition, suggesting that category labels, unlike feature labels, initiate reasoning processes.

Although the mean polarity scores obtained in the two conditions were roughly equivalent, the homogeneity in the category condition was still substantial. The individual responses observed in the category condition were highly similar as compared with those in the DNA condition, suggesting that the distinction between category and feature labels cannot be attributed solely to the relevance of the labels in predictive inference.

GENERAL DISCUSSION

We attached arbitrary labels to imaginary insects and manipulated the meanings associated with those labels, in order to investigate the manipulation’s influence on how participants predicted the characteristics of body parts of the insects. In one condition, the instructions characterized the arbitrary labels as the names of categories to which the imaginary insects belonged (a category condition). In the other conditions, they characterized the same labels as the names of diseases, foods, or islands associated with the insects (attribute conditions). These manipulations were introduced solely in the instructions that the participants received. All participants were shown the same stimuli and answered the same questions.

Despite these subtle manipulations, there were significant disparities in the ways category labels and feature labels influenced predictive inference. Overall, participants tended to predict the feature values of test stimuli on the basis of the matched or mismatched status of labels attached to the sample and test stimuli when those labels carried category membership information. This tendency was far less pronounced for feature labels, which conveyed attribute information. Furthermore, whereas category labels made the participants’ responses highly homogeneous, this tendency was absent in feature labels.

The disparity between category and feature labels goes beyond the contextual information with which the individual labels are associated. Even when the feature labels were highly relevant to the inference questions, there was a systematic difference in the ways that category and feature labels influenced reasoning processes. In Experiment 2, we made feature labels highly diagnostic to the prediction of body parts by stating that they represented the names of DNA components that affect the physical characteristics of the insects. In this case, both the category and the feature labels were clearly helpful, as observed in the high polarity scores in the DNA condition as well as in the category condition. However, participants in the DNA and category conditions derived their predictions in drastically different manners. In the category condition, the participants who responded quickly were most likely to show high polarity scores, suggesting that they used the labels to ease the reasoning process. Such a use of the labels was absent in the DNA condition.

The present results extend the findings from the Yamauchi et al. (2007) study in two ways. First, the distinction between category labels and feature labels in inductive inference is quite robust, in that it can be generalized to feature labels that are basic and central to the inference task (in this case, when the attributes were associated with the biological and behavioral characteristics of the insects). Second, the influence of category labels is fundamental because they influence reasoning processes by providing an initial assessment for inductive generalization. In other words, category labels point out not only what features to project but also how to project them.

Where does the inductive potential of category labels come from? We offer three speculations. The first reason is cognitive economy (Rosch, 1978). Research has shown that people focus on a single category and make a predictive inference on the basis of the category that is immediately recognizable (Lagnado & Shanks, 2003; Murphy & Ross, 1994; Ross & Murphy, 1996). By forming concepts, we treat individual objects as a group and deal with the characteristics of the group as a whole, rather than of the individual objects separately. “Grouped” representation can expedite many cognitive tasks. This may be the primary reason why category labels play a pivotal role in inductive inference, because they can dramatically improve cognitive economy.

Another reason is the communicative constraints that category labels receive. Category labels, which generally correspond to count nouns, are subject to communicative constraints to a larger degree than are feature labels, which generally correspond to adjectives. Because count nouns

vastly outnumber adjectives in linguistic communication, many constraints, such as the intention of a speaker (Bloom, 1996), conversational agreements between interlocutors (Malt & Sloman, 2004; A. B. Markman & Makin, 1998), and cultural and historical precedents (Malt, Sloman, Gennari, Shi, & Wang, 1999) are likely to influence what category labels stand for. Because of these constraints, category labels deviate from "category representation" and affect inductive inference separately from similarity information (see, e.g., Gelman & Markman, 1986; Malt et al., 1999).

Finally, we think that category labels are also different from feature labels in the framework knowledge that noun labels evoke. An intuitive belief that categories are created because of some "essence" (e.g., psychological essentialism; Gelman, 2003; Medin & Ortony, 1989), a general assumption that one object has only one label (the "mutually exclusive constraint"; E. M. Markman, Wasow, & Hansen, 2003), and a sense of generality associated with generic noun labels (Gelman, Hollander, Star, & Heyman, 2000; Prasada, 2000; Yamauchi, in press) inform basic knowledge about how category members are organized (Heys, 2006). This framework knowledge can be accentuated by category labels, which thereby influence inferential predictions differently from feature labels (see, e.g., Gelman & Heyman, 1999).

Conclusion

This study contrasted the influence of category labels to that of feature labels by manipulating the meanings associated with arbitrary labels. The two experiments showed that fundamental differences exist in how category and feature labels influence inductive inference. We suggest that category labels, unlike feature labels, initiate inductive inference and expedite decision processes. As a result, category labels tend to make inductive generalizations more homogeneous and polarized.

AUTHOR NOTE

This research was supported by a College Faculty Research Enhancement Award, a Glasscock Center Faculty Fellow Award, and a Developmental Grant by the Mexican American and U.S. Latino Research Center (Texas A & M University), all given to the first author. The authors thank Art Markman and Wooyoung Jung for their comments. Please address correspondence to T. Yamauchi, Department of Psychology, Mail Stop 4235, Texas A&M University, College Station, TX 77843 (e-mail: tya@psych.tamu.edu).

REFERENCES

- ANDERSON, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- BLOOM, P. (1996). Intention, history, and artifact concepts. *Cognition*, **60**, 1-29.
- BROWN, R. W. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal & Social Psychology*, **55**, 1-5.
- CLAPPER, J. P., & BOWER, G. H. (2002). Adaptive categorization in unsupervised learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 908-923.
- GELMAN, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford: Oxford University Press.
- GELMAN, S. A., & HEYMAN, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science*, **10**, 489-493.
- GELMAN, S. A., HOLLANDER, M., STAR, J., & HEYMAN, G. D. (2000). The role of language in the construction of kinds. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 39, pp. 201-263). San Diego: Academic Press.
- GELMAN, S. A., & MARKMAN, E. M. (1986). Categories and induction in young children. *Cognition*, **23**, 183-209.
- HEIT, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review*, **7**, 569-592.
- HEIT, E., & RUBINSTEIN, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 411-422.
- HEYS, B. K. (2006). Knowledge, development, and category learning. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 46, pp. 37-77). San Diego: Academic Press.
- JOHANSEN, M. K., & KRUSCHKE, J. K. (2005). Category representation for classification and feature inference. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 1433-1458.
- KRUSCHKE, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.
- LAGNADO, D. A., & SHANKS, D. R. (2003). The influence of hierarchy on probability judgment. *Cognition*, **89**, 157-178.
- LOVE, B. C., MEDIN, D. L., & GURECKIS, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, **111**, 309-332.
- LUHMANN, C. C., AHN, W.-K., & PALMERI, T. J. (2006). Theory-based categorization under speeded conditions. *Memory & Cognition*, **34**, 1102-1111.
- MALT, B. C., & SLOMAN, S. A. (2004). Conversation and convention: Enduring influences on name choice for common objects. *Memory & Cognition*, **32**, 1346-1354.
- 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 & Language*, **40**, 230-262.
- MARKMAN, A. B., & MAKIN, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General*, **127**, 331-354.
- MARKMAN, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.
- MARKMAN, E. M., & HUTCHINSON, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, **16**, 1-27.
- MARKMAN, E. M., WASOW, J. L., & HANSEN, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, **47**, 241-275.
- MEDIN, D. L., COLEY, J. D., STORMS, G., & HAYES, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review*, **10**, 517-532.
- MEDIN, D. L., & ORTONY, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). Cambridge: Cambridge University Press.
- 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.
- OSHERSON, D. N., SMITH, E. E., WILKIE, O., LÓPEZ, A., & SHAFIR, E. (1990). Category-based induction. *Psychological Review*, **97**, 185-200.
- PALMERI, T. J., & BLALOCK, C. (2000). The role of background knowledge in speeded perceptual categorization. *Cognition*, **77**, B45-B57.
- PRASADA, S. (2000). Acquiring generic knowledge. *Trends in Cognitive Sciences*, **4**, 66-72.
- ROSCHE, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.
- ROSS, B. H., & MURPHY, G. L. (1996). Category-based predictions: Influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 736-753.
- SLOMAN, S. A. (1993). Feature-based induction. *Cognitive Psychology*, **25**, 231-280.

- SLOMAN, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, **119**, 3-22.
- SLOMAN, S. A., LOVE, B. C., & AHN, W.-K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, **22**, 189-228.
- SLOUTSKY, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, **7**, 246-251.
- SLOUTSKY, V. M., & FISHER, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, **133**, 166-188.
- SMITH, E. E., & SLOMAN, S. A. (1994). Similarity- versus rule-based categorization. *Memory & Cognition*, **22**, 377-386.
- WALDMANN, M. R., & HAGMAYER, Y. (2006). Categories and causality: The neglected direction. *Cognitive Psychology*, **53**, 27-58.
- WALTON, G. M., & BANAJI, M. R. (2004). Being what you say: The effect of essentialist linguistic labels on preferences. *Social Cognition*, **22**, 193-213.
- WAXMAN, S. R., & BOOTH, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology*, **43**, 217-242.
- WHITMAN, W. (1985). *Walt Whitman's Leaves of grass* (Malcolm Cowley, Ed.). New York: Penguin. (Original work published 1855)
- YAMAUCHI, T. (2005). Labeling bias and categorical induction: Generative aspects of category information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 538-553.
- YAMAUCHI, T. (in press). Linking syntax and inductive reasoning: Categorical labeling and generic noun phrases. *Psychologia*.
- YAMAUCHI, T., KOHN, N., & YU, N.-Y. (2007). Tracking mouse movement in feature inference: Category labels are different from feature labels. *Memory & Cognition*, **35**, 852-863.
- YAMAUCHI, T., & MARKMAN, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 776-795.

NOTES

1. The skewness of a distribution is defined as

$$y = \frac{E(x - \mu)^3}{\sigma^3},$$

where μ is the mean of x , σ is the standard deviation of x , and $E(t)$ represents the expected value of t (MATLAB Statistics Toolbox; www.mathworks.com/access/helpdesk/help/toolbox/stats/). The confidence intervals were obtained from 1,000 bootstrap data samples. Note that skewness was measured relative to the mean of a given distribution; therefore, the overall means of the consistency scores obtained in each label condition did not affect the degree of skewness.

2. The responses made by these participants were $\leq .3$ for all four of the within-subjects manipulations—namely, (1) stimuli in the *same* feature set with *matched* labels, (2) stimuli in the *same* feature set with *mismatched* labels, (3) stimuli in the *different* feature set with *matched* labels, and (4) stimuli in the *different* feature set with *mismatched* labels. Although that cutoff point was arbitrary, we reasoned that these participants simply misunderstood either the response key assignments or the instructions. If the participants had responded randomly, their response proportions should have been about .5. Because their proportions were far below that chance level (.3 or below) with all independent variables, we reasoned that these participants simply misunderstood the instructions.

3. The p values were estimated from bootstrap confidence intervals obtained from 1,000 bootstrap data samples.

APPENDIX

Excerpts of the Instructions Given in Experiments 1 and 2

Italics are added in the category condition excerpt to indicate that the instructions in the other conditions differed only in the highlighted sentences.

Category Condition

In this experiment, we are interested in the way you make judgments. As you start this experiment, you will see new bugs that you've never seen before. *Scientists divided these new bugs into two types, which are called "monek" and "plaple."* On your left side, you see a sample of one new bug, and on your right side you see a question about another new bug. Each bug will be depicted with 5 different body parts—horns, head, torso, legs, and tail. *These new bugs will be shown with a sign that describes the type to which each bug belongs. In the example below, one bug belongs to one type, and the other bug belongs to another type. . . .*

Disease-Attribute Condition

Scientists found that these new bugs carry two kinds of disease, which are called "monek" and "plaple."

These new bugs will be shown with a sign that describes the disease that each new bug carries. In the example below, one bug carries one kind of disease, and the other bug carries another kind of disease. . . .

Food-Attribute Condition

Scientists found two kinds of food that these new bugs eat every day, which are called "monek" and "plaple."

These new bugs will be shown with a sign that describes the kind of food each new bug eats every day. In the example below, one bug eats one kind of food, and the other bug eats another kind of food. . . .

Island-Attribute Condition

Scientists found that these new bugs live on two different islands, which are called "monek" and "plaple."

These new bugs will be shown with a sign that describes the island where each new bug lives. In the example below, one bug lives on one island, and the other bug lives on another island. . . .

DNA Condition (Experiment 2)

Scientists found two kinds of DNA components that these new bugs have, which are called "monek" and "plaple." These DNA components are known to affect the physical characteristics of these bugs.

These new bugs will be shown with a sign that describes the kind of DNA components each new bug has. In the example below, one bug has one kind of a DNA component, and the other bug has another kind of a DNA component. . . .