Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Are Category Labels Features or Naïve Assumption?

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 33(33)

ISSN
1069-7977

Authors

Yu, Na-Yung
Yamauchi, Takashi

Publication Date
2011

Peer reviewed

Are Category Labels Features or Naïve Assumption?

Na-Yung Yu (nayungyu@gmail.com)
Takashi Yamauchi (tya@psyc.tamu.edu)
Department of Psychology, Mail Stop 4235
Texas A&M University, College Station, TX 77843 USA

Abstract

Category labels are known to guide inductive generalizations by modifying the representation of stimuli (i.e., the labeling effect); yet, the mechanisms of this effect remain unclear. One view suggests that shared category labels increase overall similarity between items as shared physical features do. The other view suggests that category labels are qualitatively different from category features, as category labels help integrate prior knowledge. The present study examined these two competing views with respect to two types of background knowledge – domains of categories (living things vs. man-made objects) and the amount of knowledge (the number of listed exemplars). The results from two experiments suggest that category labels are likely to be used as features for man-made objects, while category labels help create a naïve assumption for living things, implicating that similarity-based statistical processes and knowledge-based structured processes are employed interactively to cope with different domains of knowledge.

Keywords: Category Labels; Background Knowledge, Inductive Inference

One critical question in inductive generalization research is how the similarity-based statistical mechanism (e.g., Rodger & McClelland, 2003; Sloutsky & Fisher, 2004) and structured prior knowledge (e.g., intuitive theories and assumptions about domain knowledge) interact (Gelman & Heyman, 1999; Markman & Ross, 2003; Medin, Coley, Storms & Hayes, 2003; Griffiths & Tenenbaum, 2003; Kemp & Tenenbaum, 2009; Yamauchi, 2009). The former provides a vehicle for detecting association of stimuli, and the latter offers boundaries on the search space. The two have to work together but how? When does one strategy become dominant while the other is relegated to the background?

This article examines the capacity of shared category labels. Studies have shown that when two objects carry the same label, we tend to assume that these objects have some important characteristics in common; when objects carry different labels, we tend to think that they have some distinctive characteristics (i.e., labeling effect) (Gelman & Heyman, 1999; Gelman & Markman, 1986; Lupyan, 2008; Murphy, 2003; Sloutsky & Fisher, 2004; Waxman & Braun, 2005; Yamauchi & Yu, 2008; Yu, Yamauchi & Schumatcher, 2008). To explain the labeling effect, researchers suggested two hypotheses: the naïve-theory hypothesis and the label-as-feature hypothesis. The naïve-theory hypothesis states that people tend to make an intuitive assumption that category labels are qualitatively different from physical features (Gelman & Heyman, 1999; Gelman & Markman, 1986; Murphy, 2003; Waxman & Braun, 2005; Yamauchi & Yu, 2008). When physical similarity of stimuli was pitted against shared category labels, participants tended to judge that shared category labels are more important than shared physical features in predicting characteristics of the objects (Gelman & Markman, 1986; Yamauchi & Yu, 2008). The other dominant view, the label-as-feature hypothesis, argues that there is no indication that people have such a naïve theory about category labels (Anderson, 1990; Sloutsky & Fisher, 2004). According to this view, matching labels increase the perception of overall similarity just like matching physical features do. In both the naïve-theory and label-as-feature hypotheses, category labels are expected to influence inductive generalization. However, the naïve theory approach argues that labels are used to form naïve assumptions, while the label-as-feature approach suggests that labels are used as a feature.

This article attempted to find the boundary conditions where these two conflicting views can complement each other. We suggest that these two conflicting views can complement each other with respect to two types of general knowledge that an observer has – domains of categories and the amount of categorical knowledge. Previous studies have shown that people have general awareness of how living things and man-made objects differ, and the two broad categories of concepts seem to be separate in our semantic knowledge (Caramazza & Mahon, 2003; Moss, Tyler, Durrant-Peatifield & Bunn, 1998; McRae, Cree, Seidenberg, & McNorgan, 2005). Studies also suggest that the amount of knowledge is one important variable of the labeling effect. For example, subjects who do not have much specialized domain knowledge tend to make inferences about unfamiliar diseases using taxonomic relations, while experts employ causal, anatomical, and ecological variables to make inferences (Lopez, Atran, Coley, Medin, & Smith, 1997; Shafto & Coley, 2003).

Hypothesis and Predictions

Natural kind categories are highly correlated in their attributes as compared to artifacts (Malt & Smith, 1984; McRae & Siderberg, 1998), and correlated features are the critical factor that promotes category-based induction (Sloman, Love, & Ahn, 1998; Gelman, 1988). In this regard, we predict that the labeling effect is particularly strong when labels represent living things as opposed to man-made objects. If the labeling effect arises because people treat labels as a feature, the labeling effect would be contingent on the amount of knowledge that an observer has. In contrast, if the labeling effect arises from people's naïve assumptions, the labeling effect would be relatively independent from the amount of knowledge that an observer has. We suggest that for concepts representing living things, people have a naïve assumption that category labels are special; thus, we predict that a strong labeling effect would emerge for concepts related to living things. Here, we expect that the labeling effect is relatively independent of the amount of knowledge that an observer has. In contrast, for man-made objects, we suggest that people treat category labels as a feature. Thus, we predict that the labeling effect observed for the man-made objects would be related to the amount of knowledge an observer has.

Two experiments were conducted to test this idea. Labeling effect was measured in a triad-based similarity judgment task (e.g., Gelman & Markman, 1986; Sloutsky & Fisher, 2004; Yu, Yamauchi & Schumatcher, 2008). Participants were presented with three pictures of animal tissues —a target placed at the top and two base pictures placed at the bottom (Figure 1), Their task was to decide which base picture, left or right, was more similar to the original picture (i.e. a triad task). In each stimulus frame, one base picture was more similar to the target than the other base picture (Figure 1; later the more *sim*ilar base picture is called the similar base picture; the less similar—thus more *dis*similar—base picture is called dissimilar base picture). The similar base picture always had a different label from the target, and the dissimilar base picture always had the same label as the target (Figure 1). Thus, the labeling effect was measured by the proportion of participants selecting dissimilar base pictures (that has the same label as the target) as more similar to the target picture. For example, in Figure 1, we examined the proportion of participants selecting the picture on the right (i.e., dissimilar base picture) as more similar to the target than the picture on the left.

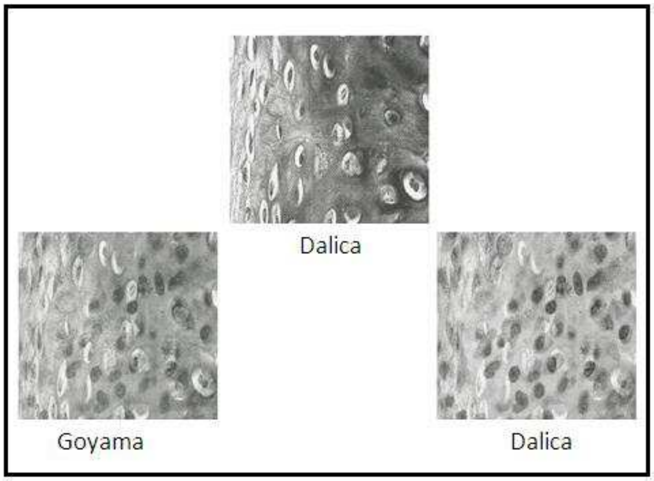

Figure 1: Sample trial shown in Experiment 1

Experiment 1

To compare labeling effect of living things and man-made objects, we used descriptive labels as a means of manipulating the type of categories. Throughout the experiment, the same set of pictures was presented to participants with the same set of arbitrary nonsensical labels (e.g., "Dalica," or "Goyama"). The type of category associated with the labels was manipulated solely in the instructions that participants received in each condition. In the diseased cell condition, participants were told that the pictures were diseased cells and the labels (e.g., "Goyama") represented names of diseases. In the painting style condition, participants were told that pictures represented abstract paintings and that the labels (e.g., "Goyama") indicated the names of painting styles. In the painter condition, participants were told that the pictures represented paintings and that labels (e.g., "Goyama") referred to the painters' names. Note that in the painter condition, labels were indexical in the sense that the label indicated a painter who created the given pictures; however, in the painting style condition, labels were categorical because the label represented a painting style that could include other paintings that were drawn by many painters in that style.

As previous studies suggested, the amount of knowledge can be one important factor in explaining labeling effect (Lopez, Atran, Coley, Medin, & Smith, 1997; Shafto & Coley, 2003). One way to capture the amount of background knowledge is to ask participants to list as many exemplars of a category as possible (Barsalou, 1985; Smith, Ward, Tindell, Sifonis, & Wilkenfeld, 2000; Ward, Patterson, Sifonis, Dodds, & Saunders, 2002). Folwing a procedure a previous study suggested (Smith et al., 2000), we asked participants to list as many exemplars as possible associated with the categories (diseases, painting styles, or painters) after the similarity judgment task. The amount of knowledge was measured by counting the number of exemplars that each participant generated.

Method

Participants A total of 304 undergraduates participated for course credit. They were randomly assigned to one of three conditions: diseased cell (*n* = 93), painting style (*n* = 106), and painter (*n* = 105) conditions.

Materials Stimuli were triads of monochrome pictures (Figure 2). The target was an original picture of animal tissues, and the two base pictures were of morphed images of two original pictures. Five pairs of original cell pictures were selected from a well-known textbook of veterinary histology (Bacha & Bacha, 2000). From these 10 pictures, five pairs were created. For each pair, one original tissue picture was merged with the other original picture to 18 different degrees using MorphMan 4.0 (2003) software. Altogether, 90 morphed pictures (18 morphed pictures from each of five pairs) were created.

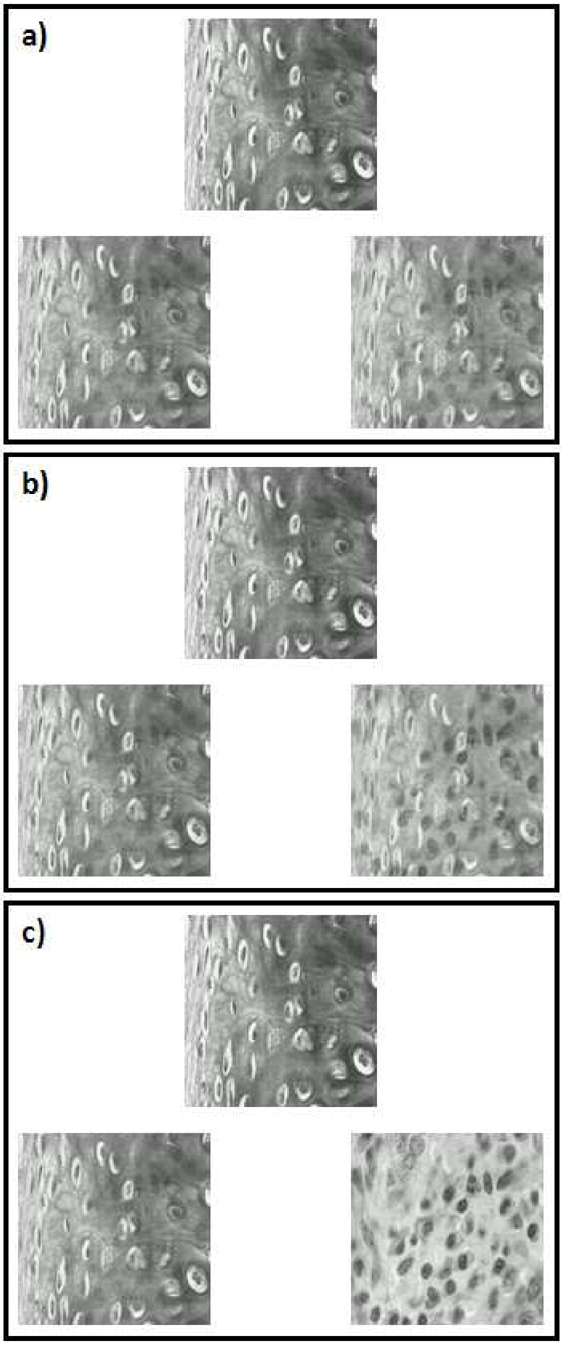

Figure 2: Three levels of physical difference. (a) low-difference, (b) medium- difference, (c) high-difference

From the morphed pictures, three levels of physical difference—low-, medium-, and high-difference—were created based on the degree of merging of the two original pictures (Figure 2). In the low-difference condition, the target picture and the dissimilar base picture were not very different (Figure 2a); in the medium-difference condition, the target picture and the dissimilar base picture were moderately different (Figure 2b); and in the high-difference condition, the target picture and the dissimilar base picture were highly different (Figure 2c). Two sets of base pictures were randomly selected at each level of physical difference and were combined with one of two original pictures in each pair, yielding 12 triads for each pair (a total of 60 triads = 5 pairs × 12 triads).

Design The experiment had a 2 (Label Condition; no-label vs. same-label conditions; between-subjects) × 3 (Physical Difference; low-, medium-, or high-difference; within-subjects) factorial design. The dependent measure was the proportion of participants selecting the dissimilar base pictures as more similar to the target than the other base pictures.

Procedure Participants were presented with 60 triads of pictures one at a time and judged which base picture was more similar to the target using left or right arrow key (E-Prime 1.1, Psychology Software Tools Inc., 2002). The order of presenting stimuli was determined randomly. The dissimilar base picture was presented on the left or the right side an equal number of times. The experiment lasted about 10 minutes. After carrying out the similarity judgment task, participants were asked to list as many exemplars of the category as possible in 2 minutes. Participants in the diseased cell condition were asked to list disease names, participants in the painting style condition were asked to list names of painting style (e.g., impressionist paintings, classical painting, etc.), and participants in the painter condition were asked to list names of painters.

Results & Discussion

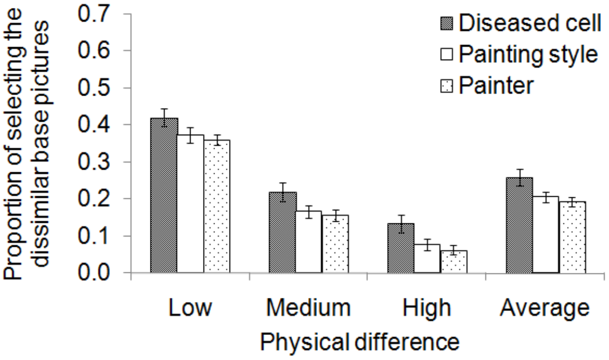

Figure 3: Results from Experiment 1

Figure 3 summarizes the main results from Experiment 1. When the labels indicated living things (diseased cell condition), participants used the labels significantly more often than when the labels represented man-made things (painter and painter style condition) (Figure 3). The main effect of label condition was significant, $F(2, 301) = 4.22$, $MSE = .08$, $p = .01$, $\eta^2 = .03$. The proportion of participants selecting the dissimilar base pictures was significantly higher in the diseased cell condition ($M = 0.26$) than in the painting style condition ($M = 0.21$) or the painter condition ($M = 0.19$): diseased cell vs. painting style, $t\ (197) = 2.02$, $SE = .03$, $p = .05$, $d = 0.29$, diseased cell vs. painter, $t\ (196) = 2.68$, $SE = .02$, $p < .01$, $d = 0.38$. The labeling effect between the painting style and painter conditions was statistically indistinguishable, *t* (209) = .63, *SE* = .02, *p* = .53, *d* = 0.09. The results suggest that, at least for the man-made objects, categorical labels (painting style condition) are not considered more special than indexical labels (painter condition). There was no interaction effect of physical difference and label condition, *F*(2,392) < 1.0.

Participants knew more about diseases than about painters/painting styles (Table 1). As participants listed more disease names (*M* = 9.47, *SD* = 3.72) than painting style names (*M* = 4.64, *SD* = 2.36) and painter names (*M* = 3.22, *SD* = 2.35), The number of exemplars participants listed for disease names, painting style names, and painter names was significantly different from each other, *F*(2, 301) = 129.42, *MSE* = 8.07, *p* < .01, η^2 = .46. More disease names were generated than names of painting styles or painter names, diseases vs. painting styles, *t* (197) = 11.09, *SE* = .44, *p* = .001, *d* = 1.58, diseases vs. painters, *t* (196) = 14.32, *SE* = .44, *p* < .001, *d* = 2.04. Also, more painting styles were generated than painter names, *t* (209) = 4.39, *SE* = .33, *p* < .001, *d* = 0.60.

Table 1: The number of exemplars participants listed for disease, painting styles, and painters in Experiment 1.

Label condition	Result
Disease	9.47 (3.72)
Painting style	4.64 (2.36)
Painter	3.22 (2.35)

The labeling effect might have been in the disease condition because our participants knew more about disease cells than painter names and painting style. Thus we performed ANCOVA with the number of listed exemplars as a covariate. The result show that even after the amount of background knowledge (i.e., the number of exemplars associated with labels) was controlled, the difference between the diseased cell, the painting style and painter conditions remained robust: *F*(2, 300) = 2.84, *MSE* = .08, *p* = .06, η^2 = .02. The results suggest that, at least for diseases, labels are used as a type of naïve assumption rather than a part of attributional information/knowledge people have.

In Experiment 1, the stimuli were produced by morphing pictures of animal tissues; thus it is possible that participants did not use labels in the painting style condition and the painter condition because these stimuli did not look like actual paintings. Therefore, Experiment 2 tested with stimuli obtained from real paintings.

Experiment 2

The goal of Experiment 2 is to replicate the results of Experiment 1 using a different set of pictures (Figure 4). One possible explanation for the differential effect of labels in Experiment 1 is that the stimuli looked more like cells than paintings. Experiment 2 was designed to rule out this explanation. The design, task, and materials in Experiment 2 were identical to those described in Experiment 1, except that real abstract paintings were used as stimuli in Experiment 2.

The idea that a naïve assumption is formed for category labels representing living things predict that the labeling effect should be still strong in the disease cell condition as compared to the painter and painting style conditions, even though the stimuli used in Experiment 2 barely resemble biological cells. The hypothesis that labels are treated like features for artifact concepts suggests that the amount of background knowledge affects the painter condition and painting style condition but not the disease cell condition.

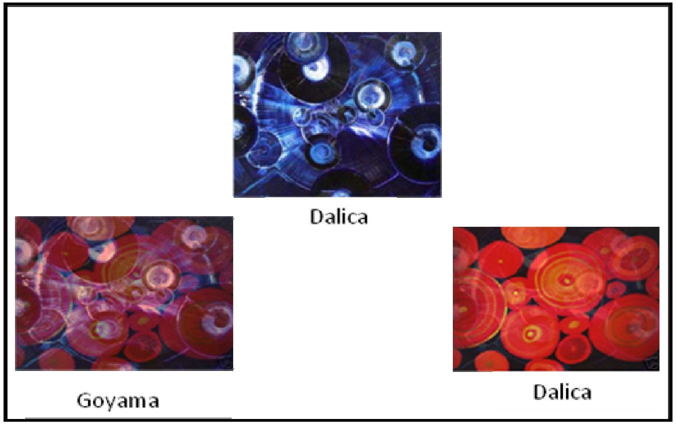

Figure 4: Sample trial shown in Experiment 2

Method

Participants Three hundred eighty-seven undergraduates participated for course credit. They were randomly assigned to one of three conditions: diseased cell (*n* = 130), painting style (*n* = 121), and painter (*n* = 136) conditions.

Materials Five pairs of abstract paintings were selected from various websites. The procedure of creating 60 triads of original and morphed pictures (see Figure 2 for the sample trial) was identical to the procedure described in Experiment 1. As in the previous experiment, the triads of pictures had three levels of physical difference (low-, medium-, and high-difference).

Design & Procedure The design and procedure of Experiment 2 was identical to those described in Experiment 1. Participants were presented with the same 60 triads of paintings with labels and were asked to judge which base picture was more similar to the target. Participants in the diseased cell condition were instructed that the stimuli were cell pictures infected by diseases and that the labels specified the names of the diseases that infected the cells. Participants in the painting style condition were informed that the stimuli were paintings and that the labels indicated the names of painting styles. Participants in the painter condition were also told that the stimuli were paintings and that the labels indicated the painters' names. As in Experiment 1, participants were asked to judge similarity of pictures and then to list as many exemplars of the category as possible right after the similarity judgment task. All participants received the same stimuli with the same label names. The meanings of labels were manipulated only in the instructions.

Results

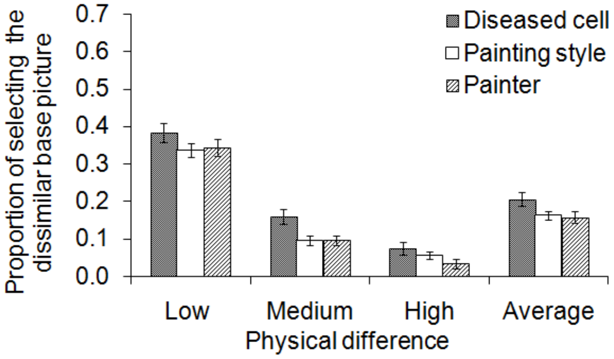

Figure 5: Results from Experiment 2

Though pictures shown in Experiment 2 barely resemble biological cells, when labels represented living things (diseased cell condition), participants used the labels considerably more often than when the labels indicated man-made objects (painter and painting style conditions) (Figure 5). The main effect of label condition was significant, *F* (2, 384) = 5.24, *MSE* = 0.05, *p* = .006, η^2 = .03. The proportion of participants selecting the dissimilar base pictures (the tendency to use category labels) was significantly higher in the diseased cell condition (*M* = 0.21) than in the painting style condition (*M* = 0.16) or in the painter condition (*M* = 0.16): diseased cell vs. painting style, *t* (249) = 2.30, *SE* = .02, *p* = .02, *d* = .29; diseased cell vs. painter condition, *t* (264) = 2.80, *SE* = .02, *p* = .006, *d* = .34. The proportion of selecting dissimilar base pictures was not different in the painting style condition compared to the painter condition, *t* (255) = 0.38, *SE* = .01, *p* = .71, *d* = .05.

As in Experiment 1, participants listed more disease names (*M* = 9.69, *SD* = 3.65) than painting style names (*M* = 4.74, *SD* = 2.41) and painter names (*M* = 3.54, *SD* = 2.67) ; *F*(2, 384) = 159.09, *MSE* = 8.73, *p* < .001, η^2 = .45. Participants listed significantly more disease names than painting style names, *t* (249) = 12.20, *SE* = .41, *p* < .001, *d* = 1.54; they generated significantly more disease names than painter names, *t* (264) = 16.27, *SE* = .38, *p* < .001, *d* = 2.00. They also listed significantly more painting style names than painter names, *t* (255) = 3.76, *SE* = .32, *p* < .001, *d* = .47.

Table 2: The number of exemplars participants listed for disease, painting styles, and painters in Experiment 2.

Label condition	Result
Disease	9.69 (3.65)
Painting style	4.74 (2.41)
Painter	3.54 (2.67)

We hypothesized that people would treat category labels special for living things thus labeling effect would be relatively independent from the amount of knowledge. There was no correlation between the listed exemplars and the labeling effect in the disease cell condition, *r*(128) = -.04, *p* = .64. The labeling effect in the disease-cell condition was strong even after controlling the amount of knowledge they had. ANCOVA by controlling the number of listed exemplars as a covariate showed that even when the number of listed exemplars was controlled, the difference between the diseased cell and the painter and painting style conditions remained strong, *F*(2, 383) = 6.88, *MSE* = 0.05, *p* = .001, η^2 = .04.

We also hypothesized that people would treat category labels as a type of feature for man-made objects, thus the labeling effect would be related to the amount of knowledge. A correlation analysis revealed that there was a significant negative correlation between the amount of background knowledge (i.e., the number of exemplars listed by participants) and the amount of labeling effect (i.e., the proportion of participants selecting the dissimilar pictures) for man-made objects (painters, painting styles), *r*(255) = -.16, *p* = .01. This result is consistent with the general finding that the labeling effect is higher when background knowledge is limited, suggesting that given man-made objects, people tend to use labels when they did not know much about the category. However, this is not the case in the living-thing concepts.

Note that participants in Experiment 2 judged the stimuli more likely to be abstract paintings than animal cells. After the similarity judgment task, participants in the diseased cell and painter conditions viewed 10 original painting stimuli one at a time and rated the extent to which the pictures actually looked like diseased cells or paintings. Participants in the diseased cell condition estimated the likelihood of each original painting stimulus to be a disease cell with a 0-100 scale. Participants in the painter condition estimated the likelihood of each original painting stimulus to be a painting with a 0-100 scale. The average rating score in the painting style condition (*M* = 62.37) was significantly higher than that in the diseased cell condition (*M* = 48.82), *t*(221) = 7.84, *p* < .001, suggesting that the proportion of participants who considered the stimuli to be actual “paintings” was much higher than the proportion of participants who considered the stimuli to be “cells.”

General Discussion

How are labels used in inductive generalization? The results from present study indicated that category labels are used more for predictions of living things than for man-made objects, and this tendency is relatively separated from the amount of knowledge. However, for man-made objects, the tendency to use category labels is related to the amount of knowledge (the amount of listed examples) a participant possesses, and there is no clear distinction between category labels and indexical labels. Experiment 1 showed that people use category labels in inductive inference of living things more than man-made objects. Also the tendency to use labels was relatively independent of the amount of knowledge. Results of Experiment 2 also showed that, when using real paintings, the amount of knowledge was negatively correlated with the tendency to use category labels on similarity judgments involving man-made objects (i.e., paintings). The relation between the amount of knowledge and the tendency to use labels was absent in the disease condition. These results suggest that types of knowledge and the amount of knowledge interacted with the labeling effect.

The current finding extends the naïve theory-based account and similarity-based views of inductive inference. The naïve theory approach suggests that people form a naive assumption that shared category membership is qualitatively different from physical features (Gelman & Heyman, 1999; Gelman & Markman, 1986; Lupyan, 2008; Murphy, 2003; Rhemtulla & Hall, 2008; Waxman & Braun, 2005; Yamauchi & Yu, 2008; Yamauchi & Yu , 2008; Yamauchi & Yu, 2008; Yu et al., 2008). The similarity-based approach argues that people treat shared labels the same as other shared features (Sloutsky & Fisher, 2004; Sloutsky, Kloos & Fisher, 2007). Our findings suggest that the naïve assumption and similarity jointly influence the labeling effect depending of types of categories (living things vs. man-made objects) and the amount of knowledge (number of listed exemplars). For living things, the labeling effect was strong and relatively independent from the amount of knowledge; for man-made objects however, the labeling effect was weak and depended on the amount of knowledge. The results suggest that people are willing to evaluate shared category labels as a type of feature for man-made objects more than for living things.

References

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

Bacha, W. J., Jr., & Bacha, L. M. (2000). *Color atlas of veterinary histology* (4th ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate–inanimate distinction. *Journal of Cognitive Neuroscience*, *10*, 1-34.

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

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*(6), 489-493.

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

López, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology, 32*, 251-295.

Lupyan, G. (2008). From chair to "chair": A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General, 137*, 348-369.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129*, 592-613.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, & Computers*, 37 (4), 547-559.

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review, 10*(3), 517-532.

Moss, H. E., Tyler, L. K., Durrant-Peatfield, M., & Bunn, E. M. (1998). 'Two eyes of a see-through': Impaired and intact semantic knowledge in a case of selective deficit for living things. *Neurocase, 4*, 291-310.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803-814.

Shafto, P., & Cole, J. D. (2003). Development of Categorization and reasoning in the natural world: Novices to experts, naïve similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, and Cognition*, *29*, 641-649.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133*(2), 166-188.

Waxman, S. R., & Braun, I. (2005). Consistent (but not variable) names as invitations to form object categories: new evidence from 12-month-old infants. *Cognition, 95*, B59-B68.

Yamauchi, T., & Yu, N. Y. (2008). Category labels versus feature labels: Category labels polarize inferential predictions. *Memory & Cognition, 36*(3), 544-553.

Yamauchi, T. (2009). Finding abstract commonalities of category members. *Journal of Experimental & Theoretical Artificial Intelligence, 21*, 155-180.

Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science, 22*, 189-228.

Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science, 18*, 179-185.

Yu, N. Y., Yamauchi, T., & Schumacher, J. (2008). Rediscovering symbols: The role of category labels in similarity judgment. *Journal of Cognitive Science, 9*, 89-109.

Acknowledgement

We would like to thank Amanda Hahn and Casady Bowman for their thoughtful comments.