UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Sizable Sharks Swim Swiftly: Learning Correlations through Inference in a Classroom Setting

Permalink
https://escholarship.org/uc/item/3r39d73m

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 28(28)

ISSN
1069-7977

Authors
Love, Bradley C.
Sakamoto, Yasuaki

Publication Date
2006

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Sizable Sharks Swim Swiftly: Learning Correlations through Inference in a Classroom Setting

Yasuaki Sakamoto (yasu@psy.utexas.edu)
Bradley C. Love (love@psy.utexas.edu)
Department of Psychology, The University of Texas at Austin
Austin, TX 78712 USA

Abstract

Fifth-graders' results from a category learning experiment suggest that inferring stimulus properties given the category membership leads to better acquisition of category knowledge than classifying stimulus items. Fifth-graders liked better and learned more properties of the shark categories acquired through inference than those acquired through classification. Classification promoted learning of only the property that was most diagnostic in discriminating among categories. Inference facilitated learning of all properties associated with each category, including properties not queried during training. Seven to 33 days after training, fifth-graders who inferred still had more information about properties of each category than fifth-graders who classified. Classroom teaching should emphasize reasoning from the category to multiple properties rather than from a set of properties to the category.

Category learning researchers seek to understand better how humans encode, organize, and use knowledge. Given these objectives, research in category learning should have important implications for education. However, the link between the two fields is not as solid as one might expect in part because category learning research has focused mostly on classification learning (e.g., Shepard, Hovland, & Jenkins, 1961) despite the fact that humans learn through a variety of interactions with their environment. As shown in Figure 1, in classification learning, participants predict the category membership of a given stimulus item and then receive corrective feedback. The focus on classification learning has advanced the development of theories that predict how people classify stimulus items in the laboratory. Unfortunately, many of these theories do not generalize to category learning in the real world, such as in a classroom.

More recent work has began to address the limitation of focusing on a single learning task by comparing learning and transfer performances for different category learning tasks, such as inference and classification learning (see Markman & Ross, 2003 for a review). Inference learning is closely related to classification learning. As shown in Figure 2, in inference learning, participants predict a single unknown property of an item given the remaining properties and the item's category membership. Different properties are queried on different inference learning trials. Inference learners receive the same corrective feedback as classification learners consisting of the category label and all perceptual properties.

In the present work, findings from category learning research with adults comparing classification and inference learning are extended to fifth-grade children with class related materials. Many classroom exercises can take the forms of classification and inference learning. For instance, children may learn about different animals in a science class by

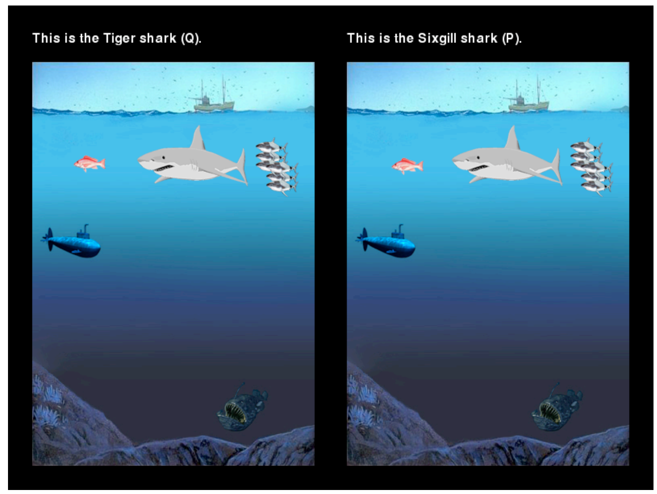

Figure 1: A classification learning trial is shown. The description on the left differs from the description on the right only on the category label (in this case, Tiger vs. Sixgill shark).

classifying a series of animals presented to them. In the next class session, children may infer properties associated with the animals whose category memberships they already know.

Consideration of related work in category learning leads us to advance that inference learning should result in better acquisition of category knowledge in a classroom than classification learning. To foreshadow our results, children liked better and learned more properties of the shark categories acquired through inference than those acquired through classification. Whereas children acquired only the most diagnostic property discriminating among categories when they learned through classification, they acquired all properties associated with each category, including properties not queried during training, when they learned through inference.

Related Work in Category Learning

The basic finding from work with adults is that the difference between classification and inference (see Figure 1 and Figure 2) leads participants to focus on different sources of information (e.g., Chin-Parker & Ross, 2004; Yamauchi & Markman, 1998). Whereas classification learners focus mostly on diagnostic information that discriminates among categories, inference learners focus on each category's prototype.

For example, Chin-Parker and Ross (2004) asked classi-

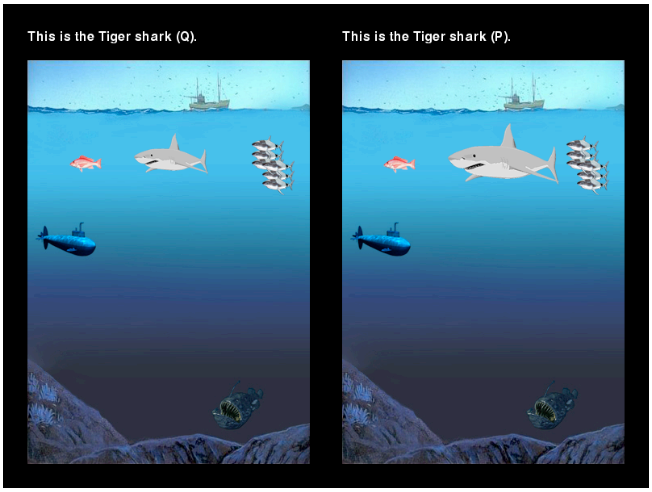

Figure 2: An inference learning trial is shown. The description on the left differs from the description on the right only on the queried dimension (in this case, smaller vs. larger body size).

Table 1: The abstract category structures used in the current experiment

Training item	Dimension value	Novel item	Dimension value
A1	21111	N1	21112
A2	12112	N2	12111
A3	12211	N3	11112
A4	11221	N4	12121
A5	11122	N5	11111
B1	12222	N6	12221
B2	21221	N7	21222
B3	21122	N8	22221
B4	22112	N9	21212
B5	22211	N10	22222

Note. During training, participants classified the training items or inferred the middle three dimensions of the training items. Both training and novel items were presented in the typicality phase.

fication and inference learners, following training, to chose which one of two choices was more typical of a given category, to draw each type of bug stimulus from training, and to rate typicality of each category member. In all these measures, classification learners' responses were based on diagnostic information, whereas inference learners' responses were based on both diagnostic information and information typical of each category.

A related result is that a family resemblance structure (Rosch & Mervis, 1975) is easier to master through inference than classification learning (Yamauchi & Markman, 1998). Table 1 shows an example of a family resemblance structure. In accord with the rich structure of natural categories, value 1 tends to co-occur with value 1 across the five dimensions, just as having wings tend to co-occur with flying, but no single dimension defines the categories. Each category's prototype in inference is useful in learning about a family resemblance structure because the category prototypes successfully distinguish members of the contrasting categories. However, prototype knowledge in inference learning interferes with learning about a nonlinear category structure, in which the category prototypes do not help in discriminating members of different categories (Yamauchi, Love, & Markman, 2002).

Furthermore, uncovering each category's prototype during inference learning should support subsequent classification learning because the knowledge of the prototypes should be sufficient to determine the category memberships when categories have family resemblance structures. In contrast, knowledge about the diagnostic information in classification learning should not promote later inference learning because the diagnostic information is not appropriate for inference learning. Indeed, Yamauchi and Markman (1998) found that inference learning-then-classification learning was easier to master than the other ordering.

These studies with adults have clear implications for classroom learning. Many categories children study at school likely have family resemblance structures. Thus, learning about categories in a classroom will involve discovering how properties are correlated within categories, which will be easier through inference learning. Though ease of learning does not necessarily mean ease of processing, if people find things more attractive that are easier to process (Rolf, Schwarz, & Winkielman, 2004), it is possible that inference learning could increase children's liking of learned materials and thus motivation to learn. Moreover, prototype knowledge in inference should contain more information about properties associated with each category than knowledge about diagnostic properties in classification learning.

Two studies have examined the effects of category use on children's category knowledge (Hayes & Younger, 2004; Ross, Gelman, & Rosengren, 2005) and showed that, like in adults' (e.g., Ross, 1999), making inferences plays an important role in children's acquisition of category knowledge. These studies however did not compare category knowledge resulting from classification and inference learning.

Current Experiment

Participants learned about the categories under the heading Training item in Table 1 through classification or inference learning. Each item (e.g., item A1) has 5 binary-valued dimensions (e.g., 21111) – each column under the heading Dimension value in Table 1 is a dimension. For example, the first column might be the size dimension, where value 1 could be small and value 2 could be large. If this was the case, item A1 was large. The five dimension values were mapped to five properties of animated sharks as described in Materials in Method. Training items A1–A5 belong to category A, whereas items B1–B5 belong to category B.

The categories are defined by both family resemblance and rule-plus-exception structures. The modal prototype of category A is 11111 (i.e., Novel item N5) and the modal prototype of category B is 22222 (i.e., N10). The first dimension is the most diagnostic in distinguishing members of the two categories as 4 of 5 items in each category follow an imperfect regularity on the first dimension. For instance, A2–A5 displays value 1 on their first dimensions and B2–B5 displays value 2 on theirs.

Special training items A1 and B1 in Table 1 are of interest because these items violate the most diagnostic regularity on the first dimension. At the same time, the special items display the category-typical values on the remaining four dimensions and are overall the most typical training members. The **regular** training items (i.e., A2–A5 and B2–B5) display the category-typical values on the first dimension and on two of the remaining four dimensions (see Table 1). Classification learners should focus on the first dimension and make more errors during training on the special than the regular items. In contrast, inference learners should focus on the prototypical nature of the special items.

Typicality ratings for the special and regular items, after training, should vary for classification and inference learners. The special items should be poor category examples for classification learners as these items violate the regularity on the first dimension, whereas the special items should be good category examples for the inference learners as these items are most typical training items. Furthermore, the inference learners should rate the novel items as better examples of the categories than the studied items. The novel items contain the category **prototype**s and **other** items that overall display more category-typical values than the studied items (see Table 1). Classification learners should be only sensitive to information about the first dimension.

The values of only three of 5 properties were queried in the current experiment (cf. Anderson, Ross, & Chin-Parker, 2002). If inference learning focuses learners on each category's prototype, inference learners should acquire the category-typical values of all five dimensions. In contrast, if inference learners are merely memorizing each category's correct value for each queried dimension, they will not learn about the non-queried properties. Classification learners will acquire the category-typical value of only the first dimension. Participants' knowledge about the sharks' 5 properties was measured a few minutes and a few weeks after training.

Method

Participants Twenty-eight fifth-graders from St. Francis School at Austin [1] participated in the experiment as computer-based exercises during a science class at St. Francis School. In addition, 54 University of Texas undergraduates participated as a control and for developmental comparisons.

Materials Each stimulus was animated picture describing a shark. Animations were used as they are common in educational materials (e.g., Lowe, 2003). The 5 binary-valued dimensions were habitat (near the surface or bottom), diet (fish or shrimp), litter size (a few or many pups), body size (small or large), and shade (light or dark). The five dimensions were mapped randomly onto the logical structure shown in Table 1. For example, for some participants the first dimension was the habitat dimension, for others it was the diet dimension. The dimension values were assigned according to the properties of the sharks used in the experiment as described below.

One set of categories contrasted Sixgill and Tiger sharks. The Sixgill sharks tend to be in deep water, feed on various animals but often on shrimp, give birth to 22 to 108 pups, are 1.5 to 5 m long, and have dark body shade. In contrast, the Tiger sharks go from the surface to 340 m, feed on anything but often on fish, deliver 10 to 80 pups, are 3 to 6 m in size, and have lighter shade.

In Table 1, value 1 on each dimension signifies the value common to the Sixgill sharks when category A is the Sixgill shark. Item N5 in Table 1 is a typical Sixgill shark that displays the category-typical values on all five dimensions (i.e., lives near the bottom, eats shrimp, delivers many pups, is small, and has dark shade). Item N10 is a typical Tiger shark. Participants were informed that the sharks vary in their properties and the two categories' members could display overlapping properties. Another set of categories, Greenland vs. Soupfin sharks, was prepared in a similar fashion.

Design and procedure The fifth-graders and the undergraduates learned about the shark categories described in Materials. Whereas each undergraduate completed a single session, each fifth-grader completed two sessions. In the initial session, the participants were randomly assigned to either the classification or the inference training condition, consisting of the familiarization, training, interruption, test, and typicality phases. On average 20 days (the range was 7 to 33 days) after completing the initial session, the fifth-graders (but not the undergraduates) completed a second session, in which they learned about a different set of sharks through a different learning mode. For example, fifth-graders who learned about the Sixgill and Tiger sharks though classification learning in the initial session learned about the Greenland and Soupfin sharks through inference learning in the second session. The second session featured a retention phase after the same 5 phases as the initial session.

Participants were familiarized with the sharks' 5 dimensions by the sequential presentation of 15 pairs of stimuli, in which selected whether the left or right stimulus correctly depicted the queried property.

Then, participants completed six blocks of training trials. In each training block, the training items in Table 1 were presented sequentially once in a random order. On each trial, two descriptions of the sharks were presented side by side. In classification, the two descriptions were identical to each other except for the category label (see Figure 1). In inference, the two descriptions were identical to each other except for the value of a queried dimension (see Figure 2). On each inference trial, participants predicted one of the middle three dimensions' values. Exception values (e.g., value 2 of item A1) were never queried as in most inference learning procedures (see however Nilsson & Olsson, 2005). Participants in both conditions predicted whether the left or right description was correct description and received the same corrective feedback.

Participants then observed an animation of 12 sharks swimming in the ocean one by one to prevent rehearsal of information from the training phase.

Then, participants were tested on two-alternative forced-choice questions about the properties of the two categories from the training phase without corrective feedback. Multiple-choice tests are still commonly used in educational

[1] See http://www.stfrancis-school.org/ for descriptions of the school.

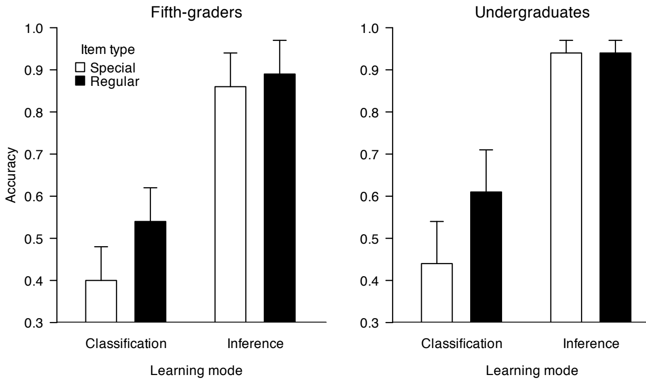

Figure 3: Mean training phase accuracies are shown. Error bars represent the upper bounds of the 95% confidence intervals (see Loftus & Masson, 1994).

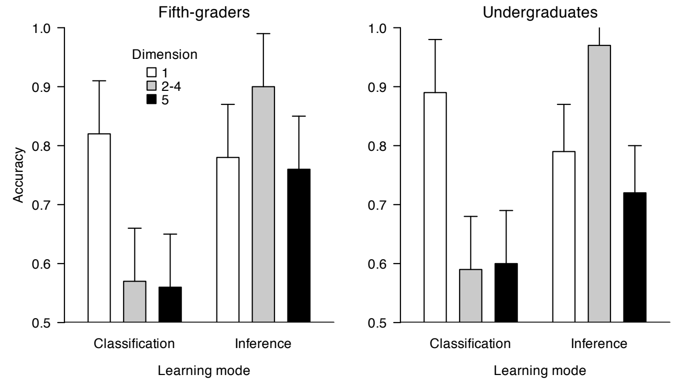

Figure 4: Mean test phase accuracies are shown.

settings (e.g., Roediger & Marsh, 2005). For example, "Tiger sharks:" was presented above "A: tend to be smaller" and "B: tend to be larger" to query the size of the Tiger shark. All five dimensions were queried for the two categories. Another set of 10 questions was queried in the opposite fashion (e.g., "tend to be larger:" was displayed above "A: Tiger sharks" and "B: Sixgill sharks"). These 20 questions were sequentially presented once in a random order.

After the test phase, all items in Table 1 were sequentially presented once in a random order in the typicality phase. Participants moved a red ball on a continuous (300-point) rating scale to indicate how good an example they think the item is of the category. The ends of the scale were labeled VERY GOOD and VERY POOR (randomly assigned for each participant).

In the second session, fifth-graders completed a retention phase, which was identical to the test phase from the initial phase, following the 5 phases described above. After the retention phase, the experimenter asked 11 fifth-graders "which sharks did you like better, those from this time or those from the last time?". For 6 fifth-graders the initial session was classification and the second session was inference, for others it was the other order.

Results

All participants were included in the analyses. We only report analyses relevant to our predictions. For fifth-graders, no significant effects involving session (initial or second) were found in all analyses.

Training As shown in Figure 3, as predicted, fifth-graders were significantly more accurate on the regular than special items, $t(27) = 2.32$, $p < .05$ when they learned through classification, whereas they were not when they learned through inference ($t < 1$). Like the fifth-graders, the undergraduates in classification were significantly more accurate on the regular than special items, $t(26) = 2.55$, $p < .05$, whereas the undergraduates in inference were not ($t < 1$).

Test For analyses, the five shark dimensions were grouped into D1 (first dimension), D2–4 (second, third, and fourth dimensions), and D5 (fifth dimension). For example, when the first dimension was size, D1 included the test trials involving "large/small." The answer is "correct" in the test phase when participants selected the property which is typical of the given category or the category for which the given property is typical. D1 and D5 were never queried in inference. D1 was the most diagnostic dimension. Accuracy data for D2–4, which were queried in inference, were collapsed.

As shown in Figure 4, as predicted, fifth-graders were more accurate on D1 than D2–4 and D5 when they learned through classification, $t(27) = 4.18$, $p < .001$ and $t(27) = 2.83$, $p < .01$, respectively. D2–4 and D5 did not differ significantly ($t < 1$). When fifth-graders learned through inference, they were more accurate on D2–4 than D1, $t(27) = 2.08$, $p < .05$. The difference between D2–4 and D5 did not reach significance, $t(27) = 1.69$, $p = .10$. D1 and D5 did not differ significantly ($t < 1$). As predicted, when fifth-graders learned through inference, they performed significantly above chance for the non-queried D1 and D5, $t(27) = 4.77$, $p < .001$ and $t(27) = 3.71$, $p < .001$, respectively. For comparison, when fifth-graders learned through classification, they performed significantly above chance for D1, $t(27) = 6.97$, $p < .001$, but not for D5 ($t < 1$).

As shown in Figure 4, undergraduates showed a similar pattern to fifth-graders. As predicted, the undergraduates in classification were more accurate on D1 than D2–4 and D5, $t(26) = 4.97$, $p < .001$ and $t(26) = 2.67$, $p < .05$, respectively. D2–4 and D5 did not differ significantly ($t < 1$). The undergraduates in inference were more accurate on D2–4 than D1 and D5, $t(26) = 2.55$, $p < .05$ and $t(26) = 3.41$, $p < .01$, respectively. Their accuracies for D1 and D5 did not differ significantly ($t < 1$). Like the fifth-graders, the undergraduates in inference performed significantly above chance for D1 and D5, $t(26) = 3.82$, $p < .001$ and $t(26) = 3.02$, $p < .01$, respectively, whereas the undergraduates in classification performed significantly above chance for D1, $t(26) = 7.98$, $p < .001$, but not for D5, $t(26) = 1.17$, $p = .25$.

Typicality As shown in Figure 5, the ratings for special, regular, prototype, and other items did not differ significantly when fifth-graders learned through classification ($F < 1$). As predicted, when fifth-graders learned through inference, they rated the special items as better examples than the regular items, $t(27) = 2.38$, $p < .05$, and the prototype items as better

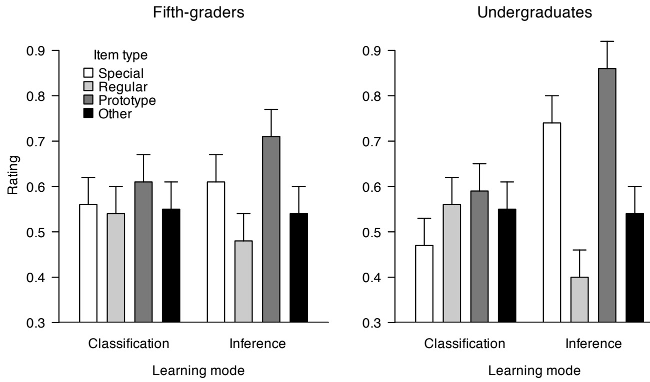

Figure 5: Mean typicality ratings (0: atypical, 1: typical) are shown.

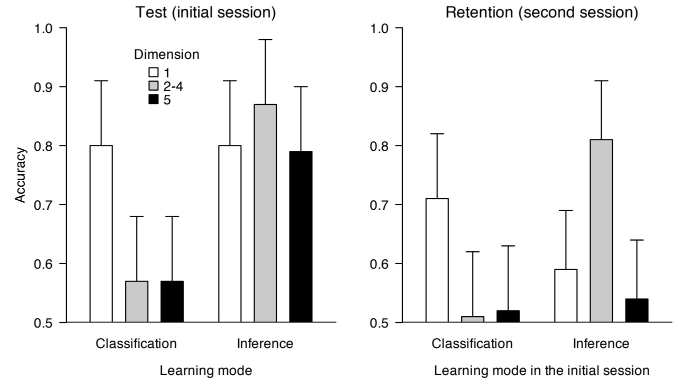

Figure 6: Fifth-graders' mean test phase accuracies (left figure) and the same students' retention phase accuracies (right figure) are shown.

examples than the other novel items, $t(27) = 3.66$, $p < .01$. The special and prototype items did not differ significantly, $t(27) = 1.48$, $p = .15$. As predicted, the other novel items were rated as better examples than the regular studied items when fifth-graders learned through inference, $t(27) = 2.09$, $p < .05$.

For the undergraduates in classification, although the special and regular items did not differ significantly, $t(26) = 1.71$, $p = .10$, the prototype items were rated as better examples than the special items (see Figure 5), $t(26) = 3.52$, $p < .01$. As predicted, no other comparisons approached significance for classification learners ($t < 1$ for both comparisons). For the undergraduates in inference, as predicted, the special items were rated as better examples than the regular items, $t(26) = 6.65$, $p < .001$, and the prototype items were rated as better examples than the other novel items, $t(26) = 6.96$, $p < .001$. Moreover, the prototypes were rated as better examples than the special items, $t(26) = 2.22$, $p < .05$. As predicted, the undergraduates in inference rated the other novel items as better examples than the studied regular items, $t(26) = 5.96$, $p < .001$.

Retention and shark preference Figure 6 shows the fifth-graders' accuracies in the initial session's test phase (left side) and the same participants' accuracies in the second session's retention phase (right side). For analyses, the fifth-graders are grouped according to the learning mode in the initial session. The fifth-graders in inference in the initial session were significantly more accurate on D2–4 in the retention phase than those in classification in the initial session (see the right side of Figure 6), $t(26) = 4.42$, $p < .001$. However, the two groups did not differ significantly in their retention phase accuracies on D1 and D5 ($t < 1$ for both comparisons). The fifth-graders preferred (10 of 11) sharks learned through inference to sharks learned through classification, exact binomial $p = .01$ (two-tailed).

General Discussion

The current experiment demonstrates that inference learning benefits classroom learning by leading to better acquisition of category knowledge and better liking of categories. Children and adults learned about two shark categories through inference or classification learning. The categories were defined by both family-resemblance and rule-plus-exception structures. A special item from each category violated the regularity on the most diagnostic dimension but possessed category-typical values on all of the other dimensions. These items resulted in more errors during training in classification learning but did not in inference learning. The special items were rated as better examples than regular items by inference learners but were not by classification learners.

Whether inference learners are forming a prototype of each category or constructing a set of rules for each category (cf. Johansen & Kruschke, 2005) was examined by querying only three of five properties during training. Although inference learners had more knowledge about the queried than the non-queried properties, they showed some learning of the non-queried properties (cf. Anderson et al., 2002), consistent with the idea that inference learners are forming a prototype of each category.

These results suggest that in accord with the findings from prior work with adults, whereas classification learning encourages the acquisition of only the information that discriminate among categories, inference learning facilitates the acquisition of all properties associated with each category, including to some extent properties not queried during training. The acquired properties guided the processing of subsequent experiences, especially for older population. Information that distinguished among categories became central for classification learners, whereas prototype information for each category became central for inference learners. Though inference learners' knowledge about the non-queried dimensions were lost after a few weeks, they still possessed more knowledge than classification learners. Further, children liked better shark categories learned through inference than those acquired through classification.

Implications and Future Research

The findings from the present experiment suggest that classroom exercises should emphasize reasoning from the category to multiple properties rather than from a set of properties to the category. Though the benefit of inference learning is clear in the present work, both classification and inference learning operate in the real world. Thus, it is important to examine the interaction between these two learning modes (e.g., Hayes & Younger, 2004; Ross et al., 2005). In the current work, half the fifth-graders completed classification-then-inference. The other half completed in the other order. Although no effects of task ordering were found in the present work, there was a relatively long delay between the first and second session. Future work should examine whether the inference-then-classification advantage observed in adults (e.g., Yamauchi & Markman, 1998) applies to classroom learning with children.

Furthermore, children's learning likely involves unsupervised learning, in which there is no teacher-provided corrective feedback, as well as supervised learning, in which corrective feedback is provided as in the classification and inference tasks in the present work. Previous work with adults suggest that there are important differences between unsupervised and supervised learning (Love, 2003). Future work should examine how unsupervised learning interacts with other learning modes in a classroom. For example, including unsupervised learning trials within supervised inference and classification learning may have differential effects on children's learning. Using the current category structures, an unsupervised learning trial, in which a stimulus is simply shown with its category label, will likely facilitate inference learning but may actually hinder classification learning. The unsupervised learning trials will reinforce the prototype representations of inference learners and may improve their acquisition and retention of properties that are not queried during training. Classification learners will be adversely affected by the unsupervised learning involving the deviant (special) items.

Final Note

Although the current work's focus was to apply the findings from category learning research to classroom learning, the present results have important implications for work in category leaning. Work in category learning typically offers a great deal of experimental control, but it is not clear whether such work has ecological validity. The work presented here showed that laboratory findings could be successfully applied to a real world classroom setting. The present results will allow us to conduct new laboratory experiments that explore factors that could enhance classroom learning, instruction, and assessment. Laboratory findings will be tested in the classroom to examine their applicability in the real world. Work along this line should benefit both education and research.

Acknowledgments

This work was supported by AFOSR FA9550-04-1-0226 and NSF CAREER 0349101 to B. C. Love. We thank Momoko Sato for her help in preparing materials used in the experiment and collecting data.

References

Anderson, A. L., Ross, B. H., & Chin-Parker, S. (2002). A further investigation of category learning by inference. *Memory & Cognition*, *30*, 119–128.

Chin-Parker, S., & Ross, B. H. (2004). Diagnosticity and prototypicality in category learning: A comparison of inference learning and classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 216–226.

Hayes, B. K., & Younger, K. (2004). Category-use effect in children. *Child Development*, *75*, 1719–1732.

Johansen, M. K., & Kruschke, J. K. (2005). Category representation for classification and feature inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1433–1458.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476–490.

Love, B. C. (2003). The multifaceted nature of unsupervised category learning. *Psychonomic Bulletin & Review*, *10*, 190–197.

Lowe, R. K. (2003). Animation and learning: selective processing of information in dynamic graphics. *Learning and Instructions*, *13*, 157–176.

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

Nilsson, H., & Olsson, H. (2005). Categorization vs. inference: Shift in attention or in representation? In *Proceedings of the 27th annual conference of the cognitive science society.* Stresa, Italy: Cognitive Science Society.

Roediger, H. L. I., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1155–1159.

Rolf, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality & Social Psychology Review*, *8*, 364–382.

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

Ross, B. H. (1999). Postclassification category use: The effects of learning to use categories after learning to classify. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 743–757.

Ross, B. H., Gelman, S. A., & Rosengren, K. S. (2005). Children's category-based inferences affect classification. *British Journal of Developmental Psychology*, *23*, 1–24.

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*(13, Whole No. 517).

Yamauchi, T., Love, B. C., & Markman, A. B. (2002). Learning nonlinearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 585–593.

Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language*, *39*, 124–148.