Diagnosticity in Category Learning by Classification and Inference

Seth Chin-Parker and **Brian H. Ross**
(chinpark@s.psych.uiuc.edu) (bross@s.psych.uiuc.edu)
Beckman Institute, 405 North Matthews Avenue
University of Illinois, Urbana-Champaign
Urbana, IL 61801 USA

Abstract

Categories are learned in many ways, but the focus of much category learning research has been on classification learning. In classification learning, the diagnosticity of features is a primary influence on learning and the category representation. In this paper, we assess this influence of diagnosticity on a different means of category learning, inference learning. In two experiments, each with a different dependent measure, we found the expected result that classification learning led to strong sensitivity to the diagnosticity of the features, even to the exclusion of prototypicality (when controlled for diagnosticity). However, inference learners were significantly less sensitive to the diagnostic value of the features, and were sensitive to the prototypicality. This result provides further evidence for the idea that different types of category learning differentially influence the category representation and provides a better understanding of inference learning.

Introduction

Categories are critical for a wide variety of cognitive tasks, such as classification, inference, explanation, communication, and problem solving. Category learning reflects how it is that people acquire knowledge of the categories that will successfully support these uses. Any intelligent system that extends information from specific examples to other related occurrences needs to account for the processes related to category learning. Thus, developing an understanding of category learning is an important research endeavor in cognitive science.

Although categories may be learned in a number of ways, the focus of category learning research has been on classification, how items are assigned to categories (e.g., Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986). In classification learning, a subject is shown an item, asked to indicate its category membership (usually from a set of two possible categories), given feedback on their choice, and then allowed to study the item before the next item is presented. Through the learning trials of this classification task, the subject learns what items go into what category, thus developing a category representation that can be used later to answer questions about the category members or to classify novel items.

Category learning, however, does not just consist of classification learning. We learn categories for a variety of purposes, and how we learn categories is often tied to these uses. Category learning is based on not just classification, but on inference or explanation or problem solving (among other possibilities). A complete understanding of category learning requires considering additional category learning tasks and how they influence the category representation. The idea underlying this research is that different ways of learning about categories lead to different category representations, and that our real-world representations often derive from a variety of ways in which categories are learned and used. Although this notion has a certain intuitive appeal, only a small number of category learning studies have examined it.

One category learning task that has received attention over the last few years is inference learning (Anderson, Ross & Chin-Parker, 2002; Yamauchi, Love & Markman, 2002; Yamauchi & Markman, 1998, 2000). In this task, a classified item is presented with a category feature missing and the task of the learner is to choose the appropriate feature for that item. For example, if one were making inferences about different types of birds, one might be given a classified bird (e.g., yellow-rumped warbler) with a number of its features and asked to choose its food preference. Inference is a critical component of category use. Since people learn about categories as they make inferences and receive feedback on their predictions, inference learning is a natural direction to follow in category learning research. This task has also been the focus of recent research because it has many similarities to classification learning and is formally equivalent to classification if the category label is treated simply as another feature (Yamauchi & Markman, 1998).

Diagnosticity

A critical aspect of current theories of classification learning is the focus on the *diagnosticity* of the features, how predictive they are of category membership (Tversky, 1977). As people learn to classify category members, they learn to attend more to those features that help to distinguish the categories (e.g., Kruschke, 1992). In the simple case in which the categories may be distinguished on the basis of a single feature, all the attention may be focused on that feature. For more complex cases, the attention is distributed across the diagnostic features to maximize classification performance.

Inference learning, however, may not lead to such an exclusive attention to diagnostic features. In inference learning, the item is already classified, so the learner can focus on *that* category and what features occur with members of that category. This focus on a single category at a time makes information about the prototypical feature values more available (Anderson et al., 2002; Yamauchi & Markman, 1998, 2000). This information about the prototypical feature values for each category means that the category representation emphasizes the internal structure of the category, what it is that coheres the members of the category. This focus during inference learning suggests that different information about the category would be acquired during learning when compared to classification learning.

Current Experiments

The goal of the current experiments is to investigate the role of diagnosticity in category learning with two different category learning tasks, classification and inference. Based on a large body of previous research, we expect that diagnosticity will be the primary influence in classification learning. Thus, the question of most interest is how inference learning is affected by feature diagnosticity (versus the internal structure of the category). The hypothesis is that inference learning will not lead to as strong an influence of diagnosticity as does classification learning. This hypothesis is of importance for two reasons. First, it questions a major assumption of current models of category learning, that the diagnosticity of features is the most important determinant of category learning. Second, it helps to provide a further understanding of another type of category learning, inference learning.

We used a common category structure, family resemblance, as shown in Table 1. In this structure, the prototype is chosen and the learning items from that category consist of items that are similar to the prototype, though they may be different from one another. In the experiments reported here, all the learning exemplars match the prototype on all but one of the features.

The diagnosticity of the features is manipulated by varying the overlap of the prototypes. With this manipulation, to be described in detail for each experiment, we could separately vary the prototypicality of the item (i.e., how similar it was to the prototype, reflecting the internal structure) and the diagnosticity of the features (in terms of how predictive they were of category membership, reflecting the relation of the two categories).

Based on results from previous studies, we can anticipate some of the results of these experiments. Classification learners should show a strong effect of diagnosticity, but not prototypicality. We also know the inference learners should show a sensitivity to prototypicality. The question remains to be answered as to whether the inference learners will show any effect of diagnosticity. If they do show a sensitivity to diagnosticity, it should be significantly less than that of the classification learners.

Experiment 1

Experiment 1 investigated the influence of diagnosticity on classification and inference learning with a forced-choice test at transfer. The critical test trials varied the diagnosticity and typicality to examine the influence of the learning conditions. The categories learned were fictional "bugs".

The manipulation of diagnosticity as a function of prototype overlap can be seen in Table 1. The target category is the one on the left, the Deegers (prototype 11111, indicating a particular set of values for each of the five binary dimensions). Along with this category, subjects either learned the Lokads (prototype 00011) or the Koozles (prototype 11000). Those features common to both prototypes are not diagnostic because they do not help one

Table 1: Category Structure for Experiment 1.

	Deeger	Lokad	Koozle
Learning Exemplars	11110	00010	11001
	11101	00001	11010
	11011	00111	11100
	10111	01011	10000
	01111	10011	01000
Prototype	11111	00011	11000

to determine category membership. As can be seen, both contrast categories overlapped Deegers on two features, but two different features. Thus, for subjects learning Deegers and Lokads, the last two features were not diagnostic, whereas the first two were not diagnostic for the subjects learning Deegers and Koozles. By varying the test features for Deegers, we can determine the extent to which diagnosticity is being used, as described shortly.

Method

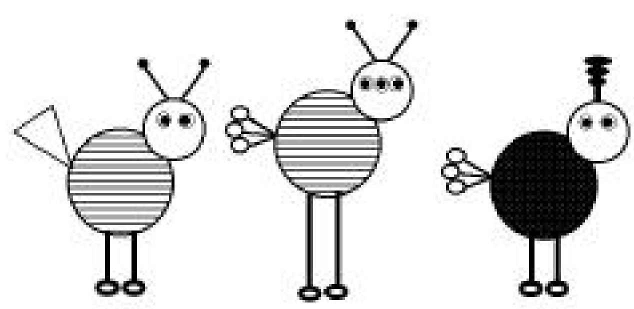
Design There were two learning conditions, classification learning and inference learning. Within each learning condition half of the subjects learned about Deegers and Lokads, and the other half learned about Deegers and Koozles. The position of the bug attributes within the category structure was balanced across subjects. This resulted in a total of ten experimental groups. Within the study and transfer blocks, items were randomly presented.

Subjects Fifty undergraduates from the University of Illinois participated for either course credit or pay. Ten subjects who did not meet the learning criterion (two classification, eight inference) were replaced.

Materials The materials were drawings of "bugs" labeled Deegers, Lokads or Koozles (see Figure 1). These bugs varied along the binary attributes of their antenna, legs, tail, body coloring and eyes.

Each subject learned two categories, the target category (Deegers) and one of the two non-target categories (either Koozles or Lokads). The abstract structures for the three categories can be found in Table 1. Each of the non-target category prototypes overlapped with the target category prototype on two features. The features that overlapped were

Figure 1: Example Prototypes for Bug Categories in Experiment One



Deeger Lokad Koozle

not predictive of category membership as they occurred equally often in each category. The remaining three features were diagnostic in that they occurred 80% of the time with one of the categories. Whether a particular feature was an overlap or diagnostic feature depended on which of the two non-target categories was learned with the Deeger category. For example, the Deeger specified by 11110 was seen by all subjects during learning. For a subject that learned the Lokad category along with the Deeger category, the first feature was a diagnostic feature, in that it was consistent with the Deeger prototype but not the Lokad prototype. For a subject that learned the Koozle category along with the Deeger category, the first feature was an overlap feature because it occurred in both the Koozle and Deeger prototypes. This design of the category structures allowed the same item to vary in terms of its diagnosticity between subjects, depending only on the category set learned, while the item prototypicality remained constant.

Table 2 specifies the terminology used to describe the bugs. The bugs varied on two dimensions, their relation to the category prototype (prototypicality) and the number of diagnostic features maintained. Along the first dimension, there were three possible levels: prototype, typical and atypical, which indicated whether there were 5, 4, or 3 features consistent with the prototype. For the second dimension, the item could maintain 1, 2, or 3 of the diagnostic features. For a subject that learned the Deeger-Lokad combination, 11110 would be a typ3 Deeger: it is typical because it has four features consistent with the prototype 11111, and it has a diagnosticity of 3 because it has all three of the diagnostic features for Deeger (111--). However, for a subject that learned the Deeger-Koozle combination, 11110 would be a typ2 Deeger: again it is typical because it has four of the prototype features, but now it only has 2 of the diagnostic features (--111).

The study items were the five typical bugs for each category being learned. For the classification subjects, these bugs were always seen complete. In the inference condition, the bugs were missing one feature that was consistent with the prototype. The inference subjects also saw the two possible features that were choices for the missing feature.

The transfer items were pairs of bugs from a given category. There were two critical types that allowed an examination of the influence of diagnosticity: typ/typ and typ/atyp items. The six typ/typ transfer items were pairs of bugs that were matched in prototypicality (both were typical) but varied in the number of diagnostic features. One of the bugs in the pair would be a typ2, while the other was a typ3, the specification being dependent on which category

Table 2: Abbreviations for Items

	# of Prototype Consistent Features	# of Diagnostic Features Maintained
Proto	5	3
Typ3	4	3
Typ2	4	2
Atyp3	3	3
Atyp2	3	2
Atyp1	3	1

set had been learned. The typ/atyp pairings varied both in terms of how many features were consistent with the category prototype and the number of diagnostic features. Of the six typ/atyp transfer items, three of the pairs consisted of a typ2 and an atyp3. These transfer items pitted typicality to the category prototype against the diagnosticity of the features. The other three pairs consisted of a typ3 and an atyp1. For these pairs, both typicality and diagnosticity were in agreement as to which bug was the better category member. The content of the typ/atyp pairs was also dependent on the category set learned; the typ2/atyp3 pair for someone who learned Deeger-Koozle was the typ3/atyp1 pair for someone who learned the Deeger-Lokad set, and vice-versa. The critical typ/atyp pairs are the typ2/atyp3, but the typ3/atyp1 pairs allow full counterbalancing of items.

There were nine additional filler transfer pairs in the Deeger task, as well as 12 pairs for the other category that did not address the issues of interest.

Procedure Subjects were given verbal instructions prior to the study phase. All subsequent instructions and reminders appeared on the computer screen. Learning and testing were done on Macintosh computers using Psyscope (Cohen, MacWhinney, Flatt, & Provost, 1993). All subjects were debriefed both verbally and with a written statement as to the general intent of the experiment.

In the classification condition, subjects saw a bug presented in the center of the screen. Subjects indicated which category they thought the bug belonged to by pressing the “D” key for Deeger, the “K” key for Koozle or the “L” key for Lokad. Feedback was given as to whether the choice was correct or incorrect, and the subject was shown the bug again along with the correct name to study. This study time was self-paced. The learning phase continued for a minimum of four blocks, ten exemplars per block, until the subject was able to correctly identify nine of the ten bugs within a block.

In the inference condition, the subjects were presented with an incomplete bug (one of the five attributes was absent from the bug picture) centered on the screen. The category label was presented to the left of the bug, and the two possible features for the missing attribute were presented on the right side of the screen, one above the other. The subject indicated a choice by clicking the mouse on one of the two features. The position of the correct feature was randomized across learning trials. Feedback was given, and self-paced study time was allowed. The learning criteria were the same as in the classification condition.

Following learning, all subjects completed the same forced choice test. No feedback was given to the subjects during the transfer phase. Each subject completed the target (Deeger) test first. Subjects saw two possible Deegers on the screen, one centered on the right-hand side of the screen and the other centered on the left-hand side of the screen. Below the pictures appeared the question, “Which of these bugs is most typical of a Deeger?” Once the subject clicked on one of the pictures, a box appeared around the choice and the prior question was replaced by “How confident are you of your choice?” along with a number scale going from one (guessing) to seven (very sure). The subject clicked on a number to indicate his or her confidence. Once the target category transfer was completed, the subject was informed that the same type of questions would be asked about the other bug category. The procedure for the non-target category transfer was identical.

Results and Discussion

For the typ/typ transfer items, the mean proportion of choice for the typ3 Deeger was calculated. Also, the mean confidence scores were determined; confidence scores when selecting the typ2 were multiplied by –1 (no preference between the bugs would result in a mean confidence of zero). These results examine how important the diagnosticity was to each condition when the prototypicality is held constant. The classification learners chose the bug with more diagnostic features well above chance (m = .79), $t(19) = 4.74$, $p < .01$, while the inference learners did not (m = .56), $t(19) < 1$. These results are also reflected in the group confidence ratings. The classification learners' confidence rating (m = 3.44) was significantly greater than 0, $t(19) = 5.38$, $p < .01$, while the inference mean confidence rating (m = 0.70), was not, $t(19) = 1.22$, $p > .10$. The proportion of the classification learners who chose the typ3 bug over the typ2 bug was significantly greater than the proportion of inference learners, $t(38) = 2.70$, $p < .01$. The mean confidence score of the classification learners was also significantly different from the score of the inference learners, $t(38) = 3.19$, $p < .01$.

For the typ/atyp items, the mean proportion of times the typical Deeger was chosen over the atypical Deeger was determined. The mean confidence rating was calculated by multiplying the confidence scores when choosing the atypical bug by –1. Of interest are the bug pairs that pitted typicality against diagnosticity, the typ2/atyp3 pairs; choosing the typ2 bug meant that overall typicality to the prototype was of primary importance while choosing the atyp3 bug indicated that diagnosticity was driving the decision. The classification learners chose the typ2 bug (m = .35) marginally less than the inference learners (m = .57), $t(38) = 1.81$, $p < .10$. However, the mean confidence score for the classification learners (-2.22) was significantly lower than that of the inference learners (0.76), $t(38) = 2.49$, $p < .05$.

For both transfer measures, the typ/typ and typ/atyp items, the classification learners showed significantly more dependence on diagnostic information than the inference learners when making decisions about the category members. In the typ/typ measure, the inference learners did not show an influence of diagnosticity. The typ/atyp results are more difficult to assess in this regard since there is no clear baseline to compare performance to. These results show a clear difference in the role of diagnosticity for the two different types of category learning.

Experiment 2

Experiment 2 examined the same issues, but had three differences from Experiment 1 to increase generality and the number of critical observations. First, during transfer, a typicality rating task was used rather than a forced-choice task. Second, to allow us to use all the transfer data to test

Table 3: Category Structure for Experiment 2

	Deeger	Lokad	Koozle	Himlit
Learning	11110	00010	11001	10100
Exemplars	11101	00001	11010	10111
	11011	00111	11100	10001
	10111	01011	10000	11101
	01111	10011	01000	00101
Prototype	11111	00011	11000	10101

the hypothesis, we changed the design so that each category had a critical contrast (see Table 3). As before, one subject might learn Deegers and Lokads, whereas another learned Deegers and Koozles. However, by adding two more counterbalancing groups (Himlit and Lokads, Himlit and Koozles), all the categories had critical test items. Third, although the items were again fictitious bugs, new features were constructed.

Method

Design As in Experiment 1, there were a classification and an inference condition. There were four possible category combinations a subject could learn [Deeger-Koozle, Deeger-Lokad, Himlit-Koozle, Himlit-Lokad]. Within each possible combination, there was a balancing as to the order of the categories presented during the typicality rating task. The presentation of items within both study and transfer blocks was random.

Subjects Sixty-one undergraduates from the University of Illinois participated for either course credit or pay. One subject's data were lost due to a computer error and 12 subjects (five inference, seven classification) were replaced who did not meet the learning criterion.

Materials The materials for this experiment were again drawings of "bugs" (Deegers, Himlits, Koozles and Lokads) that consisted of five binary attributes: legs, wing, eyes, antenna, and tail. Across subjects, each attribute was balanced as to whether it served as an overlap or diagnostic attribute. The bugs seen during learning were the five typical bugs (one feature of each was not consistent with the prototype) from each of the two categories being learned, and they were presented as in Experiment 1.

The typicality rating task consisted of 16 bugs for each category learned. Five of the bugs were the typical bugs (one inconsistent feature each) which had been seen during learning. The other 11 bugs were novel to the subject at the time of the typicality rating task. These bugs consisted of the category prototype along with 10 atypical bugs, each of which had two features that were inconsistent with the prototype. The levels of prototypicality and diagnosticity were specified as in Experiment 1 (refer to Table 2).

Procedure The procedures during the learning phase for this experiment were very similar to those in Experiment 1. The primary differences were that the classification subjects used the mouse to click on the category label and the study time was restricted to two seconds per bug. The learning criteria were the same as Experiment 1.

Following the learning phase, the subjects were given instructions on the computer screen explaining the typicality rating task. During this task, a single bug appeared centered on the screen along with the question, "How typical is this bug of a [Deeger, Himlit, Koozle, Lokad]?" Underneath the picture were the numbers from one ("Not at all typical") to seven ("Very typical"). The subject indicated their rating of each bug by clicking the mouse on one of the numbers.

Results and Discussion

What influences the typicality ratings for the two learning groups: diagnosticity, prototypicality or both? The short answer is that the ratings of classification learners were influenced only by diagnosticity, whereas both diagnosticity and prototypicality affected the ratings of the inference learners.

The group means for each of the bug types are provided in Table 4. The effect of diagnosticity becomes clear when the column means are examined. Collapsing across the levels of prototypicality, it is evident that classification learners showed a large effect of diagnosticity, whereas inference learners showed some effect, but a smaller one. The ANOVA supports this interpretation with no main effect of learning, $F(1, 46) < 1$, a significant main effect of diagnosticity, $F(2, 92) = 147.87$, $p < .001$, and an interaction between the factors, $F(2, 92) = 13.17$, $p < .001$. As can be seen at the column mean level (and also within each row), the loss of a diagnostic feature reduced typicality ratings almost 1.5 (on a 1-7 scale) for the classification learners, whereas it had much less of an effect on inference learners .

The effect of prototypicality requires a more careful examination. Collapsing over the diagnosticity levels, the two learning conditions show very similar row means, supported by the ANOVA indicating that there is no main effect of learning condition, $F(1, 46) < 1$, a significant effect

Table 4: Mean Typicality Ratings from Experiment 2

Classification Learners

	Number of Diagnostic Features			
	3	2	1	mean
Prototype	5.92	-----	-----	**5.92**
Typical	5.98	4.39	-----	**5.02**
Atypical	5.88	4.51	2.93	**4.18**
	5.94	**4.47**	**2.93**	

Inference Learners

	Number of Diagnostic Features			
	3	2	1	mean
Prototype	6.06	-----	-----	**6.06**
Typical	5.24	4.98	-----	**5.08**
Atypical	4.63	4.07	3.66	**4.00**
	5.29	**4.37**	**3.66**	

of prototypicality, $F(2, 92) = 102.96$, $p < .001$, and no interaction between the two, $F(2, 92) < 1$. However, the difference between conditions is hidden by the fact that prototypicality in this analysis is confounded with diagnosticity as can be seen in Table 4. For classification learners, within each column there is no effect of prototypicality. For example, as long as the item has all three diagnostic features, the rating given by the classification learners does not depend at all on whether it has the two non-diagnostic features consistent with the prototype (prototype, 5.92), just one of them (typical, 5.98), or neither of them (atypical, 5.88). A similar result is seen with the items that maintained two diagnostic features; on average, the classification learners actually rated the atypical items 0.12 higher than the typical items. However, inference learners show large effects of prototypicality at each level of diagnosticity. For the test items with all three diagnostic features, each non-diagnostic feature makes a difference of about 0.7. The inference learners dropped their typicality rating about 0.9 when a non-diagnostic feature was removed from the items that maintained two diagnostic features.

Thus, in Experiment 2, the typicality ratings of classification learners were influenced only by diagnosticity, whereas both diagnosticity and prototypicality affected the ratings of inference learners.

General Discussion

The diagnosticity of features plays a critical role in current theories of category learning. These experiments investigated the role of diagnosticity for two different means of category learning, classification and inference, and found an important difference. For both experiments, classification learners relied on the diagnostic features when making decisions about category members. Inference learners were sensitive to both the diagnosticity of the features (although much less so than the classification learners) and the relationship of the item to the category prototype. These results indicate that these two different ways of category learning lead to different emphases in the category representation. Yamauchi, Love, and Markman (2002), using a non-linearly separable category structure, also found that inference learners did not show an effect of feature diagnosticity, while the classification learners did, when predicting missing features of items following learning.

The results of this study rule out the strong hypothesis that the inference learners would not be at all sensitive to the diagnostic value of the features. However, this conclusion needs to await further research for two reasons. First, there were only two categories, so the probability that cross-category comparisons might be made (e.g., about the internal structure learned by inference) was probably much greater than in many more realistic situations. When we learn about items there is normally not such an obvious and closely related contrasting category being learned. Second, it is possible that some or all of this influence occurred at test. For example, the effect of diagnosticity in inference learning was most evident when the items were less typical category members. Since these items would have had more features in common with the non-target category, their presence may have prompted the inference learners to consider that other category (although it was not necessary to do given the design of the transfer tasks). Again, this seems to be much less likely in more realistic category situations. It is important to keep in mind that although the inference learners did show some sensitivity to the diagnosticity of the features, it was sometimes tiny and always significantly less than the classification learners. Despite these possibilities, it is important that future research more fully examine how the category representation is influenced by inference learning.

It is also interesting to consider the classification learners. Their lack of sensitivity to the non-diagnostic features that were part of the prototype suggests an extreme focus in the representation they develop. Other results also point out some difficulties that arise from category learning based solely on classification. Yamauchi and Markman (1998, Exp. 2) found that varying the order of classification and inference learning resulted in a situation where a block of classification learning prior to inference learning was not beneficial (although inference learning prior to classification learning was). Chin-Parker and Ross (in press) showed that classification learners were not sensitive to within-category correlations whereas inference learners were sensitive to this relational structure. Anderson et al. (2002) also found that classification learners were less accurate than inference learners when classifying individual features of category members. These lines of research suggest that classification learning may lead to a category representation that is good at determining category membership of items but is impoverished with regards to other category information.

The question remains as to why classification and inference learning lead to such different category representations. Even though the two learning tasks can be considered formally equivalent, they impose very different demands on the learner. In the classification learning task, a subject is shown an item and predicts the category label using the available information. If a piece of information is not diagnostic, such as "flying" when learning to distinguish birds and bats, it is not important and not incorporated into the category representation. The current experiments and formal modeling (Krushke, 1992) have shown that diagnosticity is the primary concern during classification learning. As noted earlier, the inference learning task focuses the learner on a single category, promoting the acquisition of information about the internal structure of that category. This would make the inference subjects sensitive to what are the most likely features given the category membership. If an item is labeled as a "bird" and then a prediction is made about how it will get from one tree to another, the correct prediction would most likely be "flying", and that piece of information is incorporated into the category representation. Recognition of the features that distinguish birds from bats is not important in this situation, so those diagnostic features would not be made salient.

A major challenge now will be to formalize the differences that exist between the various means of category learning and to incorporate that information into a category learning model. Such a model would be useful for any endeavor within the cognitive sciences concerned with learning from past experience.

In closing, it is important to remember that when we learn about categories in more realistic situations, it is by a combination of different tasks, such as classification, inference, and explanation. The limited representation that is developed as a result of classification learning does not appear to be much like the flexible, dynamic representations that underlie our knowledge of real world categories. However, the same may be true of a category representation that is developed as a result of any single category learning task. It may the combination of various learning tasks that creates a flexible and dynamic representation. To understand category learning as it exists outside of the laboratory, we are going to have to develop a more integrated approach to category learning (e.g., Solomon, Medin, & Lynch, 1999).

Acknowledgments

This work was supported by Grant NSF SBR 97-20304 from the National Science Foundation.

References

Anderson, A., 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. (in press). The effect of category learning on sensitivity to within-category correlations. *Memory & Cognition.*

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments, & Computers, 25,* 257-271.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 1922-1944.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 1207-1238.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39-57.

Solomon, K. O., Medin, D. L., & Lynch, E. (1999). Concepts do more than categorize. *Trends in Cognitive Sciences, 3*, 99-104.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

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.

Yamauchi, T., & Markman, A. B. (2000). Learning categories composed of varying instances: The effect of classification, inference, and structural alignment. *Memory & Cognition, 28*, 2064-2078