Title
Intentional and Incidental Classification Learning in Category Use

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

ISSN
1069-7977

Author
Romano, Michael

Publication Date
2006

Peer reviewed

Intentional and Incidental Classification Learning in Category Use

Michael Romano (mrr233@nyu.edu)
Department of Psychology, New York University, 6 Washington Place
New York, NY 10003 USA

Abstract

Traditional classification learning studies tell us that people learn to attend to the diagnostic features of exemplars (Kruschke, 1992; Medin & Schaffer, 1978; Shepard, Hovland, & Jenkins, 1961). But recent research has discovered that the learning task influences what information people learn about categories (Markman & Ross, 2003). A learning task can either be the primary goal or be incidental to some other larger goal. This study investigated how intentional vs. incidental classification changes the kind of category information learned. The intentional classification group replicated previous studies by learning the diagnostic features, while the incidental classification group acquired information beyond the diagnostic features.

Keywords: categorization; implicit learning; implicit memory

Introduction

For much of the last 30 years, classification learning has been considered the central strategy to forming concepts (Barsalou, 1990; Chin-Parker & Ross, 2004; Estes, 1994; Kruschke, 1992). But outside the laboratory, people do not always learn categories as a main goal; they learn categories for some kind of use (Brooks, 1999). Markman and Ross (2003) argue that traditional classification learning studies overemphasize explicit classifications, even though people often make implicit classifications outside the laboratory. For example, customers do not examine merchandise in a sports store and explicitly ask, "Is this a running shoe or a boxing glove?" Rather, they implicitly classify items as part of the larger goal of shopping.

In this light, classification learning can be split into (at least) two types: *intentional classification* and *incidental classification*. Intentional classification learning occurs when one is explicitly aware of the classification task, such that it becomes the primary goal. Incidental classification learning is performed in support of some other goal.

Category information can be split into (at least) two types: *diagnostic features* and *prototypical features*. Diagnostic features are those that inform us how to select members that belong to a category while excluding those that do not. The diagnostic features of a category are defined by other categories. For example, having hair is useful in distinguishing between dolphins and sharks, but not between dolphins and whales. Thus, diagnostic features provide between-category information.

In family resemblance category structures (Rosch & Mervis, 1975), prototypical features are those commonly found in the members of a category. The prototypical features of a category stay the same regardless of the contrasting category. For example, some prototypical features of a dog are that it has four legs, fur, and a tail that it wags when it's happy. Thus, prototypical features provide within-category information.

Because intentional classification learning encourages people to explicitly focus on between-category information, this task facilitates the learning of diagnostic features (Kruschke, 1992; Medin & Schaffer, 1978; Shepard, Hovland, & Jenkins, 1961), but not the non-diagnostic, yet prototypical, features (Chin-Parker & Ross, 2004). On the other hand, incidental classification does not as strongly stress the importance of distinguishing between categories, which could create the opportunity to learn more within-category information (the prototypical features).

In the current research, it was hypothesized that the intentional classification group would learn a high amount of diagnostic information and a low (close to zero) amount of non-diagnostic information. It was also hypothesized that the incidental classification group would learn less diagnostic information than the intentional group, but more of the non-diagnostic (prototypical) information. In other words, each learning task would benefit from a gain in one type of information and a loss in the other.

There has been some previous research investigating how intentional vs. incidental classification affects what is learned about categories. For example, incidental learners are more likely to claim that they discovered a single defining feature that perfectly predicted category membership (Brooks, Squire-Graydon, & Wood, 1998). Three groups learned family resemblance categories by explicitly analyzing items for rules, by memorizing items, or by learning items incidentally to playing a board game. The incidental group made an average of 2.5% errors in the 10 trials preceding the test phase, but any single-feature rule would have resulted in at least 20% errors. That is, incidental learners believed that the categories had simple defining features, despite the fact that their categorization behavior was more complex. Brooks et al. called this the "simpler than it is" phenomenon, in that a person's belief about the nature of categories is not necessarily consistent with actual categorization behavior (Murphy, 2002).

Current Research

In this experiment, the performance of participants in the intentional classification task was compared with the performance of participants in the incidental classification task. Borrowing from the Brooks et al. (1998) study, participants in both conditions learned two categories of bugs while playing a board game. But in each trial, the intentional group classified exemplars before playing the

Table 1: Example category structures.

Item Type	Kez	Dax
Prototype	11111	11000
Learning exemplars	01111	01000
	10111	10000
	11011	11100
	11101	11010
	11110	11001

game, while the incidental group performed classifications at the moment of category use. This manipulation had the intentional group make explicit classifications as a primary goal, and had the incidental group make classifications in support of the main goal of playing the game.

The results of Brooks et al. suggest that the incidental classification group did not explicitly analyze their categorization behavior because of their false belief in defining features, even though the incidental group was able to perform correct classifications. If there is this difference between learning tasks, then there might also exist a difference in what information is learned and how it is learned. Two transfer tasks were used to test for explicit and implicit learning of category information.

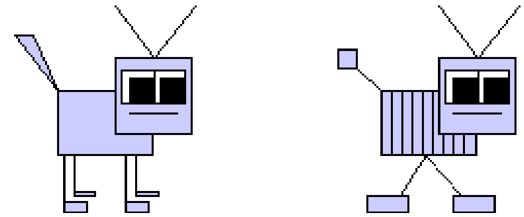
Method

Design This experiment had two between-subjects factors. First, participants were assigned to either the intentional classification learning condition or the incidental classification learning condition. Second, assignment of physical feature dimensions to dimensions of the abstract category structures was counterbalanced.

Participants Sixty-nine undergraduates from New York University participated for course credit. Twenty-nine participants, 12 from the intentional condition and 17 from the incidental condition, did not meet the learning criterion, and their data was excluded in the analyses. Non-learners were replaced with new participants from the same population. Participants were randomly assigned to a learning condition, and within each learning condition a participant was randomly assigned to 1 of 10 counterbalancing groups. This resulted in an equal number of participants (2) in each cell of this experiment's 20 cell design.

Table 2: Exemplar terms (Chin-Parker & Ross, 2004).

Term	# of Prototypical Features	# of Diagnostic Features
Prototype	5	3
Close3	4	3
Close2	4	2
Far3	3	3
Far2	3	2
Far1	3	1



Kez Dax

Figure 1: Example prototypes
(antennae & eyes are non-diagnostic).

Materials The materials were drawings of bugs (Figure 1) on a chess-style game board, presented on a computer monitor. Each participant learned 2 categories of bugs called Kez and Dax. Each bug category varied on 5 binary dimensions: antennae, eyes, body stripes, legs, and tail. The bug categories were both family resemblance structures; each prototypical feature occurred 80% of the time among a category's exemplars. Two features of the Kez and Dax prototypes overlapped causing them to be non-diagnostic, and the remaining 3 features were diagnostic (Table 1). Across participants within both learning conditions, the feature dimensions were counterbalanced such that each played the role of a diagnostic dimension an equal number of times.

Category exemplars varied on the number of diagnostic and prototypical features (Table 2). An exemplar could have 1, 2, or 3 of its diagnostic features. When an exemplar had 5 features consistent with its prototype, it was labeled *Prototype*. With 4 consistent features, it was labeled *Close*. With 3 consistent features, it was labeled *Far*.

Using the example category set in Table 1 (with Kez and Dax prototypes of 11111 and 11000, respectively), the Kez exemplar 01111 would be a *Close3* item because it has 4 prototypically-consistent features and all 3 diagnostic features. But if a participant learned the Dax category structure with the prototype 00011 instead of 11000, that same Kez exemplar 01111 would be a *Close2* item because it still has 4 prototypically-consistent features, but it only has 2 diagnostic features.

In the learning phase, all participants studied the 5 *Close* bugs from both categories, resulting in 10 learning exemplars. In the transfer phase, participants were given a typicality-ratings task for 16 exemplars from both categories, resulting in 32 separate ratings. These 16 exemplars consisted of the 5 *Close* exemplars from the learning phase, and 11 previously unseen exemplars including the category *Prototype* and 10 *Far* exemplars.

Procedure All phases of the experiment were conducted on PCs running Windows 98 using a custom-developed game programmed in C++ and OpenGL. All participants were verbally debriefed and provided with a written statement that described the purpose of the experiment.

Prior to the learning phase, all participants were told that they would play a board game where they would be required

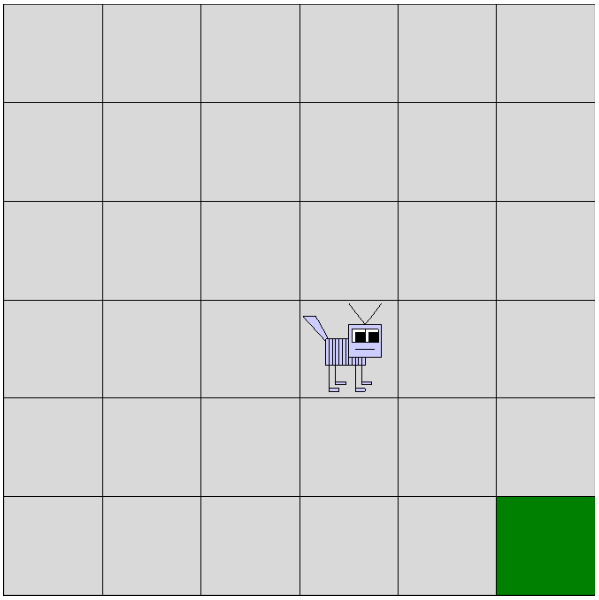


Figure 2: Example game board trial.

to move a game piece (an exemplar) to a goal position in the least number of moves. That is, they were under the assumption that this was a path-finding experiment, even though path-finding played no role in meeting the learning criterion. It was also explained to participants that there were 2 kinds of game pieces called Kez and Dax bugs, and that the Kez bugs could only move in straight lines horizontally and vertically, and that the Dax bugs could only move in diagonal lines.

For each trial, the game piece and the goal position were randomly placed such that it was possible to move the game piece as either a Kez or a Dax to the goal position. This fact precluded the determination of category membership based on starting positions. The category exemplar was rendered on the game board (Figure 2), representing the participant's game piece. The game piece was moved with the mouse by clicking on one of the adjacent squares, one square at a time, until a path was drawn from the starting position to the goal position.

In every trial of the intentional learning condition, participants were first presented with a screen displaying only the category exemplar and the explicit question, "Is this a Kez or a Dax?" The game board was not present during intentional classifications, which separated the explicit learning from the category use. Participants made their selection with the mouse by clicking one of two buttons labeled "Kez" and "Dax", and received feedback. After each intentional classification, participants played the game, moving the game piece to the goal position, with the intention of drawing a path in the least number of moves. If a participant tried to move in a direction not suited to the exemplar's category, nothing would happen. That is, no feedback occurred after the initial classification, limiting feedback to only once per trial in the intentional learning condition.

In every trial of the incidental learning condition, participants were never explicitly questioned about category membership. Instead, they only played the path-finding game in each trial. The incidental group also received feedback at a maximum of once per trial, but only on the first mistake made in moving the game piece. If a participant made no errors in moving during a trial, no feedback was given (although the absence of feedback was itself a form of feedback). In order to know how to move a game piece correctly, participants had to incidentally classify it as a Kez or a Dax. Making classification judgments at the moment of category use diverted attention to a goal other than explicit categorization in the incidental learning condition.

All participants played the game for a minimum of 4 blocks and a maximum of 40 blocks. There were 10 trials per block, and the game was self-paced. The learning criterion was passed when a participant successfully classified at least 9 of 10 exemplars for 2 consecutive blocks, and path-finding ability was irrelevant.

Following learning, all participants performed two transfer tasks. The first transfer task consisted of typicality ratings for 32 different exemplars, and tested recognition ability for diagnostic and prototypical features. Participants randomly saw 16 Kez exemplars of varying diagnosticity and prototypicality, and for each one they were asked, "How typical is this Kez?" They responded with the mouse by clicking 1 of 7 buttons (1 being "not at all typical" to 7 being "very typical"). Participants then performed the same ratings task with the 16 Dax exemplars.

The second transfer task had participants generate what they thought were the most typical Kez and Dax bugs, and tested recollection ability for diagnostic and prototypical features. First, participants drew the most typical Kez with the mouse by clicking on verbal descriptions of the binary features, such as "two eyes" or "three eyes," and "two legs" or "four legs." After drawing the most typical Kez, participants clicked a button to continue and drew the most typical Dax. It was not possible to continue until a feature for each dimension was selected.

Results and Discussion

Learning Phase The intentional group showed a clear advantage in learning over the incidental group. To start, 62.5% of the intentional participants met the learning criterion, compared to 54.1% of the incidental participants. For those that learned the two categories, there was a significant difference between groups in the number of blocks needed to meet the learning criterion, $t(38) = -2.36$, $p < .05$. The intentional group needed an average of 15.90 ($SD = 7.48$) blocks to learn, and the incidental group needed an average of 21.75 ($SD = 8.18$) blocks to learn. So not only did more intentional learners meet the learning criterion, they also met the criterion faster than the incidental learners.

In order to analyze the number of errors per learning block over time, the number of errors over every 4 blocks was averaged into a *batch* for all 40 potential blocks, so that there were 10 batches of errors/blocks per participant. Not all participants played the game for all 40 blocks, and some batches for some participants included in the analysis contained zero errors. A two-way mixed measures analysis of variance (ANOVA), with the 10 batches of errors/block as the within-subject factor and learning condition as the between-subjects factor, revealed a significant effect of learning condition, $F(1, 38) = 7.34$, $MSE = 4.73$, $p = .01$, a significant effect of batch number, $F(9, 342) = 106.99$, $MSE = .76$, $p < .001$, and a significant interaction, $F(9, 342) = 2.84$, $MSE = .76$, $p < .01$. This result reflects the faster learning achieved by the intentional group over the incidental group, suggesting that the participants in the two learning conditions engaged distinct learning strategies (Figure 3).

Transfer Phase: Typicality Ratings Only participants who met the learning criterion were included in the following analyses. Following Chin-Parker and Ross (2004), separate measures were calculated from the typicality ratings. To measure the effect of the learned diagnostic features, the *diagnosticity drop* was calculated by averaging the typicality ratings change when diagnosticity varied but prototypicality remained constant. The average typicality ratings change between the *Close3* and *Close2* exemplars, the *Far3* and *Far2* exemplars, and the *Far2* and *Far1* exemplars, all had varied diagnosticity but identical prototypicality between comparisons. To measure the effect of prototypicality, the *prototypicality drop* was calculated by averaging the typicality ratings change when prototypicality varied while diagnosticity remained constant. The average typicality ratings change between the *Prototype* and *Close3* exemplars, the *Close3* and *Far3* exemplars, and the *Close2* and *Far2* exemplars, all had varied prototypicality but identical diagnosticity between comparisons.

The intentional group had a large diagnosticity drop of 1.03 ($SD = .70$), but their prototypicality drop was near zero,

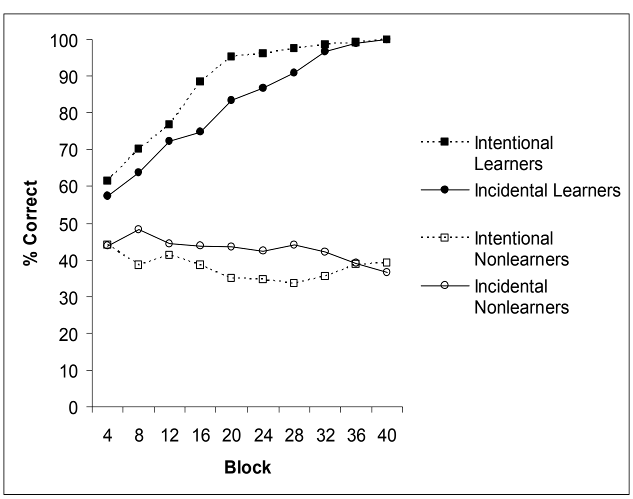


Figure 3: Learning rates.

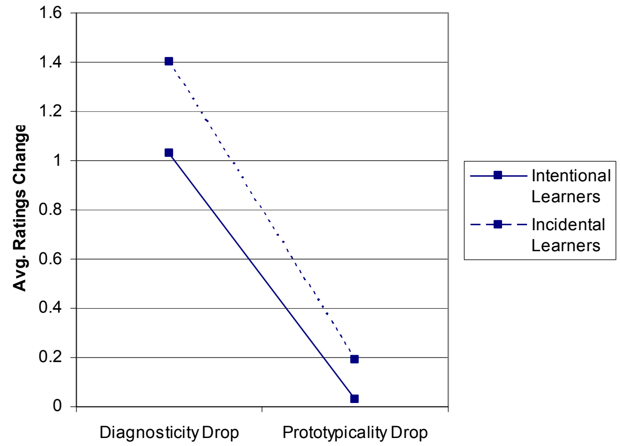


Figure 4: Average typicality ratings change.

at .03 ($SD = .52$). The incidental group had an even larger diagnosticity drop of 1.40 ($SD = .88$), and a prototypicality drop of .19 ($SD = .47$). A diagnosticity or prototypicality drop of zero indicates that no information was learned. A positive diagnosticity drop indicates learned diagnostic features, and a positive prototypicality drop indicates learned non-diagnostic features. In Figure 4, it can be seen that both the intentional and incidental group learned diagnostic features, and that only the incidental group learned some non-diagnostic features.

For the intentional group, separate *t*-tests showed that their diagnosticity drop was significantly different from zero, $t(19) = 6.60$, $p < .001$, but that their prototypicality drop was not, $t < 1$. In other words, the intentional group was good at acquiring diagnostic features, but not non-diagnostic features. For the incidental group, their diagnosticity drop was also significant, $t(19) = 7.14$, $p < .001$, and their prototypicality drop bordered on significance, $t(19) = 1.82$, $p < .08$. Upon removing a single outlier[1] from the incidental group, their prototypicality drop became significant, $t(18) = 2.37$, $p < .05$. In other words, the incidental group was also able to acquire diagnostic features, and they were better at learning the non-diagnostic features than the intentional group who learned none.

A two-way mixed measures ANOVA, with the type of drop as the within-subject factor and learning condition as the between-subjects factor, revealed a significant effect of learning condition, $F(1, 38) = 6.28$, $MSE = .22$, $p < .05$, a significant effect of drop type, $F(1, 38) = 37.75$, $MSE = .65$, $p < .001$, and no interaction, $F < 1$. These results show that learning condition affected what was learned, and that there was a significant difference in the amount of learned diagnostic vs. non-diagnostic features within both conditions. Contrary to the hypothesis, there was no

[1] The outlier was chosen due to its highly negative prototypicality drop of -.67, which was unusual compared to the rest of the incidental group. An individual who learned no non-diagnostic information would be expected to exhibit a prototypicality drop of zero, while a person who learned some would have a positive drop.

Table 3: Average typicality ratings.

	Intentional Group			Incidental Group		
Diagnosticity	3	2	1	3	2	1
Prototype	5.68			6.20		
Close	5.64	4.55		5.99	4.74	
Far	5.40	4.73	3.39	5.58	4.80	2.61

interaction. The hypothesis predicted that each learning condition would provide a gain in one type of category information and a loss in another. Incidental classification learning appears to have gains in both types of category information over intentional classification learning.

Independent samples *t*-tests showed that only the diagnosticity drop between the *Far2* and *Far1* items was significant between groups, $t(38) = -2.21$, $p < .05$ (Table 3). All other separate diagnosticity and prototypicality drops were not significantly different between groups, $t < 1$. This means that the difference between learning conditions in diagnosticity drop is due to how participants reacted to the previously unseen *Far2* and *Far1* items in the typicality ratings task.

The intentional group rated the *Far1* items as 3.39, which indicates that they perceived those items as fairly ambiguous (a rating of 4.0 meant it had a 50% likelihood of being a category member). However, the incidental group rated the *Far1* items as 2.61, which indicates that they perceived those items as poor category members (below 50%).

In the typicality ratings task, it was clear that the intentional group was only influenced by diagnosticity. These results replicated the findings of Chin-Parker and Ross (2004). In contrast, the typicality ratings of the incidental group indicated that they acquired both diagnostic and some non-diagnostic information. However, it must also be noted that the incidental group required significantly more blocks to reach the learning criterion. Could their increased sensitivity to overall category information be due to the fact that they spent, on average, more time exposed to the exemplars in the learning phase?

To answer this question, a linear regression of diagnosticity drop on the number of blocks to learning was performed. There was no significant effect of block for either the intentional group or the incidental group, both $F < 1$. A similar regression for prototypicality drop also revealed no significant effect of block for either group, both $F < 1$. Therefore, one can conclude that the differences in typicality ratings performance are a result of the differences in learning behavior rather than a consequence of prolonged exposure to the learning exemplars.

Transfer Phase: Picture Generation In the picture generation task, the accuracy of the diagnostic and non-diagnostic features was measured. The intentional group successfully drew a high proportion of diagnostic features ($M = .83$, $SD = .22$), and they drew an above-chance (50%) proportion of non-diagnostic features ($M = .59$, $SD = .15$). The incidental group also successfully drew a high proportion of diagnostic features ($M = .79$, $SD = .19$), but they drew a near-chance proportion of non-diagnostic features ($M = .53$, $SD = .11$).

A mixed measures ANOVA, with the type of proportion as the within-subject factor and learning condition as the between-subjects factor, revealed no effect of learning condition, $F(1, 38) = 1.99$, $MSE = .03$, $p = .16$, a main effect of feature diagnosticity, $F(1, 38) = 43.01$, $MSE = .03$, $p < .001$, and no interaction, $F < 1$. These results indicate that despite the differences in performance in the typicality ratings task, there was no significant difference between learning conditions in the picture generation task. The intentional group drew a significantly above-chance proportion of non-diagnostic features, $t(19) = 2.67$, $p = .01$, while the incidental group did not, $t(19) = 1.00$, $p = .33$.

General Discussion

The results of this experiment differed from the original hypothesis. Although, as predicted, the incidental classification group learned some of the non-diagnostic information, they were also seemingly better at acquiring diagnostic information. What is interesting is that the incidental group appears to have learned more overall category information compared to the intentional group. While it should be noted that the statistical effects were small, they replicated a pilot study that produced similar significant results.

Both the intentional and incidental participants whose transfer data were analyzed had met the learning criterion. Obviously all participants in both conditions had learned the 3 diagnostic features necessary to pass the criterion. Why, then, is there a difference in diagnosticity drop between groups if they learned (at least) the same diagnostic features?

This result may be another indication that the incidental group acquired more prototypical information than did the intentional group; not that they learned more diagnostic information. This claim may seem counterintuitive, but it is important to remember that although the typicality ratings allow us to separate effects *between* ratings, the amount of influence both types of category information have on individual ratings is confounded. That is, when a participant rates any specific item, the experimenter cannot calculate exactly how much influence the learned diagnostic and non-diagnostic information had on that single rating.

Recall that the significant difference between groups occurred in the typicality ratings change between the *Far2* and *Far1* items' diagnosticity drop. The intentional group rated the *Far1* items as ambiguous category members (3.39 on a 7-point scale), and the incidental group rated the same items as poor category members (2.61 on a 7-point scale). The incidental group used both diagnostic and non-diagnostic information, both providing reasons for giving poor ratings on *Far1* items because those items have low diagnosticity and low prototypicality. But the intentional group only had diagnostic information to aid in their typicality ratings, which by itself provided a weaker basis for negatively rating *Far1* items. Thus, the availability of non-diagnostic information to the incidental group could have influenced their diagnosticity drop in this way.

But if the acquisition of non-diagnostic information explains the results of the typicality ratings task, then how does this explanation handle the fact that the incidental group performed at chance-levels in the picture generation task for non-diagnostic features? The incidental group was never asked to explicitly analyze the category exemplars, and they also engaged the exemplars with divided attention, which might have been compensated for with implicit learning. The involvement of implicit memory explains why the incidental group was influenced by prototypicality in their typicality ratings but not in their picture generations, because implicit memory is useful in cued recognition tasks like typicality ratings, but not in recollection tasks like picture generation (Lockhart, 2000).

Recall the Brooks et al. (1998) study which found that the incidental group was more likely to believe (incorrectly) that categories possessed defining features. This belief suggests that the incidental classification task does not encourage explicit analysis of either the exemplars or one's learning behavior such that incidental learners are unaware of the complexity of their actual categorization behavior. The fact that the incidental group in the Brooks et al. study still successfully used the categories hints at the idea that they, too, might have benefited from implicit learning. Implicitly learned categories would not be accessible to explicit analysis, which might be the cause of the incorrect belief in defining features. Future studies will explore the possible relationship between implicit memory and the belief in defining features.

In a similar study, Minda and Ross (2004) investigated indirect category learning, which they defined as not informing participants that there are categories to be learned, but learning those categories improves performance on a feedback-driven task. The indirect learning group only made predictions about categories, but the direct learning group first classified each exemplar before making a prediction. The categories could be learned by attending to either a single criterial attribute or the overall family resemblance structure. Minda and Ross found that the indirect learning group was more likely to learn the family resemblance structure, and that the direct learning group was more likely to focus on the single criterial attribute. These results run parallel to the current research in that both incidental classification and indirect learning result in the acquisition of prototypical information, while both intentional classification and direct learning focus attention on diagnostic information.

In summary, this research expands on Markman and Ross's (2003) argument that the learning task used influences what a person learns about a category. Intentional classification learning emphasizes explicit distinctions between categories with the primary goal of learning the categories. The most efficient method of achieving this goal is to focus on and learn the diagnostic features, as the participants did. On the other hand, incidental classification learning does not emphasize explicit distinctions between categories, and it is done in support of some other goal. Future research will explore whether it is the incidental aspect of this learning task that promotes the learning of prototypical features, or if it is the main goal of category use (e.g., path-finding, prediction, shopping, etc.) that determines if prototypical features are learned. Perhaps there are some category uses, like bird watching and rock collecting, which stress explicit distinctions between categories such that incidental classification would also result in only learning diagnostic features.

Acknowledgments

I thank Bob Rehder for his advisement in this study.

References

Barsalou, L. W. (1990). Access and inference in categorization. *Bulletin of the Psychonomic Society, 28,* 268–271.

Brooks, L. R. (1999). Orientation to meaning suppresses explicit knowledge of surface cues in category learning. *Abstracts of the Psychonomic Society, 4,* 51.

Brooks, L. R., Squire-Graydon, R., & Wood, T. J. (1998). The role of inattention in everyday concept learning: Identification in the service of use. March 6, 1998 draft paper.

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.

Estes, W. K. (1994). *Classification and cognition*. New York: Oxford University Press.

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

Lockhart, R. S. (2000). Methods of memory research. In E. Tulving and F. I. M. Craik (Eds.), *The Oxford Handbook of Memory*. New York: Oxford University Press.

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

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

Minda, J. P., & Ross, B. H. (2004). Learning categories by making predictions: An investigation of indirect category learning. *Memory & Cognition, 32,* 1355–1368.

Murphy, G. L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

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

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