Feature Inference and Eyetracking

Bob Colner (bob.colner@nyu.edu)
Bob Rehder (bob.rehder@nyu.edu)
Department of Psychology, New York University
6 Washington Place, New York, NY 10003 USA

Aaron B. Hoffman (aaron.hoffman@mail.utexas.edu)
Department of Psychology, University of Texas at Austin
1 University Station, Austin, Texas 78712 USA

Abstract

In addition to traditional supervised classification learning, people can also learn categories by predicting the features of category members. It has been proposed that feature inference learning promotes the learning of more within-category information and a prototype representation of the category, as compared to classification learning that promotes learning of diagnostic information. We tracked learners' eye movements during inference learning and found (Expt. 1) that they indeed fixated other features (even though those features were not necessary to predict the missing feature), providing the opportunity to extract within-category information. But those fixations were limited to only those features that needed to be predicted on *future* trials (Expt. 2). In other words, inference learning promotes the acquisition of within-category information not because participants are motivated to learn that information, but rather because of the anticipatory learning it induces.

Whenever a person classifies an object, describes a concept verbally, engages in problem solving, or infers missing information, they must access their conceptual knowledge. As a result, the study of concept acquisition has been a critical part of understanding how people experience the world and how they interact with it in appropriate ways.

Concept researchers have developed sophisticated formal theories that explain certain aspects of concept acquisition. These theories are largely based on the study of what has come to be known as standard supervised classification—a task that occupies the majority of experimental research in this area (Solomon, Medin, & Lynch, 1999). However, an emerging literature is focused on expanding the range of tasks that can be used to inform our models of concept acquisition. By studying different learning tasks we can understand other aspects of concept acquisition, including the interplay between category use and the type of concept learned (Brooks, 1978; Yamauchi & Markman, 1998, 2000, 2002; Chin-Parker & Ross, 2002). Within this research, the distinction between inference and classification tasks has received the most attention, perhaps because those two tasks can be more easily equated. In fact, Anderson (1991) has argued that inference and classification can be treated identically if category labels and category features are interchangeable (however see Yamauchi & Markman, 2000).

Research on classification versus inference learning has revealed apparent differences in the types of category representations formed. Whereas classification promotes learning the most diagnostic features for determining category membership, inference may foster learning additional category information (Chin-Parker & Ross, 2004; Medin et al., 1987; Shepard, Hovland, & Jenkins, 1961; Rehder & Hoffman, 2005a). Classification versus inference learning also affects the ease with which different category structures are acquired. Linearly separable (family-resemblance) category structures are more easily acquired through inference relative to classification (Yamauchi & Markman, 1998). However, when a comparable non-linearly separable category structure is used, classification yields a significant learning advantage (Yamauchi & Markman, 2002).

Differences in how category information is acquired across classification and inference tasks have been explained in terms of exemplars and prototypes. Yamauchi and Markman have argued that inference learners form representations consistent with prototype models because they seem to extract family-resemblance information such as typical features and typical feature relations. In contrast, by focusing on diagnostic information, classification encourages representations consistent with learning rules and exceptions (perhaps via exemplar memorization). Nevertheless, this interpretation has been challenged by arguments noting the many differences between the classification and inference tasks. This debate is worth discussing in detail.

Yamauchi and Markman (1998, Exp. 1) contrasted classification and inference learning by training groups of participants on a family resemblance category structure, consisting of four exemplars per category (see Table 1). Each item consisted of a label and four binary feature dimensions. The members of both categories were derived from category prototypes, A = 0000 and B = 1111. All items had one dimension that contained a feature value taken from the opposite category prototype, i.e., an *exception feature*. Participants either classified the eight exemplars into two categories or they predicted a feature missing from every exemplar. One critical aspect of their design was that participants were never required to predict a missing exception feature. For example, they were never presented with the item 000*x* labelled as a member of category A and asked to predict (on

Table 1. Yamauchi & Markman category structure.

Cat. Label	D1	D2	D3	D4
Prototype A	0	0	0	0
A1	0	0	0	1
A2	0	0	1	0
A3	0	1	0	0
A4	1	0	0	0
B1	1	1	1	0
B2	1	1	0	1
B3	1	0	1	1
B4	0	1	1	1
Prototype B	1	1	1	1

the basis of A1 in Table 1) a '1' for the unknown value *x* on dimension 4. The reason for this choice was to keep the classification and inference tasks as closely matched as possible during learning. Following learning, all participants completed a transfer test in which participants made inferences on all feature dimensions. During this phase, learners were not only asked to infer typical features (just as they had during training), they were also presented with *exception feature trials* (e.g., they predicted *x* in item A000*x*). Participants were told to respond based on the categories they had just learned and did not receive feedback.

Yamauchi and Markman observed that the inference participants required fewer blocks to reach the learning criterion. Perhaps this should not come as a surprise, because whereas classification required integrating information across four feature dimensions, none of which were perfect predictors alone, the inference learners had access to a perfect predictor, namely, the category label.

A second important result concerned how people responded to the exception feature trials during test. Again, strict adherence to exemplars in Table 1 requires one to predict a value typical of the opposite category (e.g., predict $x = 1$ for A000*x*). In contrast, responding on the basis of the category prototype requires responding with a typical feature. In fact, Yamauchi and Markman found that whereas classification learners generally responded with the exception feature, inference learners generally responded with the category's typical feature. In other words, whereas classification learners made inferences according to the training exemplars, inference learners made inferences consistent with the category's prototype. This result, coupled with formal model fits, led Yamuachi and Markman to conclude that inference learners represent prototypes and classification learners represent exemplars.

Subsequent investigations with the inference task have supported and expanded this conclusion. For example, Chin-Parker and Ross (2004) manipulated the diagnosticity and prototypicality of a feature dimension and found that categorization learners were only sensitive to the diagnostic features whereas inference learners were also sensitive to nondiagnostic but prototypical features. In addition, Chin-Parker and Ross (2002) demonstrated that inference learning but not classification learning results in sensitivity to within-category correlations. This latter finding suggests that inference learning not only promotes learning of the category's prototype, it results in better learning of the category's internal structure (including interfeature correlations) more generally. We'll refer to the proposal that inference learning promotes learning of categories' internal structure as the *category-centered learning hypo*thesis (CCL).

However, Johansen and Kruschke (2005) offered an alternative explanation of some of these data. They argued that inference learners did not learn prototypes, but rather a set of category-to-feature rules that the prototype model mimics. According to Johansen and Kruschke, the rule-based explanation is possible because exception-feature inferences were excluded from the learning phase. As a result, learners in Yamauchi and Markman (1998) and Chin-Parker and Ross (2004) could succeed in the inference task even if they ignored everything but the category label. In contrast, the classification learners were forced to either memorize exemplars or learn an imperfect rule with exceptions. (Note that this account does not explain the learning of within-category correlations in Chin-Parker & Ross, 2002, a point we return to in the Discussion.)

The current study is designed to differentiate between the CCL hypothesis and the alternative category-to-feature rule hypothesis. It turns out that these theories make unique predictions regarding the allocation of attention during the course of inference learning. Under the rule hypothesis, attention should eventually be limited to just the category label and to the to-be-predicted feature. However, under the CCL hypothesis, attention should be allocated to within-category information, including the multiple feature dimensions throughout the course of learning. Thus, CCL and the rule account of inference learner offer opposite predictions regarding how learners will allocate attention during the learning task.

Previous research with supervised classification tasks has used eye tracking to assess learners' attention allocation (Rehder & Hoffman, 2005a; b). For example, in Rehder and Hoffman (2005a), participants' eye movements were recorded as they learned Shepard et al's Type I category structure in which one dimension is completely predictive of category membership and the other two dimensions are irrelevant. Eye movements revealed that attention was eventually allocated exclusively to the diagnostic dimension, an account consistent with participants acquiring a simple feature-to-category label rule. A similar result is predicted by the rule account in the inference task, as learners should restrict attention to the perfectly predictive category label.

Experiment 1

Method

Participants. A total of 44 New York University undergraduates participated in the experiment for course credit. Two participants did not complete the experiment.

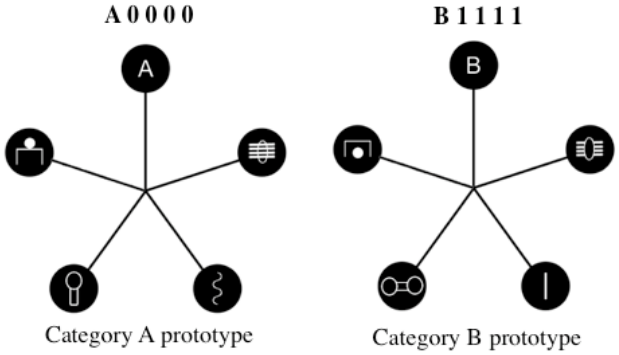

Figure 1. Stimuli for Experiments 1 and 2.

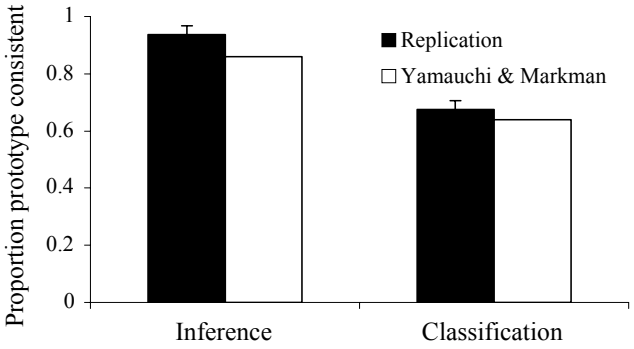

Figure 2. Exception feature trial results from Expt. 1.

Materials. The category structure was identical to that used by Yamauchi and Markman (1998) (Table 1). The stimuli were designed to facilitate the recording of eye movements with dimensions separated in space and features pretested for nearly equal discrimination times. They consisted of a category label, the letter A or B, and four binary shape features. An example is shown in Fig. 1. The category label and features were equidistant from the center of the display. The position of the features and category label on the screen were counterbalanced with a Latin square design so that the category label and features appearing an approximately equal number of times in each of the five positions on the screen. An SMI Eyelink I system was used to record eye movements.

Design. Participants were assigned randomly and in equal numbers to inference learning or classification learning. Participants were also assigned to one of five conditions determining the physical position of features.

Procedure. The design and procedure replicated Yamauchi and Markman (1998) with an eye tracker. Yamauchi and Markman also had a third, "mixed" condition, but this was omitted. The experiment had a learning phase followed by a test phase. In the learning phase all participants responded until they reached a *learning criterion* of 90% accuracy on three consecutive blocks, or until they completed 30 blocks.

For classification learners, possible category labels were presented side by side (in a random position) at the category location. Participants responded by using the left or right arrow key to select one of the category labels.

Inference learners were presented with stimulus items with intact category labels but with one missing (queried) feature. A dashed line terminated at the queried location with the two possible features. Once the stimulus appeared, participants responded with the arrow key to select the correct feature. Following each response the stimulus would disambiguate, i.e., the correct feature or category label would replace the queried location. There were also feedback tones associated with correct and incorrect responses. Throughout learning, a classification block consisted of classifying all eight exemplars but no prototypes. An inference block consisted of two feature inferences on all of the four features, on all eight exemplars, but never on an exception feature (as in Yamauchi & Markman, 1998).

The test phase was identical for both conditions and included both classification and inference trials. First, all participants made 10 classification judgments on all eight exemplars from the learning phase and the two novel category prototypes. Following classification, participants made 32 feature inferences; they inferred every feature of every exemplar (including typical and exception features).

Results and Discussion

We analyzed data from the 38 participants (19 in each condition) who reached the learning criterion. The average number of blocks needed to reach the learning criterion was computed. Replicating the result from Yamauchi and Markman, inference participants required fewer learning blocks (m = 7.9, sd = 4.0), than classification participants (m = 13.2, sd = 7.8), $t(36) = -2.64$, $p < .01$.

We next examined participants' accuracy during the test phase, as a function of whether the test task matched their training task. As expected, performance was superior when the learning and test tasks matched. Classification participants were significantly more accurate (m = 0.98) than inference participants (m = 0.77) when classifying, $t(36) = -6.55$, $p < 0.001$, but category-prototype classification was not reliably different between classification (m = 0.97) and inference (m = 1.0) conditions, $t < 1$.

The inference test of old stimuli confirmed that participants in the inference condition are slightly more accurate (m = 0.93) than classification participants (m = 0.90). Unlike Yamauchi and Markman this difference was not significant. Importantly, the current study replicated the results of the exception-feature inference trials (Fig. 2). Recall that on these trials responding in manner faithful to the exemplars in Table 1 requires inferring a feature typical of the opposite category. Inference participants instead inferred features consistent with the category prototype (m = 0.93), and did so significantly more often than classification participants (m = 0.67), $t(36) = 2.68$, $p < 0.01$.

We next examined eye movements to understand why inference learners are likely to infer prototypical features. Is it because they are learning simple rules between category labels and features, or because inference learning promotes learning about the internal structure of categories? Our analysis used the binary measure of whether a participant fixated a particular area of interest to produce the probability (over participants) that an area was fixated. Five separate

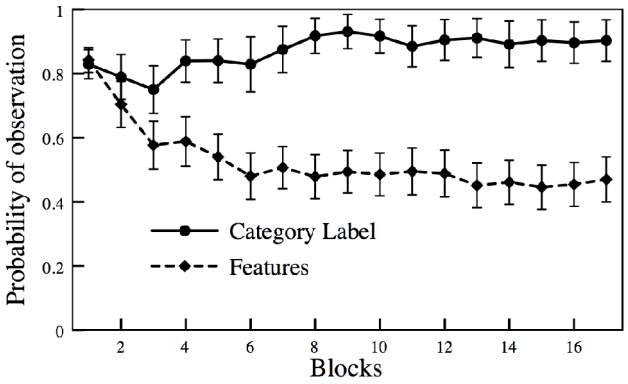

Figure 3. Observation probability in Expt. 1.

regions around the four feature dimensions and the category label were defined and we recorded whether these areas were fixated. A dimension was considered *observed* on a trial if its region received one or more fixations. This variable was averaged over participants and blocks. Fig. 3 plots the probability of observing the category label and features over the course of learning in the inference condition, excluding fixations to the to-be-predicted feature. To generate Fig. 3, the averaged eye movement data from the last two blocks of learning were used to pad a participant's data if they completed the experiment in fewer than 17 blocks.

Fig. 3 shows a reduction in the probability of observed features during the first six blocks, from .84 in the first block to .48 in the sixth. In contrast, the probability of observing the category label remained stable at around .82. But although there was a reduction in fixations to the features, they remained substantial throughout the learning phase. This result is important for two reasons. First, it reflects a pattern of attention allocation unlike what has been observed in standard classification tasks. In such tasks, attention is optimized exclusively to perfectly predictive dimensions (Rehder & Hoffman, 2005a). Second, the observed pattern of attention allocation is unlike what was predicted from the simple rule account. Indeed, learners attended in a way that suggests active information extraction about the internal structure of the categories and not (just) learning simple rules.

A closer analysis of attention allocation at the end of learning revealed substantial differences among participants. The histogram in Fig. 4 shows a bimodal distribution of feature observations on the last block of learning. On one side of the distribution, there is a cluster of people who

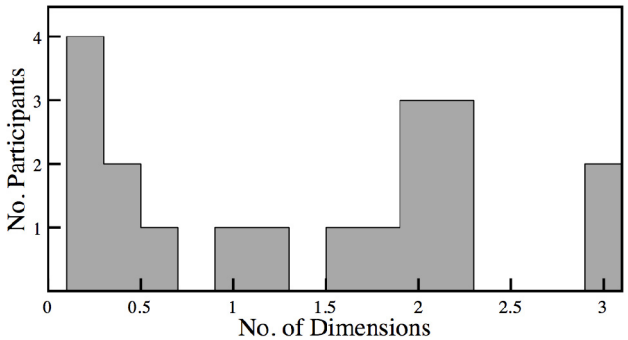

Figure 4. Observations in last block of Expt. 1.

made two feature observations. This group represents participants who attended to multiple features in addition to the category label to solve the task. The second cluster of participants made nearly zero feature observations. This group represents people who relied entirely on the category label to successfully infer missing features. Unlike the first group, this group is in fact consistent with learning simple rules. Interestingly, there was also a pair of participants who managed to solve the task without attending to the category label at all during the last learning block. This strategy would necessitate integrating the probabilistic information from all three features to achieve the 90% accuracy required by the criterion (much like the classification group). This more fine-grained analysis suggests that participants were engaging in qualitatively different learning strategies. Still, most participants were fixating most of the other feature dimensions—a result that the rule account by itself is unable to explain.

Experiment 2

Expt. 1 provided evidence that inference learners distribute attention among multiple feature dimensions. This result supports the CCL hypothesis, that inference training fosters an interest in what the categories are like, rather than simply how to discriminate objects into opposing categories. However, there are other potential explanations of the fixations to the additional feature dimensions. One possibility is that most participants were attending to unnecessary dimensions on trial *n* because they realized they would need to predict one of those dimensions on trial n+1. Such *anticipatory learning* reflects a meta-cognitive learning strategy rather than a motivation to represent the internal structure of categories. Another possibility is that while the category label was perfectly predictive of each to-be-predicted feature, the other three feature dimensions, taken together, also predicted those features. (Indeed, recall that two Expt. 1 participants reached criterion even without fixating the category label.) Thus, participants may have fixated the other dimensions in order to have a set of predictors that were redundant with the category label. Finally, it is possible participants fixated additional dimensions simply because they found the stimuli in Fig. 1 intrinsically interesting.

The purpose of Expt. 2 was to discriminate between anticipatory learning and CCL (and the other two hypotheses). Following Anderson, Ross, and Chin-Parker (2002, Expt. 3), inference learners acquired the category structure in Table 1 but this time made inferences on only 2 of the 4 the feature dimensions. The other dimensions were presented but never queried. According to CCL, inference learners' motivation to learn the internal structure of categories should cause them to fixate all stimulus dimensions. This result will also obtain if (a) the other dimensions are being used as a redundant predictor or (b) the stimuli are intrinsically interesting. But if Expt. 1's participants were engaged in anticipatory learning instead, fixations to the two never-queried dimensions should be rare.

Methods

Participants and materials. A total of 36 New York University undergraduates participated for course credit. The abstract category structure and its physical instantiation were identical to Expt. 1.

Design. Participants were randomly assigned to one of five configurations of the physical locations of the item dimensions and to one of six possible pairs of feature dimensions to serve as the never-queried features.

Procedure. The procedure was identical to Expt. 1, except that just two of the four possible feature dimensions were queried during the learning phase.

Results and Discussion

We restricted our analysis to the 32 participants who reached the learning criterion. On average less then seven blocks were required to reach the learning criterion (m = 6.4, sd = 0.79), a learning rate similar to that in Expt. 1's inference condition.

Fig. 5 shows inference performance during the test phase, as a function of whether the dimension was sometimes queried or never queried and whether it was a typical- or exception-feature trial. Fig. 5 indicates a large effect of whether dimensions were sometimes- or never-queried. Whereas participants made the prototype-consistent response 91% of the time on the sometimes-queried dimension, their accuracy dropped to 64% on the never-queried dimensions. Importantly this effect was obtained not only for typical-feature trials but also exception-feature trials in which strict adherence to the exemplars in Table 1 requires participants to predict an exception feature. That is, just as in Expt. 1, inference participants learned the typical features and predict those features even on items that displayed an exception feature during training. Nevertheless, note that participants were somewhat less likely to make a prototype-consistent response on exception feature trials than typical feature trials (m = .74 vs. m = .82), demonstrating that inference learners acquire some configural (exemplar-based) information about the category structure.

A 2 x 2 repeated measures ANOVA with dimension (sometimes- or never-queried) and trial type (typical, or exception) as factors revealed a main effect of dimension, $F(1, 30) = 39.2$, $MSE = 2.36$, $p < 0.0001$, reflecting the greater number of prototype-consistent responses on the sometimes-queried dimensions. Nevertheless, accuracy on the never-queried features was significantly better then chance, $t(31) = 4.05$, $p < 0.001$. There was also a main effect of trial type, $F(1, 30) = 5.1$, $MSE = 0.23$, $p < 0.05$, indicating that participants made more prototype-consistent responses on typical features trials than exception feature trials. The interaction was not significant, $F < 1$.

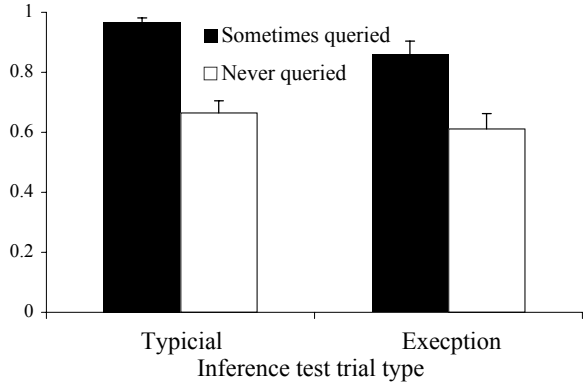

Figure 5. Inference test results from Expt. 2.

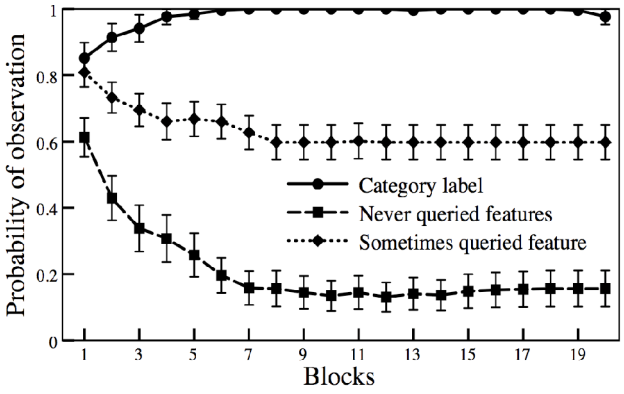

Figure 6. Observation probability in Expt. 2.

The key data of course are eye movements. Did learners' desire to learn within-category information lead them to fixate dimensions that are never queried or will anticipatory learning lead them to focus primarily on queried dimensions? The probability of observing never- versus sometimes-queried dimensions is plotted in Fig. 6. In the figure, fixations to the current, to-be-predicted dimension are omitted. In fact, as early as the first block of learning, participants showed signs of beginning to ignore the never-queried dimensions. At the end of learning, inference learners' probability of fixating never-queried dimensions was less than 0.2. In contrast, their probability of fixating the sometime-queried dimension was over three times greater, 0.6. These results suggest that the vast majority of fixations to other feature dimensions in Expt. 1 arose not because participants were trying to learn what the categories were like, but rather because they were anticipating feature inferences they would be required to make on future trials.

General Discussion

To discriminate between the CCL hypothesis and the Johansen and Kruschke's (2005) label-based rule hypothesis we replicated Yamauchi & Markman (1998) while measuring learners' attention with an eyetracker. In fact, we found that neither hypothesis provided a full explanation of our results.

First, the eye movements observed during Expt. 1 were inconsistent with the simple predictions we derived from the rule account. Despite the presence of a perfect predictor (the category label), participants generally fixated most of the other features. These fixations provided learners with the opportunity to acquire category information beyond the minimum necessary for the task. These results suggested that the inference learners were motivated to learn about the internal structure of categories, a finding in support of CCL.

However, in Expt. 2 learners made inferences on only two of the four feature dimensions. CCL predicted that the never-queried feature dimensions should continue to be fixated (to learn as much as possible about the categories' internal structure). However, fixations were generally restricted to the sometimes-queried dimensions and the category label. Our learners generally ignored dimensions that were never queried.

We can interpret the results of our two experiments as indicating that inference learners are not generally motivated to learn the internal structure of categories. Instead, they are motivated to do well on their assigned task. To accomplish this, they fixate information that will lead to a correct response on the current trial and they may also fixate information that will help them respond correctly on upcoming trials. That is, although on each trial inference learners appear to attend to both necessary and unnecessary information, as predicted by CCL, the additional attention allocated prepares them for making inferences on future trials. Recall that Expt. 2 also ruled out two other interpretations of Expt. 1, namely, that fixations to all dimensions arose because they provided a redundant set of predictors or because participants found the stimuli intrinsically interesting.

Of course, it is important to note that inference learners *did* learn something about the never-queried dimensions in Expt. 2. Participants responded with prototype-consistent responses even on those dimensions, albeit far less accurately than on the queried dimensions. Because our eye tracking data revealed few fixations to those dimensions after the first few blocks (Fig. 6), this learning must have occurred early in the experimental session. We interpret this as participants being initially unaware of which feature dimensions were sometimes-queried and which were never-queried, and thus fixated all dimensions in anticipation of future feature inference trials. But they quickly learn which dimensions are never queried and stop fixating them. Indeed, even in the first block, the never-queried dimensions are fixated significantly less often than the sometimes-queried one.

Although our results thus argue against the CCL hypothesis, it is important to recognize that feature inference learning still results in the acquisition of different sorts of category information. Our findings, like those of Yamauchi and Markman (1998), show that inference learning results in participants predicting prototype-congruent feature values more often than classification learners. As mentioned, Chin-Parker and Ross (2004) have shown that it results in them learning information that is prototypical but not diagnostic. Finally, it also results in them learning within-category correlations that are not necessary for classification (Chin-Parker & Ross, 2002). But we believe all of these results occur because inference learners engage in a form of anticipatory learning in which they fixate feature dimensions they know they will need to predict in the future. Apparently, this learning is sufficient to not only learn the features themselves but also to learn (incidentally, in an unsupervised manner) any additional internal structure involving those features, namely, the correlations between them (Chin-Parker & Ross, 2002). That is, although participants in these sorts of experimental paradigms may not be motivated to learn more than necessary, they learn more than necessary nevertheless (Brooks et al., 2007).

So inference learning may not be sufficient to energize people to learn categories. But recognize that, if you want someone to learn a family resemblance category, including its prototypical features and within-category correlations, don't have them classify items. Have them predict features.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 0545298.

References

Anderson, A. L., Ross, B. H., & Chin-Parker, S. (2002). A further investigation of category learning by inference. *Memory & Cognition, 1*, 119-28.

Anderson, J.R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*, 409-429.

Brooks, L. (1978). Non-analytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (eds.), *Cognition and categorization* (pp. 169-211). Hillsdale, NJ: Erlbaum.

Brooks, L. R., Squire-Graydon, R., & Wood, T. J. (2007). Diversion of attention in everyday concept learning: Identification in the service of use. *Memory & Cognition, 35*, 1-14.

Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within-category correlations. *Memory & Cognition, 3*, 353-62.

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, 1*, 216-26.

Johansen, M. K., & Kruschke, J. K. (2005). Category representation for classification and feature inference. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 6*, 1433-58.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 4*, 592-613.

Medin, D.L., Wattenmaker, W.D., & Hampson, S.E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19*, 242-279.

Rehder, B., & Hoffman, A. B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology, 1*, 1-41.

Rehder, B., & Hoffman, A. B. (2005b). Thirty-something categorization results explained: selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 5*, 811-29.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs , 75* 3. (13, Whole No. 517).

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

Yamauchi, T., & Markman, A. B. (1998). Category Learning by Inference and Classification. *Journal of Memory and Language, 39*, 124-48.

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology. Learning, memory, and cognition, 3*, 776-95.

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, 3*, 585-93.