

Eyetracking and Selective Attention in Category Learning

Bob Rehder (bob.rehder@nyu.edu)

Department of Psychology, New York University, 6 Washington Place
New York, NY 10003 USA

Aaron B. Hoffman (abh226@nyu.edu)

Department of Psychology, New York University, 6 Washington Place
New York, NY 10003 USA

Abstract

An eyetracking version of the classic Shepard, Hovland and Jenkins (1961) experiment was conducted. Forty years of research has assumed that category learning includes learning how to selectively attend to only those stimulus dimensions useful for classification. We confirmed that participants learned to allocate their attention optimally. However, we also found that neither associationist accounts of gradual learning nor hypothesis-testing accounts accurately predicted the pattern of eye movements leading up to successful learning. The implication of these results, and the use of eyetracking technology more generally, for categorization theory are discussed.

Selective attention has played a prominent role in most theories of categorization ever since Roger Shepard's influential work (Shepard, Hovland & Jenkins, 1961) demonstrating that a simple stimulus generalization account of category learning is untenable. The stimulus generalization account regarded category learning to be a process of simple associations between stimuli and category labels. This account predicted that it should be easy for participants to associate stimuli that shared many features with one category label, and difficult to associate such stimuli with different labels. Unexpectedly, one important determiner of difficulty was the number of stimulus dimensions needed for correct classification. It has been generally accepted that this pattern of results is best understood in terms of learners optimally allocating their *selective attention* to those dimensions diagnostic of category membership (Rosch & Mervis, 1975).

Currently, such selective attention is often explicitly parameterized in computational models of categorization, as it is in Nosofsky's generalized context model (1986), or in other descendents and extensions of Medin and Schaffer's (1978) exemplar-based context model of categorization such as EGCM (Lamberts, 1998). It is similarly defined in prototype models, be they based on additive or multiplicative similarity (Nosofsky, 1992; Smith & Minda, 1998). Rule-based models as well implicitly assume the operation of selective attention to those stimulus dimensions referred to by the current hypothesis (i.e., rule) being tested (Smith, Patalano, & Jonides, 1992).

In more recent years, categorization theory has developed

to include the mechanisms by which selective attention changes during the course of learning. One prominent example is Kruschke's (1992) ALCOVE, a connectionist implementation of the generalized context model that changes attention weights as a function of error feedback. Another is Nosofsky's (1994) rule-plus-exception (RULEX) model of classification learning which first performs hypothesis testing of single-dimension rules (then on multi-dimensional rules and/or exceptions to those rules).

Despite its prominence in all modern theories of categorization, evidence for the operation of selective attention in category learning has always amounted to demonstrations that dimensions vary in their influence on explicit categorization judgments, not the operation of selective attention *per se* (Lamberts, 1998). Accordingly, this study had two main goals. The first was to determine if eyetracking data would provide direct support for the claim that learners allocate their attention so as to optimize classification performance. The second was to determine whether the changes in the pattern in eye movements during the course of learning was well described by ALCOVE, RULEX, or either model. To these ends, we performed a replication of the Shepard et al. (1961) category learning experiment with an eyetracker.

The Shepard et al. (1961) Study

Shepard et al. (1961) constructed stimuli defined by three binary-valued dimensions, resulting in eight stimuli which could then be split into two categories. Division of stimuli into categories was determined by six unique category structures, four of which are shown in Figure 1. Here, the dimensions have been arbitrarily instantiated by shape (circle vs. triangle), color (black vs. white), and size (large vs. small).

Type I is the most basic category structure, requiring information from only a single dimension for correct classification (size in Fig. 1). The Type II structure is an exclusive-or problem along two relevant dimensions (size and shape in Fig. 1). Type IV is a single dimension plus exception category structure (as were Shepard et al.'s Types III and V, not shown in Fig. 1). Finally, the most complicated structure is Type VI in which all three dimensions are equally important for category membership. Shepard et al.'s central finding was that the ordering among the category structures from least to most difficult was Type I < II < (III, IV, V) < VI. Because this ordering mirrors the number of dimensions

needed to solve each problem, it was taken as evidence for the importance of selective attention in category learning.

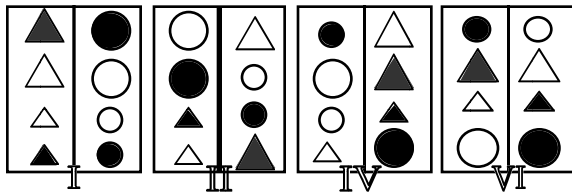


Figure 1

We tested participants wearing an eyetracker on the four category structures shown in Figure 1. Our first question was whether, as predicted by all current theories of category learning, participants would limit their attention to only those stimulus dimensions needed to solve the Type I, II, IV, and VI problems: 1, 2, 3, and 3 dimensions. Our second question was whether the changes in attention (i.e., eye movements) during learning would support a gradual or rule-based learning account. According to ALCOVE, participants should begin by examining all stimulus dimensions, and then reduce the dimensions they fixate to the minimum (to 1 for Type I and 2 for Type II). According to RULEX, all participants should begin by examining one stimulus dimension, and then increase the dimensions they fixate as needed (to 2 for Type II and 3 for Types IV and VI).

Method

Participants

A total of 72 New York University undergraduates participated for class credit or for pay. Participants were tested individually and randomly assigned to conditions.

Materials

Because we were looking at eye movement data, it was crucial for the dimensions of our stimuli to be separated in space (as in Experiment I of Shepard et al., 1961). The binary dimensions were realized by a pair of text symbols (\$ and ¢, ? and !, and + and -) which we presented on a 20 inch computer monitor set to a resolution of 1024 x 768. The symbols were a light gray, RGB (128,128,128), Times New Roman, 30 points, and bolded, situated 818 pixels apart on the monitor forming an equilateral triangle. Figure 2 provides an example of one such stimulus. The assignment of physical dimensions and location to the abstract category structure was counterbalanced across subjects.

We used the SMI Eyelink eyetracking system. For efficiency, we recorded from a single eye. All eye movements, button presses, and RTs were recorded.

Procedure

Each participant was first fitted and calibrated to the eyetracker. Each subsequent learning trial consisted of a drift correction in which the participant fixates on a small circle that appears at the center of the stimulus allowing the eyetracker to make small calibration adjustments that compensate for slight movements (drifts) of the eyetracker on the participant's head. Following the drift correction, one of the

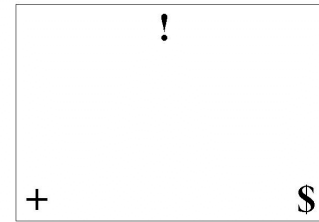


Figure 2

eight exemplars was presented on the screen. Participants classified the exemplar as belonging to either the "red" or "green" category by pressing the corresponding red and green buttons on a button box (assignment of categories to the red or green labels was counterbalanced). Immediately after their response they heard either a chime indicating that they were correct or a low buzz indicating they were incorrect. The exemplar remained visible for 4 sec. after feedback. The stimuli were presented randomly in blocks of 8. The experiment ended if the participant completed 4 blocks in a row without error, or if they completed all 28 blocks. After each block, participants were informed how close they were to this goal.

Eyetracking Dependent Variables

In addition to standard learning measures such as error rates and number of blocks to criterion, we derived a number of measures based on the eyetracking data. We first defined areas of interest (AOIs) as rectangles that encompassed the physical location of each of the symbol dimensions on the computer screen. Based on these AOIs we computed three measures: number of dimensions fixated, proportional fixation time, and dimension priority.

Number of dimensions fixated (ranging from 0 to 3). This is a basic measure which can be thought of as the number of dimensions from which a participant sought information on a particular trial. This measure is intended to tell us which dimension(s) participants use to classify a stimulus.

Proportional fixation time (ranging from 0 to 1). This is the amount of time spent fixating each dimension divided by the total time spent fixating all three dimensions. This measure is intended to provide information regarding which dimension(s) participants found to be most important for the classification decision.

Dimension priority (ranging from 0 to 1). Dimension priority captures the relative order of fixations. Priority is calculated by weighting fixations to a dimension more heavily to the extent they occurred earlier in a trial. The total priority across dimensions sums to 1.

Results

We first set out to establish that this experiment replicated the basic ordering of problem type difficulty found by Shepard et al. (1961). The number of participants out of 18 that reached the learning criterion of four perfect blocks in a row was 18, 18, 15, and 10 for problem Types I, II, IV, and VI, respectively. For each participant we also computed the number of blocks to criterion (for the Types IV and VI non-learners we made the highly conservative assumption that they would successfully solve the problem by block 29).

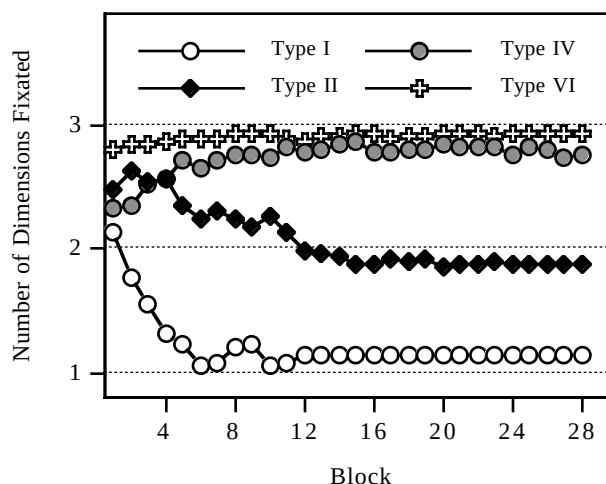


Figure 3

The average number of blocks to criterion was 7.11, 14.11, 18.11, and 22.94 for Types I, II, IV, and VI, respectively. A one-way ANOVA demonstrated that these differences were statistically reliable, $F(3, 68) = 24.8$, $MSE = 32.5$, $p < .01$; all pairwise comparisons (I vs. II, II vs. IV, IV vs. VI) were significant ($p < .05$). Finally, the total number of errors committed for the four problems types was 8.17, 31.17, 36.89, and 70.56; $F(3, 68) = 23.4$, $MSE = 510.9$, $p < .01$ (all pairwise comparisons $p < .05$, except the Type II vs. IV contrast, $p < .15$). Thus, this experiment indeed replicated the basic problem type ordering: Type I < II < IV < VI.

One primary goal of this study, and a first for the categorization field, was to determine if selective attention can be measured directly from eye movements. Figure 3 presents the average number of dimensions fixated for learners only in each category structure in each block. For participants who reached criterion before the 28th block, we assumed their eye movement data for the remaining blocks would have been identical to their last actual four blocks.

Figure 3 illustrates that learners in this experiment indeed allocated their attention (as measured by eye movements) to only those stimulus dimensions needed to solve the classification problem. By the end of learning, the Type I group was examining ~1 stimulus dimension; only 1 of the 18 Type I participants was not restricting his or her eye movements to the one relevant dimension. Similarly, the Type II group was attending to ~2 stimulus dimensions; only 2 of the 18 Type II participants were examining all three dimensions. Finally, as expected all Type IV and VI participants were fixating ~3 dimensions. These results provide direct evidence that the acquisition of categories involves selective attention to only those dimensions needed for judging category membership.

A second goal of the present study concerns the process by which participants reached their final pattern of selective attention. We considered two possibilities. The first, based on ALCOVE, was that attention would shift gradually to the relevant dimensions. The second, based on RULEX, was that attention would first be allocated to a single dimension (as simple 1D rules were being tested) and then shift to include more dimensions as needed. As Figure 3 indicates, the average group data clearly supports an ALCOVE-like gradual

learning view of selective attention. But Figure 3 is a result of averaging over participants. Does gradual learning hold when participants are considered individually?

Learning to Attend Selectively

To answer this question we examined the pattern of eye movements for each of 18 participants assigned the Type I problem. The Type I problem is ideal for this purpose because it is associated with the greatest reduction in the number of dimensions fixated, and hence the greatest change in selective attention, during the course of learning. Although at a detailed level there was of course a great deal of variety across subjects, we found that the pattern of eye movements of 15 of the 18 participants were qualitatively similar. This pattern is exemplified by the eyetracking data of the one Type I participant shown in Figure 4a-c.

Figure 4a presents the number of dimensions examined by this participant in each of the participant's 56 trials. As Figure 4a indicates, in the first 21 trials this participant typically fixated on all three dimensions (s/he fixated on 2 dimensions on 6 trials, and on 1 dimension on 1 trial). However, starting on trial 22, and continuing for the rest of the experimental session, only one stimulus dimension was fixated. Note that rather than the gradual shift of attention from ~2.5 dimensions to ~1 dimension suggested by the Type I group data shown in Figure 3, this participant exhibits a sudden shift of attention to a single dimension. (The fitted lines shown superimposed on the data in Figures 4a-d are explained below.)

Figure 4b presents the proportion of time the participant fixated on the one relevant dimension. A trial in which all three dimensions are examined an equal amount of time results in a proportion time score of 0.33; one in which only the relevant dimension is examined results in a score of 1.00. The figure indicates that in the first 21 trials the participant did not spend appreciably more time fixating the relevant dimension than the other two dimensions. Starting with trial 22 however, only the relevant dimension was fixated. This data confirms that the single dimension fixated beginning with trial 22 (Figure 4a) was indeed the relevant one (i.e., the one needed to respond correctly).

Finally, Figure 4c presents the relative priority of the relevant dimension. A trial in which the three dimensions were examined in random order would produce a priority score of 0.33, and this score increases as the participant tends to fixate relevant dimensions before the other dimensions. Figure 4c indicates that until trial 21, the participant showed virtually no preference for looking at the relevant dimension first. After trial 21, the relative priority score becomes 1.00, a result entailed by the fact that starting with that trial only the relevant dimension is fixated.

Taken together, Figures 4a-c suggest that this participant exhibits none of the signs of gradual learning suggested by the Type I group data. Up until trial 21, the participant typically examines all three dimensions, spends about as much time examining the relevant dimension as the irrelevant ones, and shows no preference for looking at the relevant one first. Starting with trial 22 and continuing until the learning criterion is reached on trial 56, only the relevant dimension is fixated. The "suddenness" of learning suggested

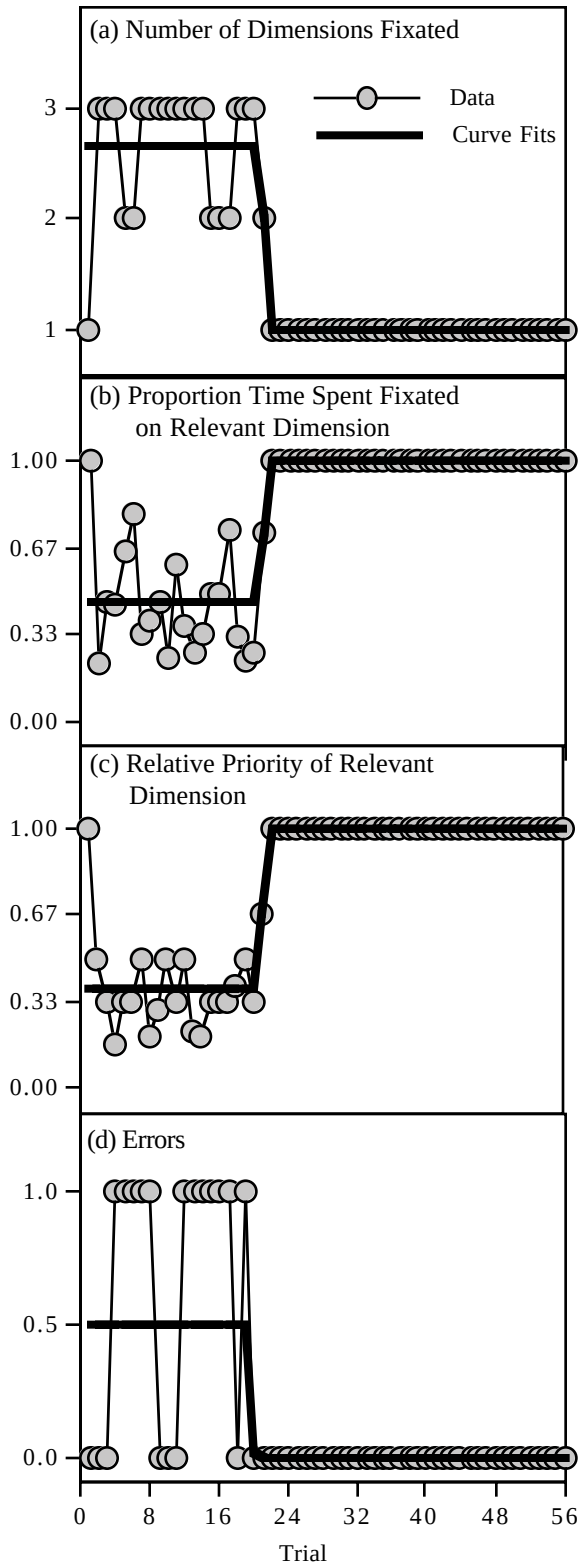


Figure 4

by these results is directly confirmed by the pattern of errors (Figure 4d). Whereas during the first twenty trials the participant shows no indication of gradually reducing his/her error rate (e.g., 5 errors committed in trials 1-10 followed by 7 in trials 11-20), errors cease entirely after trial 20.

To characterize the changes that took place in the pattern of eye movements quantitatively, we fit sigmoid functions to the results shown in Figure 4. Specifically, we fit the following sigmoid function to the participant's four dependent variables,

$$y = \text{initial} + \text{diff} / (1 + \exp(-(mt + b)))$$

where y is the dependent variable being fit, initial is the initial asymptote of the sigmoid, diff is the magnitude of the change of the sigmoid from its initial asymptote to its final asymptote, m is a measure of whether that change occurs slowly or rapidly, b is the inflection point of the curve, and t is trial number. (For the error fit, we set $\text{initial}=0.50$ and $\text{diff}=-0.50$ reflecting initial guessing and eventual learning).

The results of these fits are shown superimposed on the empirical data in Figure 4. For example, the parameters for the fit to the number of dimensions fixated (Figure 4a) was $\text{initial}=2.65$, $\text{diff}=-1.65$, $m=10$, and $b=21.0$. These parameter estimates indicate that this subject began by fixating on 2.65 dimensions, ended up fixating $2.65-1.65=1$ dimension, the transition from 2.65 to 1 occurred rapidly ($m=10$), and occurred at trial 21. The fits of the sigmoid functions in Figures 4b-d also confirm the suddenness of the transition on all three measures. Moreover, the value of the b parameter in all four fits confirms that the transitions occurred within a trial or two of one another ($b=21.0$, 20.0, 21.0, 19.7 for number of dimensions, relevant fixation time, priority, and errors, respectively).

This fitting procedure was carried out for all 18 Type I participants, and the parameter values for each of the four fits averaged over participants are presented in Table 1. Consider first the fits for the number of dimensions fixated. The averaged parameter values indicate that the average participant began by fixating 2.80 dimensions, ended fixating $2.80-1.63=1.17$ dimensions, and made the transition at about trial 13. Importantly, the average value of the m parameter (6.08) suggests that this transition from 2.80 to 1.17 dimensions occurred abruptly for most Type I subjects.

Table 1

Parameter	Dependent Variable			
	# of Dimensions	Proportion Time	Priority	Errors
<i>initial</i>	2.54	0.26	0.37	0.50
<i>diff</i>	-1.52	0.69	0.53	-0.50
<i>m</i>	7.21	5.93	7.59	6.80
<i>b</i>	16.59	16.60	18.54	14.97

To illustrate this result directly, we constructed "backward learning curves" in which the curve fits for each participant were aligned with one another by translating each participant's trial number so that 0 corresponded to the value of the b parameter, that is, the inflection point of the fitted sigmoid. These translated curves for each Type I participant are shown in Figure 5. As expected given the average values of the parameters shown in Table 1, Figure 5 indicates that most Type I participants began by fixating between 2.5 and 3 dimensions, and all ended fixating the single relevant dimension. Moreover, for all but 3 of the participants, this reduction in the number of dimensions took place within a few trials. Figure 5 also shows the sigmoid produced by the average parameter values shown in Table 1 for the number

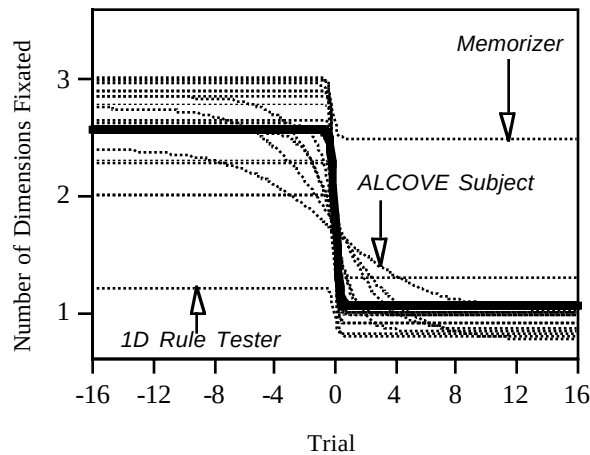


Figure 5

of dimensions fixated. This curve, which represents the performance of the average Type I participant, shows that the number of dimensions fixated undergoes a sudden decrease from 2.6 to <1.2 dimensions in only two trials.

The three exceptions to the general pattern in Figure 5 have been labeled according to the alternative category learning strategy that these participants apparently used. First, the *one-dimensional rule tester* looked at just one dimension the entire session. We speculate that this person systematically tested one-dimensional rules until the correct one was found. Second, the *ALCOVE participant* exhibited the gradual shift of attention toward the irrelevant dimensions and toward the relevant one predicted by ALCOVE. Finally the *memorizer* fixated all three dimensions the entire session. We speculate that this person systematically memorized the category membership of all eight stimuli.

The curve fits produced by the average parameter values in Table 1 for all four dependent measures were computed, and are presented in Figure 6. Figure 6 characterizes how performance according to all four measures changes as a function of trial for the average Type I participant. The figure illustrates how the four effects all occur within a few trials of one another. Before the critical learning point, the learner's chance of making an error is close to 50%, the number of dimensions fixated is close to three, the proportion of time spent fixating the relevant dimension is less than one-third, and the relevant dimensions is no more likely to be fixated before the other dimensions. After the critical learning point, errors have largely ceased, the average number of dimensions fixated is less than 1.5, the proportion of time spent on the relevant dimension is over 90%, and the relevant dimension is fixated early in the trial.

General Discussion

Since Shepard et al.'s (1961) seminal study a core assumption of virtually all theories of categorization has been that category learning involves learning to attend to those stimulus dimensions that are useful for discriminating between categories. However, evidence for this claim has consisted of demonstrations that dimensions vary in their influence on explicit categorization (and similarity) judgments, not the operation of selective attention per se. To our knowledge, the current results provide the first direct evidence for the

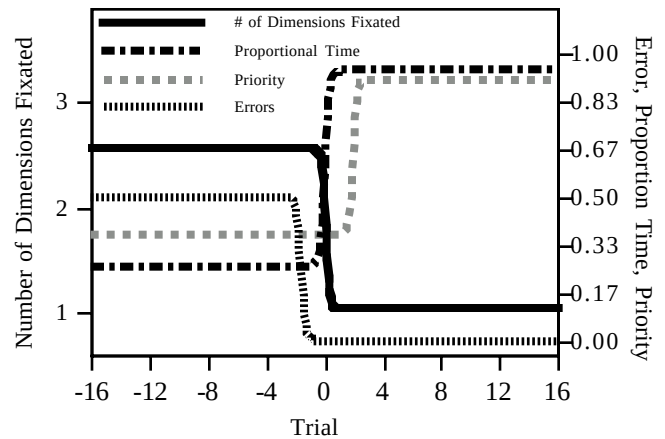


Figure 6

operation of selective attention in category learning.

Our findings provide strong support for the basic claim that categorizers learn to allocate their attention in a way that optimizes their classification performance. Of the 18 Type I participants, only one had failed to reduce the number of dimension they fixated to the single relevant dimension by the end of the experimental session. Similarly, by the end of the session all but two of the 18 Type II participants were fixating almost exclusively on the two dimensions relevant to that category structure. That is, learners apparently confine their attention to those aspects of stimuli needed to succeed at the classification task at hand.

Since Shepard et al. (1961), an important development in categorization theory has been the development of computational models that formalize the mechanisms by which selective attention changes as a result of experience classifying exemplars. One prominent example is the ALCOVE connectionist learning model that predicts that attention weights gradually shift in the direction of better performance as learning experiences accumulate. In contrast to the gradual change of attention suggested by ALCOVE however, our eyetracking data indicated that the shift of attention to the relevant dimension in the Type I problem occurred very abruptly, in only a trial or two. Before that transition, most Type I subjects showed no preference for looking at the relevant dimension any more than, or any earlier than, the other dimensions.

Of course, the possibility exists that most Type I participants were gradually assigning more weight to the relevant dimension in their explicit categorization decisions (as ALCOVE predicts), but that this change was not reflected in eye movements. However, this interpretation is undermined by the fact that most participants' error rates also showed no improvement until what we have characterized as a critical learning event occurred. That is, participants' explicit categorization behavior and their eye movements were well matched to one another: Before the critical learning event, eye movements showed no preference for the relevant dimension, and error rates were about 50%. After it, eye movements showed exclusive preference for the relevant dimension, and errors ceased. Given this apparent "all-or-none" learning, we follow Bower and Trabasso (1963) and Nosofsky et al. (1994) and suggest that results taken as support

for gradual learning may often be an artifact that results from averaging over participants.

On the one hand, the all-or-none learning we found for Type I problems is more consistent with a rule-based model such as RULEX than ALCOVE. However, our results present problems for RULEX as well. Although RULEX predicts all-or-none learning for Type I problems, it also predicts that learners will initially consider single-dimension rules as possible hypotheses regarding what defines category membership. Instead, we found that on most trials almost all Type I participants fixated on all three dimensions up until learning occurred.

Once again, it is conceivable that participants were in fact testing single dimension rules (as RULEX predicts), and that the additional two or three dimension fixated were extraneous, that is, did not involve acquiring information that was useful for the learning process. However, this suggestion is incompatible with an overall view of category learners as optimal allocators of attention. In the forty years since Shepard et al. (1961), theorists have assumed, and the present experiment has confirmed, that categorizers largely allocate their attention to just those aspects of the stimulus they believe are needed for successful classification. Analogously, we suggest that learners largely allocate their attention to just those aspects of stimuli they believe will be needed for successful learning. That is, rather than assuming that our Type I learners were committing extraneous eye fixations, we argue that it is more likely that they thought they could make use of the information on that which they fixated, and that it is incumbent on categorization theorists to account for why they sought out the sources of information that they did. On these grounds, we conclude that models like RULEX that assume that learners start off by (only) testing single-dimension rules cannot be considered complete accounts of participants' learning strategies.

If our Type I learners were not involved in ALCOVE-like gradual learning, or the testing of single-dimension rules, what were they doing? One possibility is that they were testing multi-dimension rules. Another is that they began by trying to memorize the individual exemplars, but then "noticed" (somehow) that one dimension covaried consistent with the category label. This noticing could be based on comparing the current exemplar with the single previous exemplar stored in working memory (Anderson, Kline, & Beasley, 1979), or with multiple examples stored in long-term memory (Ross, Perkins, & Tenpenny 1990). Examining all stimulus dimensions is also consistent with the growing trend toward considering category learning as involving multiple learning processes. For example, according to Kruschke (2001), acquiring categories may involve learning which of possibly several learning "experts" (e.g. a 1D rule-based learning module and exemplar learning) is producing the best performance. Examining all stimulus features allows as many experts as possible to be involved in the learning process. Finally, we note that because real-world category learning often involves categories with meaningful rather than meaningless stimulus dimensions, our participants may have been looking to find (or construct) meaningful relationships between category features.

To answer these and other questions we are continuing our

analyses of the rich set of eyetracking data produced for the Type I category structure, the other three Shepard structures, and yet other learning problems. Still, we believe even the initial results reported here establish the usefulness of eyetracking to measure selective attention in category learning. After several decades of building categorization theories on the basis of binary choice data alone, we believe that the field of categorization is ready for the introduction of more sophisticated measures. We expect that use of a head-mounted eyetracker will provide the field with a rich source of empirical data that will help discriminate among existing models and help advance cognitive theory in this area.

References

- Anderson, J. R., Kline, P. J., & Beasley, C. M. (1979). A general learning theory and its applications to schema abstraction. In G. H. Bower (Eds.), *The Psychology of Learning and Motivation*. (pp. 277-318). New York: Academic Press.
- Bower, G. H., & Trabasso, T. R. (1963). Reversals prior to solution in concept identification. *Journal of Experimental Psychology*, 66, 409-418.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22-44.
- Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, 45, 812-863.
- Lamberts, K. (1998). The time course of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 695-711.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207-238.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology*, 115, 39-57.
- Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, 101, 53-79.
- Rosch, E. H., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.
- Ross, B. H., Perkins, S. J., & Tenpenny, P. L. (1990). Reminding-based category learning. *Cognitive Psychology*, 22, 460-492.
- Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, 75, Whole No. 517.
- Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1411-1436.
- Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, 65, 167-196.