Eyetracking and Selective Attention in Category Learning

Bob Rehder

Aaron B. Hoffman

Department of Psychology
New York University

Address:
Bob Rehder
Department of Psychology
6 Washington Place
New York, NY 10003
Phone: (212) 992-9586
Email: bob.rehder@nyu.edu
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. We also found that learners tend to fixate all stimulus dimensions early in learning. This result obtained despite evidence that participants were also testing one-dimensional rules during this period. Finally, we found that the restriction of eye movements to only relevant dimensions tended to occur only after errors were largely (or completely) eliminated. We interpret these findings as consistent with multi-systems theories of learning which maximize information input in order to maximize the number of learning modules involved, and which focus solely on relevant information only after one module has solved the learning problem.
Eyetracking and Selective Attention in Category Learning

Selective attention has played a prominent role in most theories of categorization ever since Roger Shepard’s influential work (Shepard, Hovland & Jenkins, 1961) demonstrated that a simple stimulus generalization account of category learning is untenable. The stimulus generalization account took 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 (Shepard et al. 1961; Medin & Schaffer, 1978)

Currently, selective attention is an integral component of all major categorization theories. For example, in both exemplar models (Medin & Shaffer, 1978; Nosofsky, 1986) and prototype models (Nosofsky, 1992; Smith & Minda, 1998), selective attention is formalized in terms of explicit attention weights on the stimulus dimensions. Rule-based models also 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).

Moreover, in more recent years, these theories have been extended to include the mechanisms by which selective attention changes with learning. One prominent example is Kruschke’s (1992) ALCOVE, a connectionist exemplar model that changes attention weights as a function of error feedback. Another is Nosofsky, Palmeri, and McKinley’s (1994) rule-plus-exception (RULEX) model, which first performs hypothesis (rule) testing on single dimensions, then on multi-dimensional rules and/or exceptions
to those rules if needed.

Despite its prominence in modern categorization theory, however, evidence for the operation of selective attention has always amounted to demonstrations that dimensions vary in their influence on explicit categorization judgments, but not on 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 support the claim that learners allocate their attention to optimize classification performance. To this end, we replicated the Shepard et al. (1961) category learning experiment with an eyetracker. Specifically, we asked whether Shepard et al.’s claims regarding learners’ reallocation of attention to only those stimulus dimensions relevant to producing correct classification decisions would be directly corroborated by eyetracking data.

To our knowledge, the current work is the first to apply eyetracking to the domain of categorization research. At the outset then, one concern that must be addressed is the interpretation of eye movements as a surrogate measure of attention during category learning. It is of course well known that attention can dissociate from eye gaze under certain circumstances (Posner, 1980). However, in many cases changes in attention are immediately followed by the corresponding eye movements (e.g., Kowler et al., 1995), and there is evidence that attention and eye movements are tightly coupled for all but the simplest stimuli (Deubel & Schneider, 1996). Not surprisingly then, eye tracking has proven to be an effective tool in many areas of research, most notably of course reading (Ferreira & Clifton, 1986; Maki, Vonk & Schriefers, 2002; Rayner, 1998; Tanenhaus et al., 1995; Just & Carpenter, 1984) but also language production (Griffin & Bock, 2000; Meyer et al., 1998), scene perception (Biederman et al., 1982; Henderson, 1999; Loftus & Mackworth, 1978), problem solving (Grant & Spivey, 2003; Hegarty & Just, 1993), and face perception (Althoff & Cohen, 1999), to name a few. In the current study, we will take the presence of eye fixations to spatially-separated stimulus dimensions as a proxy measure of attention to those dimensions, and predict
that fixations to dimensions irrelevant to correct classification will cease as a result of classification experience. An important feature of the category learning task is the availability of an overt behavioral measure (the elimination of classification errors) as a source of converging evidence about which aspects of stimuli are being attended; specifically, learning entails that a participant attend to those stimulus dimensions needed to discriminate members of the categories. Thus, confirmation that learners only attend to relevant dimensions will not only corroborate the basic claim of Shepard et al.’s, it will also cross-validate the use of eyetracking as an index of attention in category learning.

The second goal of our study was to determine whether the manner in which attention (as measured by eye movements) changes during the course of category learning was well described by ALCOVE, RULEX, or either model. According to ALCOVE, learners will generally start off allocating attention to all stimulus dimensions equally (or perhaps in a manner that reflects differences in their physical salience), and then gradually shift attention to only relevant dimensions as a result of error feedback. In the experiment which follows, dimensions will be of roughly equal salience, and thus, our ALCOVE-based predictions are that learners will initially spend an equal amount of time fixating each stimulus dimension. As learning proceeds, fixations to irrelevant dimensions will gradually decrease until they are eliminated altogether.

In contrast, a hypothesis-testing model like RULEX makes very different predictions regarding how selective attention changes as a function of learning. According to RULEX, learners first search for a single-dimension rule that successfully discriminates members of the two categories. Thus, our RULEX-based eye movement predictions are that learners will fixate single dimensions early in learning. When no single-dimension rule is found, learners will fixate multiple dimensions as they attempt to form more complex rules (e.g., conjunctions, disjunctions, etc.), or to memorize
exceptions to an imperfect rule. That is, whereas ALCOVE predicts that learners will initially fixate all dimensions and then gradually reduce the number fixated to the minimum needed, RULEX predicts that they will first fixate one dimension, and then increase the number fixated as needed.

Once again, an important characteristic of the category learning task is the presence of an overt behavioral measure in the form of classification errors that can corroborate any conclusions we reach regarding changes in selective attention on the basis of eye movements. For example, one diagnostic feature of hypothesis-testing models is the all-or-none learning (i.e., the sudden elimination of classification errors) that obtains when a learner discovers a correct single-dimension rule (Bower & Trabaasso, 1963). Thus, the RULEX-based prediction is that the fixations to a single dimension which are supposed to reflect rule application should be closely accompanied by the elimination of classification errors when that dimension is one which can be used to discriminate category members. Similarly, an important characteristic of associationist learning models like ALCOVE is the gradual learning that obtains as a result of the incremental adjustment of connection weights on the basis of error feedback. Thus, the ALCOVE-based prediction is that a gradual shift of eye movements should be accompanied by a gradual decrease of errors. More generally, a close correspondence between error reduction and changes in eye movements will not only provide evidence for one or the other model of learning, it would also validate eyetracking as an effective measure of the changes in selective attention during category learning.

Although we believe our predictions provide a useful initial framework for the evaluation of eye movements in category learning, we acknowledge that there are good reasons to expect that the relationship between selective attention and eye movements might be considerably more complex than we have assumed. For example, eye fixations may diverge from attention weights because there are components of the category
learning process (ones not modeled by ALCOVE or RULEX) that lead learners to fixate stimulus dimensions. And, eye fixations may be influenced by goals people have other than category learning per se. Nevertheless, the application of eyetracking to category learning is new, and thus we believe that for now our (perhaps overly simplistic) predictions provide a useful starting point for evaluation of eye movements in the Shepard et al. (1961) category structures. In the General Discussion, we will reevaluate the relationship between selective attention and eye movements in light of our experimental results.

The Shepard et al. (1961) Study

Shepard et al. (1961) constructed stimuli with three binary-valued dimensions, resulting in eight stimuli split into two categories. There were six unique divisions of stimuli into categories, four of which are shown in Figure 1A. Here, the dimensions have been arbitrarily instantiated by shape, color, and size.

Type I is the most basic category structure, requiring information from a single dimension for classification (the shape dimension in Fig. 1A). The Type II structure is an exclusive-or problem along two relevant dimensions (size and shape in Fig. 1A). Type IV can be described as single-dimension-plus-exception structure (as can Types III and V, not shown in Fig. 1A), in which all three dimensions are relevant although not equally so. Type IV can also be characterized with a “2 out of 3” decision rule in which all dimensions are equally relevant. Finally, in the Type VI structure, all three dimensions are equally relevant and categorizers must memorize the category label for every exemplar. Shepard et al.’s central finding was that the ordering among the category structures from least to most difficult was Type I < II < IV < VI (also see Love, 2002; Nosofsky, Gluck, Palmeri, McKinley, & Glaughier, 1994). Because this ordering mirrors the number of dimensions for correct classification, it was taken as evidence for selective attention in category learning. (The greater difficulty of Type VI vs. IV was
taken as reflecting VI’s lower between-category and greater within-category similarity, consistent with Shepard et al.’s original predictions.)

We tested participants wearing an eyetracker on the four category structures shown in Figure 1. However, because eye movement analysis requires the dimensions of stimuli to be separated in space, our stimuli were in fact analogous to those used in Experiment I of Shepard et al. (1961). An example of the stimuli used in that experiment is presented in Figure 1B. For example, a Type I problem could be constructed from the stimuli in Figure 1B on the basis of the bottom left “dimension”: Items with a candlestick would form one category and those with a lightbulb would form the other. However, to avoid the perceptual complexity of the features in Figure 1B, in our experiment the binary dimensions were realized instead by a pair of characters ($ and ¢, ? and !, and x and o). An example of a single stimulus used in the current experiment is presented in Figure 2.

Our first question was whether, as predicted by current theories, participants would limit their attention (measured by eye movements) to only those stimulus dimensions needed to classify each structure: 1, 2, 3, and 3 dimensions. Our second question was whether the changes in eye movements during learning would support a gradual or rule-based learning account. According to ALCOVE, participants should begin by examining all dimensions and gradually reduce the dimensions they fixate to the minimum (to one for Type I and two for Type II). According to RULEX, participants should begin by examining one dimension, and increase the dimensions they fixate as needed (to two for Type II and three for Types IV and VI).

Method

Participants

A total of 72 New York University undergraduates were randomly assigned to one of the four category structures.
Materials

The characters which comprised the stimuli (\$ and \&, ? and !, and x and o) appeared as a light gray (RGB: 128, 128, 128) and within ~1/2 by ~1 degree of visual angle. The three symbols were situated ~20 degrees apart on the CRT at ~12 degrees eccentricity, forming an equilateral triangle. An example stimulus is presented in Figure 2. The assignment of physical dimensions and location to the abstract category structure was counterbalanced.

Our SMI Eyelink eyetracking system corrected for drift between trials, recording a single eye.

Procedure

Each participant was first fitted and calibrated to the eyetracker. Each subsequent learning trial began with a drift correction in which the participant fixated on a small circle that appeared at the center of the CRT allowing the eyetracker to make small calibration adjustments that compensate for slight movements (drifts) of the eyetracker on the participant’s head. Following drift correction, one of the eight exemplars was presented on the screen. Participants classified the exemplar as belonging to either a “red” or “green” category by pressing colored buttons on a button box (assignment of categories to the red or green labels was balanced). Exemplars remained visible for 4s after auditory feedback. Exemplars were presented randomly in blocks of 8. The experiment ended after four consecutive errorless blocks or after a 28 block maximum. Participants were informed how close they were to this goal after each block.

Eyetracking Dependent Variables

To derive eyetracking measures, we defined three rectangular areas of interest (AOIs) that encompassed the symbol dimensions on the CRT (Fig. 2). Based on these AOIs we computed three measures for each learning trial. The first of these is the number of dimensions fixated (ranging between 0 and 3). The second, proportion fixation
time (ranging from 0 to 1), is the time spent fixating each dimension divided by the total time spent fixating all three dimensions. It provides information regarding which dimensions participants found most important. Finally, the relative priority (ranging from 0 to 1) captures the ordering of fixations. To compute this measure we weighed each fixation on a dimension according to the terms in the arithmetic sequence, \{n, n–1, ..., 1 \}, of n ordered fixations such that the first fixation of the trial was given a weight of n, the second fixation was given a weight of n–1, and the last fixation was given a weight of 1. Thus, dimensions receive a greater relative priority score the earlier in the trial they are fixated.

Results

We first set out to establish that we replicated the basic ordering of problem 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, 16, and 10 for Types I, II, IV, and VI, respectively. We also analyzed the number of blocks to criterion (assuming, conservatively, that nonlearners would have reached perfect performance by block 29). The average number of blocks to criterion was 7.1, 14.1, 18.1, and 22.9 for Types I, II, IV, and VI, respectively; F(3, 68) = 24.8, 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 8.2, 31.2, 36.9, and 70.6; F(3, 68) = 23.4, 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.

Our primary goal, 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 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 the mean of
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 one dimension; only 1 of the 18 Type I participants did not restrict eye movements to the one relevant dimension. Similarly, the Type II group was attending to two dimensions; only 2 of the 18 participants examined all three dimensions. Finally, all Type IV and VI participants generally fixated three 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 learned to attend selectively. 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 support an ALCOVE-like gradual learning view of selective attention as participants in all groups examined between two and three dimensions. But Figure 3 is a result of averaging over participants. Does gradual learning hold for individuals? To answer this question we examined the pattern of eye movements for each participant individually, starting with those that solved the Type I problem.

**Type I Results**

The Type I problem is ideal for the purpose of examining the role of selective attention in category learning, because it is associated with the greatest reduction in the number of dimensions fixated—and hence the greatest change in selective attention—during learning. Although at a detailed level there was of course a great deal of variety across participants, we found that the pattern of eye movements of 11 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 stimulus dimensions examined by this individual on each trial. Figure 4A indicates that in the first 21 trials this participant typically fixated all three dimensions (except 2 dimensions on 6 trials, and 1 dimension on 1 trial). However, starting on trial 22, and continuing for the rest of the experiment, only one dimension was fixated. Rather than the gradual shift of attention from ~2.5 dimensions to ~1 dimension suggested by the Type I group data (Fig. 3), this participant exhibits a sudden shift of eye movements to a single dimension.

Figure 4B presents the proportion of time the participant fixated the one relevant dimension. A trial in which all three dimensions are examined equally results in proportions of 0.33; one in which only the relevant dimension is examined results in a proportion 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.

Finally, Figure 4C presents the relative priority of the relevant dimension. If the relevant dimension is fixated no earlier or later than other dimensions then its relative priority score is 0.33. This score increases if the participant fixates the relevant dimension before the other dimensions. Figure 4C indicates that until trial 21, the participant showed no preference for fixating the relevant dimension any earlier than other dimensions. After trial 21, the relative priority score becomes 1.00 since at that point it is the first and only dimension 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 by these results is directly confirmed by the pattern of errors (Fig. 4D). Whereas during the first twenty trials the participant shows no indication of a gradually improving 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 shown in Figure 4 quantitatively, we fit the following sigmoid function to the participant’s four dependent variables,

\[ y = \text{initial} + \frac{\text{diff}}{1 + \exp(-m(t - b))} \]

where \( y \) is the dependent variable being fit, \( \text{initial} \) is the initial asymptote of the sigmoid, \( \text{diff} \) 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.\(^1\)

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 (Fig. 4A) was \( \text{initial} = 2.65, \text{diff} = -1.65, m = 5.89, \) and \( b = 21.1 \). These parameter estimates indicate that this participant 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 \((m = 5.89)\), 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 \((b = 21.1, 20.0, 21.0, 19.7\) for number of dimensions, relevant fixation time, relative priority, and errors, respectively). Interestingly, the reduction in number of errors begins to occur a trial or two earlier than the change in eye movements.

This fitting procedure was carried out for all 18 Type I participants. To accommodate those instances in which a dependent measure showed no change over the course of the experimental session, we also fit an intercept model that consisted of a single parameter representing the average value of the measure over all trials. Either the
sigmoid or the intercept model was then chosen as the best fitting model according to a measure (root means square error, RMSE) that took into account the different number of parameters in the two models (4 vs. 1).

We first present the fits to the number of dimensions fixated for each Type I participant. To make these fits comparable, we constructed backward learning curves in which the fits 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 sigmoid. These translated curves are shown in Figure 5A for each Type I participant. Figure 5A shows that most Type I participants began by fixating between 2.5 and 3 dimensions, and all but one ended fixating the single relevant dimension. Moreover, for all but three participants, this reduction in the number of dimensions took place within a few trials. (We discuss the three exceptions labeled “1D rule testers” and “memorizer” below.)

The “average” sigmoid in the Type I condition was calculated by averaging the parameters of the 18 sigmoids. These averaged parameters are presented in Table 1, and the average sigmoid is shown superimposed on the individual curves in Figure 5A. The average sigmoid confirms the sudden restriction of eye movements to the single relevant dimension. The typical participant began by fixating 2.61 dimensions, ended fixating $2.61 - 1.45 = 1.16$ dimensions, and made the transition at about trial 19. Importantly, the average value of the $m$ parameter (1.43) suggests that this transition from 2.61 to 1.16 dimensions occurred abruptly for most Type I participants: An $m$ parameter of 1.43 corresponds to a 90% change in the average sigmoid occurring in just three trials. Table 1 also presents the average parameters for the sigmoids to the proportion time spent fixating on the relevant dimension, and to the relative priority. These average sigmoids corroborate a sudden shift to the single relevant dimension occurring around trial 19.

We also computed backward learning curves from the individual participants’
error sigmoids. These curves are presented in Figure 5B and show that, with a few exceptions, most Type I participants exhibited a sudden reduction in the number of errors from 50% to 0%. The parameters of the average error sigmoid—which is superimposed on the individual curves in Figure 5B—are presented in Table 1. The average value of $m$ for the error fits (2.28) indicates that the reduction in errors from 50% to 0% occurred in about two trials.

These findings indicate that the sudden reduction in number of dimensions fixated and errors exhibited by the individual in Figure 4 holds for most members of the Type I group. However, although the average parameter values presented in Table 1 provide a coarse summary of performance in the Type I condition, Figures 5A and 5B also indicate that there were some exceptions to the general pattern. To characterize this variability, we identified five groups of Type I participants that exhibited unique performance profiles. The profiles of these five groups are presented in Figures 6 and 7. Each panel in these figures characterizes how the number of dimensions fixated, proportion fixation time, relative priority, and errors change over the course of the experiment.

The two learning profiles that accounted for most Type I participants are presented in Figures 6A and 6B. The profile in Figure 6A represents the modal performance in the Type I condition, accounting for 11 participants (including the one in Figure 4). Figure 6A illustrates the sudden elimination of errors and restriction of eye movements to the single relevant dimension. Moreover, this figure illustrates how these effects all occur at about the same time: Before trial 11, the modal Type I participant’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 about one-third, and the relevant dimension is no more likely to be fixated before the other dimensions. By trial 17, errors have ceased and only the relevant dimension is being fixated. Importantly, Figure 6A indicates that the sudden restriction of eye movements to the
single relevant dimension occurs on average about 4 trials after errors have ceased; suggesting that participants focused exclusively on the single relevant dimension only after the category learning problem was already solved. Indeed, a within-subject t-test confirmed that the inflection of the sigmoid for the error fit ($b = 14.5$) occurred significantly earlier than that for the number of dimensions fixated ($b = 18.7$), $t(14) = 6.90$, $p < .05$.

The second major performance profile is presented in Figure 6B. In contrast to modal group in Figure 6A who exhibited all-or-none learning, this group of four individuals exhibited a more gradual reduction in error rates: the average error sigmoid for this group underwent a 90% change in an average of 14.6 trials ($m = 0.30$). These individuals also gradually limited their eye movements to the single relevant dimension. In this regard, the performance of these individuals accords with the predictions of ALCOVE in which error reduction and attention shifts are gradual and co-occur. Just as was the case for the all-or-none learners however, the shift in eye fixations tended to follow rather than precede the reduction in errors: By the time eye fixations begin to show a preference for the single relevant dimension (around trial 25), the average error rate has dropped almost to 0.10.

Figures 7A-C present three exceptions to the dominant profiles shown in Figure 6. Figures 7A and 7B depict the performance of the two individuals we referred to as the one-dimensional rule testers in Figure 5A, because they generally examined only one dimension on each trial of the experimental session. For the participant presented in Figure 7A, this one dimension was always the relevant one. Not surprisingly, this person solved the Type I problem almost immediately (committing only one error on trial 1). In contrast, the participant in Figure 7B began fixating one of the irrelevant dimensions, but then, after committing 7 errors in the first 9 trials, switched to examining only the relevant dimension on trial 10, after which only one additional error was committed on trial 12. Finally, the participant in Figure 7C corresponds to the one
we have labeled the *memorizer* in Figure 5, because he or she fixated all three dimensions the entire session. We speculate that this person systematically memorized all eight stimuli. Consistent with this interpretation is the fact that this individual took 10 blocks to learn the Type I problem, as compared to the group average of 7.1 blocks.

In summary, most Type I participants exhibited the all-or-none reduction in errors characteristic of hypothesis-testing accounts of learning. However, only two participants exhibited the pattern of eye movements we derived from the RULEX model, namely, fixating single dimensions while testing simple one-dimensional rules. Instead, most participants examined all three dimensions early in learning, and only restricted their eye movements to the single relevant dimension several trials after classification errors ceased. Moreover, we also observed a substantial minority of participants (5 of 18) that exhibited gradual rather than all-none-learning. Nevertheless, 4 of these 5 participants performed like the modal group in restricting their eye movements to the relevant dimension, albeit only after the learning problem was already solved.

**Type II Results**

Like the Type I category structure, the Type II structure allows an examination of how people learn to attend selectively to only those dimensions relevant to discriminating the categories, in this case, the two out of three dimensions on which an exclusive-or rule is formed. For each Type II participant, we carried out the same sigmoid fitting procedure on the four dependent measures used to analyze the Type I condition.

We again start by presenting in Figure 8 the results of one participant that exemplifies the modal pattern in the Type II condition. Figure 8 presents the number of dimensions fixated on each of this participant’s 80 trials. This figure indicates that in trials 2–33 all three dimensions were fixated. In this regard, this individual behaves like the typical Type I participant by examining all stimulus dimensions at the beginning of
the experimental session. However, during trials 34-38 the participant alternates between fixating 2 and 3 dimensions, and then, starting on trial 39 and continuing until the final trial 80, generally examines only the two dimensions relevant to solving the Type II problem. On the one hand, as was the case for the Type I results, the reduction in the number of dimensions occurred much more abruptly than implied by the Type II group data presented in Figure 3. On the other hand, the restriction of eye movements to the relevant dimensions occurs more gradually than it did in the Type I condition. This difference in the rate of change in eye movements is reflected in the value of the $m$ parameter for this participant’s Figure 8 sigmoid fit (0.64, corresponding to a 90% change occurring in 6.9 trials) versus the average value of $m$ found in the Type I condition (1.43, 3.1 trials).

The gradual change in attention is more apparent when one examines the sum of the proportion of time spent fixating the two relevant dimensions (Fig. 7B) and the sum of the priority score for those dimensions (Fig. 7C). These measures indicate that the shift to the relevant dimensions in fact began as early as trial 27. That is, even though the participant examines all three stimulus dimensions on trials 27-33, the two relevant dimensions begin to be examined earlier and for a greater proportion of time during these trials. According to both of these measures, the shift of attention to the two relevant dimensions ($m = 0.41$ and 0.39, respectively) occurs over 11 trials (27-38). In comparison, this same shift occurred in ~2 trials in the Type I condition.

Finally, the restriction of attention to the relevant dimensions during trials 27-38 is accompanied by a decrease in the number of errors committed (Fig. 7D). The error rate for this participant is 50% until around trial 27, after which it gradually decreases until the final error is made on trial 41.

In summary, all four dependent variables presented in Figure 8A-D suggest that, although learning for this participant occurred much more abruptly than implied by the average Type II data (Fig. 3), it is nevertheless more gradual as compared to the all-or-
none learning seen in the Type I condition. The decrease in errors and the increase in the proportion of time fixating on the two relevant dimensions start around trial 27, the irrelevant dimension is no longer fixated starting on trial 38, and the final error is committed on trial 41. This performance is consistent with an associationist theory of learning such as ALCOVE in which the reduction in errors and the focusing of attention on relevant information occur gradually and at the same time.

Analogous with the Type I analysis, we began to examine performance of all 18 Type II participants by constructing backward learning curves from the sigmoids for the number of dimensions examined (Fig. 8A) and errors (Fig. 8B). Figure 9A shows that most Type II participants began by fixating between 2.5 and 3 dimension and ended by fixating on 2 dimensions. Thus, as for the Type I problem, we see that learners end up fixating only the dimensions needed to solve the learning problem (in this case two dimensions). One important exception, however, is the participant we have labeled in Figure 9A as the peripheral vision user. This learner showed a gradual reduction in the number of dimensions fixated so that, by the end of the experimental session, he or she was only fixating one stimulus dimension, and doing so despite the fact that correct responding required acquiring information from two dimensions. Apparently, the information from one of the two stimulus dimensions was acquired by use of peripheral vision, that is, without any fixations to that dimension. Of the 62 individuals in the current study who learned their assigned category structure, this is the only one whose acquisition of information from the stimulus display was not accompanied by eye fixations. Because this participant’s eyetracking data is not a reliable indicator of their use of stimulus information, their data is omitted from the subsequent analyses.

The parameter values for the sigmoid were averaged over the remaining 17 Type II participants, and are presented in Table 1. Table 1 confirms that many of the performance characteristics exhibited by the individual in Figure 8 also hold at the group level. First, like the Type I group, the average Type II participant generally
fixated all (2.71) stimulus dimensions early in learning, but by the end of learning was fixating only those dimensions needed to solve the learning problem (2.71 – 0.67 = 2.04 dimensions). Second, comparison of the average $m$ parameter in the Type I and II conditions confirms the more gradual learning that occurred in the latter condition according to all four measures: number of dimensions fixated ($m = 0.77$ vs. 1.43 in the Type I condition, or 7.6 vs. 4.1 trials), proportion fixation ($m = 0.38$ vs. 1.34, or 11.6 vs. 3.3 trials), relative priority ($m = 0.34$ vs. 2.11, or 12.9 vs. 2.1 trials), and errors ($m = 0.51$ vs. 2.28, or 8.6 vs. 1.9 trials). T-tests confirmed that the (logarithm) of the $m$ parameter in the two conditions were statistically different from one another ($ps < .05$) for all dependent measures except for the number of dimensions fixated ($p > .15$). Finally, as expected given the greater number of blocks required for Type II learning, the inflection points of the sigmoids (the $b$ parameter) occurred considerably later for the average Type II versus Type I participant (around trial 60 versus 19, all $ps < .0001$).

Although the average Type II parameter values provide an overall summary of performance in that condition, the backward learning curves presented in Figure 9 indicate that there was a substantial variability over participants. For example, whereas most participants exhibited a reduction in the number of dimensions fixated from three to two (Fig. 8A), two participants consistently fixated two dimensions, and two others three dimensions. And, Figure 9B indicates that participants exhibited a variety of rates in error reduction. Thus, as we did in the Type I condition, we present a number of distinct clusters of performance presumably representing different learning strategies.

The two learning profiles that accounted for most Type II participants are presented in Figure 10. The profile in Figure 10A represents the modal performance in the Type II condition, accounting for 9 participants (including the one in Figure 8). Because the reduction in the number of errors in this group occurred in an average of ~26 trials ($m = 0.17$), participants in this cluster exhibit gradual learning. Moreover, this reduction in errors is synchronized with a shift in eye movements to the two relevant
dimensions: These participants begin to spend a greater proportion of time fixating the two relevant dimensions a few trials after the reduction in the number of errors starts, and are fixating only the two relevant dimensions by the time errors cease. Thus, the performance of these individuals accords with the predictions of ALCOVE in which error reduction and attention shifts are gradual and co-occur. Just as was the case for the gradual learners in the Type I condition, note that changes in eye fixations occur after changes in error rates: By the time eye fixations begin to show a preference for the two relevant dimensions (around trial 56), the average error rate has dropped almost to 0.22.

The second major performance profile is presented in Figure 10B. In contrast to modal group in Figure 10A who exhibited gradual learning, this group of four individuals can be characterized as all-or-none learners, because, according to the sigmoid fits to their error data, their error rate dropped from 50% to 0% in a single trial. Note however, that whereas for the four individuals in Figure 11B errors are eliminated by trial 41, the shift in eye movements to the two relevant dimensions is not complete for another 16 trials. In this regard, these individuals look like typical Type I learners in that the shift in eye movements occurs only after the learning problem is already solved.

Figure 11 presents exceptions to the dominant profiles shown in Figure 10. Figure 11A depicts two participants who displayed gradual learning like those in the modal group, but without any shift in eye movements to the two relevant dimensions. Just as was the case for the single Type I participant who consistently examined all three dimensions, we speculate that these two individuals systematically memorized each of the eight stimuli. Consistent with this interpretation is the especially slow decrease in errors exhibited by these participants (average $m = 0.05$, or 90% change in 88 trials) as compared to those in Figure 10 (average $m = 0.17$, or 26 trials), as well as the greater average number of blocks taken to reach the learning criterion by the former group.
Eyetracking and Selective Attention

(17.5 vs. 13.7).

Finally, the two individuals in Figure 11C are those we have labeled two-dimensional rule testers in Figure 9. These learners only examined two dimensions during the experimental session, and these turned out to be the two dimensions needed to learn the exclusive-or rule. Just as we did for the two one-dimensional rule testers in the Type I condition, we speculate that these individuals learned via hypothesis testing in which errors ceased when the correct exclusive-or rule was discovered.

In summary, most Type II learners exhibited the gradual reduction in errors characteristic of associationist theories of learning, although a sizable minority (6 of 17) exhibited all-or-none learning. In addition, although four participants never showed any shift to the relevant dimensions (because two consistently fixated two dimensions and two others always fixated three), most Type II participants exhibited a shift in eye movements to the two relevant dimensions that was closely synchronized with (albeit later than) the reduction in errors. Given its theoretical importance, we summarize the close relationship between shifts in eye movements and error reduction in both the Type I and Type II conditions in Figure 12. Figure 12 plots the fitted $b$ parameters (the inflection point of the sigmoid curve) for each participant’s number of dimensions fixated and error data for those participants that exhibited a shift in eye movements. As the figure illustrates, shifts in eye movements were highly correlated with error reduction ($r = .96$). In addition, the fact that most data points fall above the diagonal emphasizes that shifts in eye movements tended to follow, rather than precede, the elimination of errors.

The final issue we consider in our presentation of the Type II results concerns the ordering of the fixations to the two relevant dimensions. Although current models of categorization do not generally make predictions regarding the order in which information from stimulus dimensions is acquired, we asked whether the Type II participants exhibited a consistent scan path, that is, a consistent order in which stimulus
dimensions were fixated. To answer this question, we first computed the average relative priority for each of the three stimulus dimensions during the last four error-free blocks for each Type II participant. On the basis of these averages, the three dimensions were then designated as either high, medium, or low priority, indicating whether they tended to be fixated earlier or later in the trial. Finally, we divided each trial into 50ms buckets, and in each bucket tabulated whether the high, medium, and low priority dimensions were fixated. The result of averaging these tabulations over all Type II trials is presented in Figure 13B (for purposes of comparison the corresponding results from Type I are presented in Figure 13A). If, at the end of Type II learning, the two relevant dimensions were being examined in a random order, we would expect that the two histograms for high and medium priority dimensions to be indistinguishable. In contrast, Figure 13B indicates a clear separation between these two histograms. This result suggests that most Type II participants tended to utilize a consistent scan path: one dimension tended to be fixated in the early parts of the trial, whereas the other was examined in the latter parts.

**Types IV and VI Results**

In this final section we present the results from the two category structures which remain, Types IV and VI. Unlike Types I and II, these structures require learners to attend to all three stimulus dimensions to successfully discriminate the two categories. Type IV can be construed either as a single-dimension-plus-exception structure, or a linearly-separable problem in which all three dimensions have equal weight. For example, Type IV can be solved with a $2$-out-of-$3$ rule in which an exemplar is considered a category member if two out of three dimension values favor that category. The Type VI structure, in contrast, essentially requires learners to memorize the category membership of each exemplar.

We used our sigmoid fitting procedure to analyze the number of dimensions fixated for those Type IV and VI participants who solved the category learning
problems (15 for Type IV and 10 for Type VI). Backward learning curves for these Type IV and VI learners are presented in Figures 14A and 15A, respectively. These figures illustrate that participants began the experimental session by examining between 2.5 and 3 stimulus dimensions, just like those in the Type I and II conditions. As expected given these category structures, all learners were fixating all three dimensions by the end of learning. The average parameter values of the sigmoid fits to the number of dimensions fixated (Table 1) confirms that the average Type IV and VI learner fixated all stimulus dimensions early in learning (2.44 and 2.71, respectively), and were fixating all three dimensions by the end of learning (2.85 and 2.91). (Because all three dimensions are equally relevant in the Type IV and VI category structures, we do not define the proportion fixation and relative priority measures for the relevant dimensions in these conditions.)

Figures 14B and 15B present backward learning curves for the error sigmoids in the Type IV and VI conditions, respectively. These figures indicate how most participants in these conditions exhibited gradual learning. On the one hand, the average values of the $m$ parameter for the error fits (Table 1) indicate that learning occurred more abruptly than implied by the group level data presented in Figure 3. The average $m$ of 0.11 in the Type IV condition corresponds to a 90% reduction in errors occurring in 40 trials; an average $m$ of 0.23 in the Type VI condition corresponds to a 90% reduction occurring in 19 trials. On the other hand, Figures 14B and 15B indicate that learning occurred more gradually than in either the Type I (1.9 trials) or Type II (8.6 trials) conditions. In fact, the $m$ parameter in the Type IV condition differed significantly from that in Type I ($p < .0001$) and Type II ($p < .05$) conditions; $m$ in the Type VI condition differed significantly from $m$ in the Type I condition ($p < .0001$) although not the Type II condition ($p > .20$).

Once again, we examined individual Type IV and VI participants to identify distinct clusters of performance. In fact, the pattern of performance represented by
average parameters values in Table 1 were manifested by the large majority of learners in both the Type IV (13 of 15) and the Type VI (8 of 10) conditions. The performance of these modal groups are presented in Figures 16A and 17A, respectively, which illustrates the gradual reduction in errors manifested in both conditions. However, one notable feature of these results is that although learning was faster in the Type IV condition overall, the reduction in number of errors, once it starts, occurs more abruptly in the Type VI condition. On the one hand, an associationist account of learning like ALCOVE explains the faster learning of the Type IV category structure in terms of the larger within-category similarity (and smaller between-category similarity) found in that structure as compared to Type VI. However, this account does not explain why the rate of learning should be slower in the Type IV condition (even though it starts earlier). Instead, we speculate that some Type IV participants first discovered a single-dimension rule and then memorized the exceptions to this rule. This strategy, which accords with the predictions of RULEX, yields an initial reduction in error rate to 0.25 as the single-dimension rule produces the correct classification on 6 out of the 8 exemplars, and then a slow elimination of all errors as the two exceptions are memorized.

Although the group level performance profiles of the Type IV and VI conditions reflect gradual learning, in fact we found two all-or-none learners in each condition. The performance profiles of these two clusters are presented in Figures 16B and 17B. We speculate that the two all-or-none learners in the Type IV condition (Fig. 16B) first tested single-dimension rules on each of the three dimensions, and then, after discovering that each of these rules had some predictive validity, formed a 2-out-of-3 decision rule to solve the problem. Given the absence of any rule-like solution for the Type VI category structure, the presence of all-or-none learners in that condition (Fig. 17B) is quite surprising—especially the one individual whose last error occurred on trial 19! Conceivably, these individuals may have first learned to successfully identify each of the eight stimuli (that is, they first eliminated the problem of between-category
similarity that makes the Type VI structure so difficult), and only then learned to associate these stimuli and their correct category label. Indeed, one participant reported encoding the stimulus with the features $, !, and x as the word “six”, a mnemonic strategy likely to have accelerated the association of the stimulus with its category label (Gibson, 1940; Bower & Hilgard, 1981).

Finally, as we did for the Type II condition, we also consider the question of whether Type IV and VI participants exhibited a consistent scan path. As before, for each participant the three stimulus dimensions were initially classified as being of either high, medium, or low priority, and then we tabulated whether each of these dimensions were fixated in each 50ms bucket. These tabulations averaged over all Type IV and VI trials are presented in Figures 18A and 18B, respectively. If, at the end of learning, the three stimulus dimensions were being examined in a random order, we would expect that the three histograms to exhibit a high degree of overlap. In contrast, both Figures 18A and 18B indicate a clear separation between the histograms. This result suggests that most Type IV and VI participants utilized a consistent scan path as stimulus dimensions were examined in a consistent order.

Fixations Early in Learning

One of the most striking results from current study is the fact that learners tend to fixate all stimulus dimensions early in learning. Of the 62 participants in this experiment who reached the learning criterion, the average value of the initial parameter of the sigmoid fits to the number of dimensions fixated was 2.54, reflecting that most participants fixated most stimulus dimensions early in learning. This result goes against the predictions derived from a rule- or hypothesis-testing account of category learning in which learners begin by searching for simple one-dimension rules. However, because this result was based on our sigmoid fits to each participant’s entire eye fixation data, we sought more direct confirmation that learners fixate all stimulus dimensions early in learning. In Figure 19 we present the average number of
dimensions fixated in the first five trials of the experiment for all 72 participants. Note that during these initial trials relatively little learning has occurred, and thus Figure 19 reflects participants’ eye movements before they have acquired substantial knowledge of the correct category representation. The results are clear-cut. Out of 72 participants, only 3 consistently fixated one dimension in the first five trials, and only 6 fixated an average of less than 1.5. Instead, during these trials 85% of participants fixated two or more dimensions, and the modal number of dimensions fixated was three.

Discussion
Since Shepard et al.’s (1961) seminal study a core assumption of categorization theory has been that category learning involves learning to attend to those stimulus dimensions useful for category discrimination. However, evidence for this claim has consisted of demonstrations that dimensions vary in their influence on explicit categorization (and similarity) judgments, not on the operation of selective attention per se. Our findings provide strong support for the claim that categorizers learn to allocate their attention to optimize classification performance. Only one of 18 Type I participants and just two of 18 Type II participants failed to restrict attention to only those relevant dimensions by the end of learning. To our knowledge, the current results provide the first direct evidence for the operation of selective attention in category learning.

An important accomplishment in categorization theory has been the specification of computational models that formalize the mechanism by which selective attention changes as a result of classification experience. According to one member of the class of associationist learning models—the ALCOVE model—attention should first be allocated to all stimulus dimensions and then gradually shift to diagnostic dimensions as learning proceeds. In contrast, according to the hypothesis-testing, or rule-based, model known as RULEX, attention should first be allocated to single stimulus dimensions as one-dimension categorization rules are tested, and then to multiple
dimensions as more complex rules are considered. In fact, we found that early in
learning participants tended to fixate all stimulus dimensions, a result which ostensibly
provides support for ALCOVE. At the same time, however, we also found that
numerous participants exhibited what we referred to as all-or-none learning, the
sudden elimination in errors usually taken to be characteristic of rule-based learning.

In the two sections which follow we review the evidence in favor of both
associationist and rule-based learning. We then propose a framework for category
learning which accounts for both of these learning strategies. Finally, we conclude with
a discussion of our third notable eyetracking result: The fact that changes in eye
movements tended to follow rather than precede changes in errors.

Evidence for Associationist Accounts of Category Learning

Overall, we found that considerable numbers of learners exhibited the type of
gradual learning typical of associationist learning models like ALCOVE. There are two
sources of evidence for this conclusion. The first involved the sigmoid functions we fit
to participants’ error data which showed a gradual drop in error rates from 50% to 0%
occurring over several trials. Of those participants assigned the Type II category
structure, the majority (12 of 18) exhibited a gradual decrease in errors which occurred
over about three blocks. In addition, despite the simpler single-dimension rule required,
a substantial minority (5 of 18) participants also exhibited a gradual decrease in errors
in the Type I condition occurring over about two blocks. Finally, we found that gradual
learning was exhibited by virtually all participants in the Type IV and VI conditions.

By itself, of course, a gradual decrease in errors cannot be considered uniquely
diagnostic of associationist learning, because rule-based classification processes are also
likely to include numerous sources of stochastic variability that can also produce a
gradual decrease in errors. For example, even after the correct rule has been discovered,
the chance of successfully retrieving that rule from memory on any given trial may be
less than certain. However, the number of such retrieval failures will decrease as the
rule becomes more strongly represented in memory (Anderson, 1983). Second, application of a correct rule also requires correctly identifying the stimulus dimension values; misidentification of a feature (e.g., due to perceptual noise) will result in misclassification (Smith, Patalano, & Jonides, 1998). Third, noise at the decision stage may arise when classifiers attempt to probability match; however, responses will tend to become more deterministic as classification experience increases (Ashby & Gott, 1988; McKinley & Nosofsky, 1995; Nosofsky & Zaki, 2002). Finally, simple motoric response noise (i.e., pushing the wrong button) may combine with all these factors to produce a pattern of gradually decreasing errors.

Nevertheless, our claim of gradual learning is also supported by a second source of evidence, the eyetracking data. A model like ALCOVE not only predicts a gradual reduction in errors, but a also a gradual shift in attention to the relevant dimensions. In fact, just such a shift (as measured by eye movements) was also observed for those Type I and II participants who exhibited gradual learning. For example, in the Type I condition, the shift in eye movements to the single relevant dimension occurred over four or five trials for gradual learners. Similarly, for gradual learners in the Type II condition, the shift to the two relevant dimensions occurred over the course of ten or more trials. Taken together, the error and eye tracking data provide strong support for the presence of ALCOVE-like learning processes for a large number of our participants. It is important to note, however, that a critical feature of these data is the fact that changes in eye movements tended to follow rather than precede the reduction in errors, a result which we discuss at length below.

**Evidence for Rule-Based Accounts of Category Learning**

We also found considerable evidence for the use of hypothesis-testing or rule-based models. Although gradual learning is not by itself diagnostic of associationistic learning (as just discussed), “all or none” learning in which errors are suddenly eliminated is uniquely diagnostic of the discovery of a rule that discriminates category
members (Bower & Trabasso, 1963). In fact, we found that the error sigmoids of the majority (13 of 18) of participants assigned the Type I category structure exhibited a drop in errors from 50% to 0% in only one or two trials. Moreover, despite the more complex exclusive-or rule required, a substantial minority (6 of 18) participants exhibited all-or-none learning in the Type II condition. Finally, we found that all-or-none learning was exhibited even in the more complicated Type IV and VI conditions by a small number of learners (2 in each condition).

Our eye movement predictions for rule-based learners were derived from the RULEX model which predicts that learners begin with simple hypotheses in the form of single-dimension rules, and only consider multi-dimensional rules when simple one-dimensional rules fail to yield a solution to the learning problem. The straightforward prediction we derived is that learners will fixate a single stimulus dimension early in learning, and will only fixate multiple dimensions later in learning for those category structures that cannot be solved with a one-dimensional rule. In fact, perhaps the most striking result from the current study was the almost complete absence of evidence that category learners fixate single stimulus dimensions early in learning. Of the 72 undergraduates who participated in the current study, we found only three that fixated approximately one dimension in the first five trials of the experimental session. Instead, during these trials 85% of participants fixated two or more dimensions, and the modal number of dimensions fixated was three.

Superficially at least, this eye movement data calls into question RULEX’s claim that people first test single-dimension rules when learning categories. However, it is important to recognize that RULEX was not specifically designed to account for eye fixation data, and so we must be careful to consider possible reasons for the failure of our (perhaps overly simplistic) expectations regarding the relationship between eye fixations and rule testing. In particular, it may be that all-or-none learners were in fact testing rules in the manner prescribed by RULEX, but that eye fixations did not reflect
this fact because they were also being influenced by cognitive processes not involved in rule testing \textit{per se}. There are a number of such processes that may have been partly responsible for the observed eye movements.

One possibility is that participants fixated all stimulus dimensions at the beginning of the experimental session in order to learn the structure of the stimulus space. Although the stimulus dimensions were described to participants before the experiment started, this information may not have been fully encoded, and thus some of their initial efforts may have been devoted to more fully learn the six feature values and their locations (e.g., “$” and “¢” were the two values that appeared at the top of the screen, that “x” and “o” appeared at the bottom left, etc.). This encoding would assist in the generation and testing of future candidate rules, and/or the memorization of individual exemplars.

Another possibility is that participants construed their learning task to be broader than just classification. For example, a number of investigators have argued that the central function of categories—the reason people learn categories in the first place—is to allow them to infer the presence of features that cannot be directly observed (Anderson, 1991; Corter & Gluck, 1992; Markman & Ross, 2003). If this is correct, then during category learning a learner’s goals may not be just to determine features’ cue validity (the probability of the category given the feature), but also their category validity (the probability of a feature given the category). On this account, category learners fixate all dimensions of a stimulus in order to learn which features are characteristic of each category and thus promote the accuracy of feature inferences that may be required in the future.

Learners may also be driven by the general goal of remembering the individual instances to which they are exposed. Recognizing individual instances is likely to have adaptive advantages beyond classification performance (Palmeri & Nosofsky, 1995); more generally, such memorization may arise from a general cognitive strategy of
avoiding information loss (Medin & Florian, 1992). Finally, for completeness we note that under some conditions (ones unlikely to have obtained in the current experiment) stimulus dimensions may attract attention because learners find them intrinsically interesting, or because of preattentive processes that obligate the processing of certain aspects of stimuli (Lamberts, 1995; 1998).

However, all of these possibilities fail to account for the high correlation we found between changes in eye fixations and the elimination of classification errors. If fixations to all stimulus dimensions merely reflected participants’ attempt to encode the stimulus dimensions, then such fixations should have disappeared in the relatively small number of trials needed for such encoding to be complete. And if those fixations reflected the learning of category validities (or the memorization of individual exemplars or because the stimuli were intrinsically interesting), those fixations should have continued well after classification errors ceased. Instead, we found that fixations to dimensions irrelevant to correct classification were eliminated at the same time that errors ceased (or a few trials later). It therefore follows that those fixations must have arisen as a result of cognitive processes that were directly related to the goal of category learning.

For this reason, we conclude that models like RULEX that assume that learners start off by (only) testing single-dimension rules cannot be considered complete accounts of our participants’ learning strategies. At the same time, however, the sudden elimination of errors in our Type I condition indicates that category learners are able to easily discover single-dimension rules when they exist. The question then is: Why do learners examine all stimulus dimensions at the same time they are able to extract single-dimension rules so readily?

Implications for Multi-Strategy Approaches to Category Learning

We believe that the answer to this question lies in the recognition that category learners are often pursuing more than one learning strategy. That is, although many
participants may be able to discover single-dimension rules when they exist (as our all-or-none learning data suggests), it is likely that they are able to apply other learning approaches at the same time. This account explains the fixations to all stimulus dimensions early in learning assuming that at least one of these other approaches requires access to information on all stimulus dimensions. We can envision a number of potential strategy combinations that would require fixations to all stimulus dimensions.

Rule testing plus exemplar memorization. One possibility is that although learners began by explicitly searching for single-dimension rules, they may have recognized that a perfect single-dimension rule may not be found, and that memorizing individual exemplars may be necessary as a backup strategy. Examining all stimulus dimensions early in learning would provide the learner a head start on this memorization process. Examining all stimulus dimensions would also provide a head start on the process of memorizing exceptions to an imperfect yet predictive single-dimension rule. Consistent with this proposal is evidence demonstrating the influence of specific exemplars on classification even when a perfect classification rule is available. For example, Allen and Brooks (1991) used a novel procedure in which they provided participants with the correct rule (a 2-out-of-3 rule) to distinguish members of two categories of imaginary animals. Nonetheless, they found that performance on a transfer classification test was influenced by features unrelated to the rule, that is, by overall similarity of the test items to the training items (also see Nosofsky, Clark, & Shin, 1989; Smith et al., 1998). Erickson and Kruschke (1998, Experiment 1) found that after learning rule-plus-exception category structures, categorizers were influenced by the similarity of transfer stimuli to the exceptions, as standard exemplar models would predict. Finally, Nosofsky (1991) found that the frequency of training stimuli influenced subsequent classification performance even for one-dimensional category structures, a result explained naturally in terms of the memory traces of the individual training exemplars (also see Erickson & Kruschke, 1998, Experiment 2).
Exemplar memorization plus spontaneous rule noticing. The account just described assumes that people’s initial explicit learning strategy is to look for one-dimensional rules, and to use exemplar memorization as a backup strategy. However, it is also possible that some learners started off by trying to memorize exemplars, but that all-or-none learning arose in the Type I condition when learners “noticed” (somehow) that one dimension covaried consistently with the category label. This noticing might have been based on comparing the current exemplar with the previous exemplar stored in working memory (Anderson, Kline, & Beasley, 1979). Or, the comparison may have been between the current exemplar and one stored in long-term memory that the learner was reminded of (Ross, Perkins, & Tenpenny 1990).

Rule testing (or noticing) plus meaningful interfeature relations. Finally, learners may have examined all stimulus dimensions because they are biased to expect that features are meaningfully related on the basis of their prior knowledge. Indeed, there is considerable evidence that learners readily notice and make use of inter-feature relations when they are available. Research has shown that supervised category learning is accelerated when the categories’ features are mutually meaningful and coherent in light of existing knowledge (Kaplan & Murphy, 2000; Murphy & Allopenna, 1994; Rehder & Ross, 2001; Rehder & Murphy, 2003). Likewise, people’s unsupervised sorting of items into categories is strongly determined by their prior knowledge about the items’ features (Ahn & Medin, 1992; Kaplan & Murphy, 1999; Medin, Wattenmaker, & Hampson, 1987; Spalding & Murphy, 1996). Of course, the fact that cross-dimension feature relations influences learning entails that learners were attending to multiple dimensions in order to have noticed those relations.

Taken together, these considerations have led us to conceive of our participants as opportunistic learners who can make simultaneous use of multiple learning strategies and who select a strategy when it yields a solution to the learning problem. This account predicts the current pattern of eye movements early in learning, because eye
Eyetracking and Selective Attention

Fixations to all stimulus dimensions would be required, for example, to (a) start the process of memorizing exemplars, (b) compare the current exemplar with previous one stored in memory in order to notice commonalities, and/or (c) to search for meaningful relations among features. Stated more generally, we suggest that learners examine all stimulus features because it maximizes the number of potential learning strategies involved.

This view of learners as opportunistically pursuing multiple learning strategies is consistent with the current trend toward considering category learning as involving more than one learning module (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998; Kruschke, 2001). For example, Erickson and Kruschke’s (2001) ATRIUM model instantiates a mixture of experts architecture in which multiple modules each apply a different strategy to solve the current learning problem. In its current form, ATRIUM includes both a rule module that works to detect a single-dimension rule that solves the categorization problem, and an exemplar module (equivalent to ALCOVE) that simultaneously associates stored exemplars with their correct category label. A gating mechanism then determines how the responses from these two modules should be combined to produce an overt response for a given exemplar, with preference eventually being given to the module that is producing the fewest classification errors. Similarly, Ashby et al.’s (1998) multiple-systems COVIS model contains both a “verbal” rule-learning module that can operate on single dimensions and a “procedural learning” module that discovers an optimal decision bound using all available dimensions. These modules then compete for attention such that the more successful module is increasingly used in category decisions. In either of these frameworks, eye fixations would be made to all dimensions early in learning because one of the learning modules requires access to all the information in the stimulus.

Our discovery of distinct clusters of performance provides additional evidence
for a multi-strategy view of category learning. For example, a rule module dedicated to
discovering single dimension rules will usually be able to solve a Type I category
learning problem faster than an ALCOVE-like exemplar module engaged associative
learning. However, because of the presumably stochastic nature of the processes
involved, the exemplar module will occasionally win this race. This conjecture is
supported by our finding in the Type I condition that although the majority of
participants apparently solved the problem by discovering the correct rule, a substantial
minority (4 of 18) exhibited the gradual learning characteristic of associative learning.
Similarly, the finding in the Type II condition that learning was gradual for 9
participants but all-or-none for 4 suggests that an ALCOVE-like exemplar module will
usually be the first to arrive at the solution to an exclusive-or problem, but that a
(perhaps RULEX-like) module responsible for discovering rules that include
conjunctions, disjunctions, and combinations of the two will occasionally win the race
Considerably more work is required to clarify the relationship between multiple
learning strategies, and how those strategies interact with one another during learning.
For example, one outstanding question is how attention gets redirected toward one
learning strategy and away from others. According to ATRIUM, this change occurs
gradually as error feedback influences the gating mechanism that combines the outputs
of multiple experts in a way that minimizes error. However, this gradual shift is
incompatible with our finding of all-or-none learning (in the Type I condition for
example) that suggests that explicit classification responses suddenly come under
control of a single-dimensional rule. More recently, Kruschke (2001; Kruschke &
Johansen, 1999) has developed a series of models that incorporate rapid shifts of
attention (either between different single-dimension rules, or between rules and an
exemplar module) that have the potential of accounting for the patterns of all-or-none
learning we observed.
A final important question concerns the role of explicit strategy choice on the
part of the learner. We have suggested that learners examine all stimulus dimensions in order to involve as many learning modules in the learning process as possible. But we also found a small number of participants who did not examine all stimulus dimensions early in learning, a result we attributed to those learners adopting an explicit strategy of searching for rules. In addition, we suggested that the small number of Type I and II participants who never restricted their eye fixations to the relevant dimensions adopted an explicit strategy of just memorizing each exemplar’s category membership. That is, although we believe that most category learners start with an open mind regarding the form of the solution to the learning problem, some will begin with a commitment to a specific strategy (e.g., rule discovery, memorization, or one of the other learning strategies we have noted). In such cases eye fixations will reflect the informational requirements of that strategy alone.

**Attention as a Cognitive Resource in Category Learning**

The final aspect of our eyetracking results that we discuss is the fact that, although eye movements were generally well-synchronized with error reduction, the changes in fixations tended to follow rather than precede changes in errors. This result is a puzzle because, from the perspective of the multi-systems theories of learning we have been considering, the most natural prediction would be that the shift of “attention” toward the one expert which is producing correct classification responses should be accompanied by a shift in eye movements to only the information relevant to that expert. To take ATRIUM as an example, when presented with a one-dimensional categorization problem the rule module would discover the correct rule in fairly short order, and would quickly dominate the categorizer’s explicit classification responses. When this occurred, the withdrawal of attention from the ALCOVE module would presumably release the learner from the need to fixate dimensions irrelevant to the one-dimensional rule. But we found instead that although the modal all-or-none Type I learner discovered the correct one-dimension rule in 12 trials, he or she continued to
examine all stimulus dimensions until trial 16. The question then is: Why do learners continue to fixate stimulus dimensions which have become irrelevant to their overt classification response?

We believe that the answer to this question is that learners will continue to pursue multiple learning strategies until there is evidence that one strategy has solved (or is about to solve) the learning problem. For example, we believe that the change in eye movements of the modal all-or-none Type I learner was delayed because four error-free trials were required to conclude that the correct rule had been discovered. Only at that point was the learner willing to abandon other learning strategies and focus exclusively on the one relevant dimension.

A similar account can be given of the results in the Type II condition. The modal Type II gradual learner began to show a reduction in their error rates about trial 40. However, shifts in eye fixations to the two relevant dimensions didn’t begin until about trial 56, when the modal participant’s chance of making an error had fallen to about 0.20. Again, we believe that restriction of eye movements to the two relevant dimensions occurred only once a reduction in the rate of errors signaled that an ALCOVE-like learning module was heading toward a solution of the learning problem. At that point other learning strategies were abandoned and eye fixations began to reflect the informational needs of ALCOVE alone.

We believe this account provides a new perspective on the role of “attention” in category learning. Traditionally, categorization theories have used the concept of attention to refer to the relative influence that individual stimulus dimensions have on explicit categorization decisions. More recently, Lamberts (1995; 1998) has distinguished this notion of decision weights from that of perceptual saliency, demonstrating that more salient dimensions will exert greater influence on classification decisions at shorter response deadlines (also see Maddox 2002; 2003; Maddox, Ashby, Waldron, 2002). On our account, attention has yet a third meaning, which is the allocation of cognitive
resources to a particular learning strategy. Moreover, allocation of attention (i.e.,
cognitive resources) to a learning strategy in turn leads to the acquisition of the
stimulus information needed for that strategy to function, whereas the withdrawal of
cognitive resources from a strategy in turn removes the need for that information. On
this account, eye fixations are diagnostic of whether a particular learning strategy is still
active, that is, is still being allocated cognitive resources.

Conclusion

To our knowledge, the current experiment is the first to use eyetracking
technology to examine the question of selective attention in category learning. There
were three primary findings. The first is that category learners indeed learn to optimally
allocate attention as part of learning to discriminate categories. This finding
corroborates the assumptions of virtually all modern theories of category learning. The
second finding is that learners tend to fixate all stimulus dimensions early in learning.
This occurs despite the fact that they are also able to easily discover one-dimensional
categorization rules during this same period. The third finding is that changes in
selective attention (i.e., eye fixations) to only relevant dimensions tend to occur after
errors have been greatly reduced (or completely eliminated). We have interpreted this
set of findings as consistent with multi-systems theories of learning. In addition, we
proposed that (a) participants will initially maximize information input in order to
maximize the number of learning modules involved, and that (b) participants will
optimize attention to just the relevant information only after one module has largely
solved the learning problem.

We believe that the results reported here have established the usefulness of
eyetracking for testing existing categorization theory and forming interesting new
hypotheses regarding people’s learning strategies. Rather than just estimating
“attention weights” as free parameters in classifications models, eyetracking measures
can help decompose the heretofore monolithic construct of attention into better-defined
components such as physical salience, decision weights, and attention-as-a-cognitive resource. In addition, by providing a sophisticated online processing measure, eyetracking data may promote the development of models that specify the cognitive processes that produce classification decisions (Nosofsky et al. 1997; Lamberts 1998). We expect that eyetracking will provide a rich source of empirical data that will help discriminate among existing models and advance the construction of new theory in this area.
References


Author Note

Bob Rehder, Department of Psychology, New York University. Aaron B. Hoffman, Department of Psychology, New York University.

Be thank John K. Kruschke, Bradley C. Love, Gregory L. Murphy, Robert M. Nosofsky, and Jonathon Nelson for their comments on a previous version of this manuscript. Correspondence concerning this article should be addressed to Bob Rehder, Department of Psychology, 6 Washington Place, New York, NY 10003 (email: bob.rehder@nyu.edu).
Footnotes

1 In these fits, parameter $m$ was constrained to be $\geq 0.026$, and $\leq 5.89$. When $m = 0.026$, 95% of the change in the sigmoid from its initial to final asymptote occurs in 224 trials (the maximum length of the experiment); when $m = 5.89$, 95% of the change occurs in 1 trial. For the error fit, we set $\text{initial} = 0.50$ and $\text{diff} = -0.50$ reflecting initial guessing and eventual learning.

2 For purposes of computing the averages in Table 1, those fits for which the intercept model was the best fitting model were assigned $\text{initial} = \text{the average}$, $\text{diff} = 0$, and $m = 0.026$. No value for the $b$ parameter was assigned for these fits, which therefore have no influence on the average value of the $b$ parameters in Table 1. Note that because of the $m$ parameter’s nonnormal distribution, its average was calculated by taking the natural logarithm of the individual $m$s, averaging $\ln(m)$ over participants, and then exponentiating this average.
Table 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th># of Dimensions</th>
<th>Proportion Time</th>
<th>Relative Priority</th>
<th>Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type I</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>2.61</td>
<td>0.34</td>
<td>0.39</td>
<td>0.50</td>
</tr>
<tr>
<td>diff</td>
<td>−1.45</td>
<td>0.60</td>
<td>0.53</td>
<td>−0.50</td>
</tr>
<tr>
<td>m</td>
<td>1.43</td>
<td>1.34</td>
<td>2.11</td>
<td>2.28</td>
</tr>
<tr>
<td>b</td>
<td>18.7</td>
<td>19.2</td>
<td>19.2</td>
<td>14.5</td>
</tr>
<tr>
<td><strong>Type II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>2.71</td>
<td>0.65</td>
<td>0.72</td>
<td>0.50</td>
</tr>
<tr>
<td>diff</td>
<td>−0.67</td>
<td>0.30</td>
<td>0.21</td>
<td>−0.50</td>
</tr>
<tr>
<td>m</td>
<td>0.77</td>
<td>0.38</td>
<td>0.34</td>
<td>0.51</td>
</tr>
<tr>
<td>b</td>
<td>60.1</td>
<td>56.0</td>
<td>61.0</td>
<td>50.0</td>
</tr>
<tr>
<td><strong>Type IV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>2.31</td>
<td></td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>diff</td>
<td>0.58</td>
<td></td>
<td>−0.50</td>
<td></td>
</tr>
<tr>
<td>m</td>
<td>0.11</td>
<td></td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>24.4</td>
<td></td>
<td>59.5</td>
<td></td>
</tr>
<tr>
<td><strong>Type VI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>2.71</td>
<td></td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>diff</td>
<td>0.21</td>
<td></td>
<td>−0.50</td>
<td></td>
</tr>
<tr>
<td>m</td>
<td>0.13</td>
<td></td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>14.7</td>
<td></td>
<td>85.8</td>
<td></td>
</tr>
</tbody>
</table>
Figure Captions

Figure 1. A. Category structures Type I, II, IV, and VI from Shepard et al. (1961), Experiment II. B. The Type I stimuli used in Shepard et al. (1961), Experiment I.

Figure 2. Example of stimulus presentation.

Figure 3. The number of dimensions fixated as a function of training block and problem type.

Figure 4. Performance on the 56 trials of one Type I learner. A. Number of dimensions fixated. B. Proportion RT. C. Relative priority. D. Errors.

Figure 5. Backward learning curves for the 18 Type I participants. A. Number of dimensions fixated. B. Errors. Performance of the average Type I participant is superimposed on the individual curves.

Figure 6. Two clusters of Type I learners.

Figure 7. Three individual Type I learners.

Figure 8. Performance on the 80 trials of one Type II learner. A. Number of dimensions fixated. B. Proportion RT. C. Relative priority. D. Errors.

Figure 9. Backward learning curves for the 18 Type II participants. A. Number of dimensions fixated. B. Errors. Performance of the average Type II participant is superimposed on the individual curves.

Figure 10. Two clusters of Type II learners.

Figure 11. Two clusters of Type II learners.

Figure 12. Relationship between changes in eye fixations and changes in error rates.

Figure 13. Fixations as a function of time for the highest, medium, and lowest priority dimensions. A. Type I condition. B. Type II condition.

Figure 14. Backward learning curves for the 15 Type IV participants who reached the learning criterion. A. Number of dimensions fixated. B. Errors. Performance of the
average Type IV participant is superimposed on the individual curves.

**Figure 15.** Backward learning curves for the 15 Type VI participants who reached the learning criterion. A. Number of dimensions fixated. B. Errors. Performance of the average Type VI participant is superimposed on the individual curves.

**Figure 16.** Two clusters of Type IV learners.

**Figure 17.** Two clusters of Type VI learners.

**Figure 18.** Fixations as a function of time for the highest, medium, and lowest priority dimensions. A. Type IV condition. B. Type VI condition.

**Figure 19.** Number of dimensions fixated in the first five trials of the experiment for all 72 participants.
Figure 1

A

B
Figure 2
Figure 3
Figure 4
Figure 5

A

B

Trial
Figure 6

![Graph A](image1)

A triangle represents the number of dimensions, a diamond represents proportion time, a square represents relative priority, and a circle represents error. The y-axis shows proportion time from 0.00 to 1.00, and the x-axis shows trial numbers from 0 to 40.

![Graph B](image2)

Graph B shows similar data with a different n value, indicating a different group or condition.
Figure 7

A

B

C

Trial
Figure 8
Figure 9
Figure 10

A

B

Trial

n = 9

n = 4

Number of Dim
Proportion Time
Relative Priority
Error
Figure 11

A

B

Trial
Figure 12

![Graph showing error inflection points for Type I and Type II with corresponding X and Y axes labeled.]

Error Inflection Point (b)
Figure 13

A. Type I

B. Type II
Figure 14
Figure 16

A

B

Trial

n = 13

n = 2

Number of Dim

Error
Figure 17

A

B

\( n = 8 \)

\( n = 2 \)

Trial
Figure 18

A. Type IV

B. Type VI
Figure 19