

Running Head: Knowledge and Attention in Category Learning

How Prior Knowledge Affects Selective Attention During Category Learning:

An Eyetracking Study

ShinWoo Kim

Bob Rehder

Department of Psychology

New York University

New York, NY 10003

Send all correspondence to:

ShinWoo Kim

Dept. of Psychology

6 Washington Place

New York, New York 10003

Phone: (212) 998-3881

E-mail: shinwoo.kim@nyu.edu

Abstract

Research has shown that category learning is affected by (a) attention, which selects which aspects of stimuli are available for further processing, and (b) the existing semantic knowledge that learners bring to the task. However, little is known about how knowledge affects what is attended. Using eyetracking, we found that (a) knowledge indeed changes what features are attended, with knowledge-relevant features being fixated more often than irrelevant ones, (b) this effect was not due to an initial attentional bias toward relevant dimensions but rather emerged as a result of observing category members, and (c) this effect grew even after a learning criterion was reached, that is, despite the absence of error feedback. We argue that models of knowledge-based learning will remain incomplete until they include mechanisms that dynamically select prior knowledge in response to observed category members and which then directs attention to knowledge-relevant dimensions and away from irrelevant ones.

How Prior Knowledge Affects Selective Attention During Category Learning:
An Eyetracking Study

Because of the importance of categories for human cognition, the manner in which people learn categories has received intensive study. Among many procedures, supervised classification learning has been popular among researchers, and a number of basic facts have been established about the acquisition of categories.

One of these concerns the role of selective attention. Selective attention has played a prominent role in theories of category learning since the finding that learning difficulty correlates with the number of diagnostic dimensions needed for classification (Shepard, Hovland, & Jenkins, 1961). Currently, all major theories of categorization incorporate means of moderating which stimulus dimensions are attended. In both exemplar (Medin & Schaffer, 1978; Nosofsky, 1986) and prototype (Hampton, 1995; Nosofsky, 1992; Smith & Minda, 1998) models, selective attention is formalized in terms of different weights that stimulus dimensions have on classification. Rule-based models also assume that learners selectively attend to the dimensions referred to by the current hypothesis being tested (Smith, Patalano, & Jonides, 1998; also see Maddox, 2002; Maddox, Ashby, & Waldron, 2002; Maddox & Dodd, 2003). In addition, some models include mechanisms specifying how selective attention changes with learning. For example, Kruschke's (1992) ALCOVE model changes attention weights on dimensions as a function of error feedback. Nosofsky, Palmeri, and McKinley's (1994) rule-plus-exception (RULEX) model first performs hypothesis (rule) testing on single dimensions and then on multi-dimensional rules if needed (also see Kruschke & Johansen, 1999).

A second important finding concerns the role of prior knowledge in the acquisition of new categories. Whereas many studies have examined how categories are learned solely on the basis of observed category members, research has also shown that semantic knowledge that subjects bring to the task can have a profound effect on the ease with which learning occurs and what sort of category information is acquired (see Murphy, 2002, for a review). For example, in Murphy and Allopenna's (1994) seminal study, subjects were asked to discriminate two categories in which all the diagnostic features of each category were related to a particular theme. Features of one category included "drives on

glaciers," "made in Norway," and "heavily insulated" and those of another included "drives in jungles," "made in Africa," and "lightly insulated." Subjects learned to distinguish these categories far faster—presumably because the features could be subsumed under the themes “a sort of arctic vehicle” and “a sort of jungle vehicle,” respectively—as compared to categories whose features shared no theme (also see Rehder & Ross, 2001). In addition, when only a subset of diagnostic features were related to a theme, subjects showed better learning of those features as compared to ones that were unrelated to the theme (Heit & Bott, 2000; Kaplan & Murphy, 2000).

However, although selective attention and prior knowledge have each been shown to affect category learning, little is understood about how knowledge affects what is attended. This question is important because any theory of how knowledge influences learning is likely to be incomplete in the absence of any account of how it alters what category information is attended and thus processed. For example, computational models such as Baywatch (Heit & Bott, 2000) and KRES (Rehder & Murphy, 2003; Harris & Rehder, 2006) attempt to account for effects of knowledge by assuming that it moderates some of the same basic associative learning mechanisms that have successfully accounted for category learning in the absence of knowledge. However, neither model does so by postulating changes to what is attended—an important omission given the effects of selective attention reviewed earlier. However, this omission is readily understandable in light of the fact that virtually *nothing* is known about how knowledge alters what is attended: modeling attentional effects is impossible if there is no data to model. Thus, to further our understanding of prior knowledge effects in particular, and category learning in general, a first step is to establish some basic empirical facts regarding how attention is modulated by the prior knowledge that learners bring to the task.

The overall purpose of the current study, then, is to describe how knowledge affects attentional changes to dimensions during the course of learning thematic categories, that is, those whose dimensions are related to one another on the basis of a common theme (as in, e.g., Murphy & Allopenna, 1994). Unlike previous research that indirectly inferred dimensional attention by fitting computational models (e.g., Medin & Schaffer, 1978), in this study we used eyetracking as a more direct measure of attention. Using eyetracking, we test a number of alternative hypotheses regarding the effect of knowledge on

attention, as now described.

How Knowledge Might Affect Attention

Our study is guided by three open questions regarding how prior knowledge might affect attention during learning. The first question of course concerns whether in fact knowledge induces *any* change to what is attended. One possibility is that knowledge might affect learning not via attention but rather by changing how category information is processed and encoded. For example, knowledge might allow a category's theme-related features to be associated more strongly with its underlying representation. Indeed, a classic finding in the memory literature is that the effect of repetition or rehearsal on memory of nonsense items can be reversed depending on the “depth” or meaningfulness of the encoding process—how much the new information is related to what is already known (e.g., Bower, Clark, Lesgold, & Winzens, 1969; Craik & Lockhart, 1972; Craik & Tulving, 1975; Stein & Bransford, 1979). In addition, the presence of knowledge might allow classification to become an act of reasoning—one *infers* an item's category membership on the basis of a semantic understanding of the category (we provide examples below). As mentioned, models like Baywatch and KRES have successfully accounted for a range of results without selective attention, supporting the notion that the encoding and inferential processes they embody might be sufficient to account for knowledge effects.

On the other hand, there have been many suggestions that knowledge exerts its effects by directing attention to some sources of information at the expense of others (Murphy & Allopenna, 1994; Murphy & Medin, 1985; Pazzani, 1991; Wisniewski, 1995). For example, if a subset of a category's features can be related to a common theme, it is natural to suspect that the thematic knowledge might direct attention to those features and away from others (see Kaplan & Murphy, 2000, for discussion). Thus, our first goal is to determine whether knowledge indeed induces any change to what is attended.

Assuming that it does, a second question concerns the time course of that effect. Some theorists have suggested that the role of knowledge is to preselect dimensions (or hypotheses) for further testing (Keil, 1981). For example, Pazzani's (1991) rule-based PostHoc model selectively attends to goal-relevant features over the irrelevant ones and thus the model predicts preselection of the related dimensions. Kruschke (1993) suggested that his associative ALCOVE model can account for prior knowledge by

setting *initial* attention weights on the related dimensions higher than on the unrelated ones. Consistent with these proposals, studies have frequently found that knowledge effects appear quite early in learning. Indeed, Kaplan and Murphy found that knowledge effects are present after just one block of training. Pazzani found that his subjects learned to classify correctly in only a few trials, apparently because prior knowledge directed them to rely on features they knew were causally-related to the outcome. Furthermore, Heit (1995) reported that the largest effect of knowledge occurred before subjects saw *any* examples of category members (also see Heit, 1994, 1998).

However, in some circumstances the effect of prior knowledge might emerge gradually as a result of experience with category members. Because knowledge consists of representations in long-term memory, the number of observed category members needed for those representations to become sufficiently active will likely vary depending on (a) how strongly they're associated with observed features and (b) how strongly observed features are related to *other* representations. For example, in Murphy and Allopenna (1994), a feature such as "heavily insulated" might be related to not only arctic vehicles but also to buildings or soundproofing, "white" might also be related to clouds and weddings, "made in Norway" might also be related to smoked fish and fjords. For a theme that is *common* to many category features to become sufficiently active and thus noticed (or constructed), subjects may need to observe them multiple times, which would delay any apparent effect of prior knowledge (see Heit & Bott, 2000, for discussion). Another reason to expect a gradual (or delayed) knowledge effect is that a learner may only begin to make use of knowledge after a simpler learning strategy (e.g., one-dimensional rule-testing) fails to yield an acceptable solution. Thus, our second question concerns whether the attentional selection of the related dimensions can emerge gradually as a result of experience with category members.

Assuming that it can, the third question concerns whether error feedback is required to mediate that change. Supervised classification learning is often characterized as learners' adapting their responses to error feedback so as to minimize future errors (Kruschke, 2001). It is assumed that doing so involves redirecting attention to more diagnostic dimensions: If classification responses based on a certain dimension produce error, one should give less weight to that dimension and more to others in the future. This error-driven account might be extended to account for attentional adjustments during learning of

thematic categories, because, for example, error feedback might serve as a cue indicating to the learner to use prior knowledge (or to use *different* knowledge), which in turn might induce a shift in attention to features related to that knowledge.

However, there are good reasons to doubt that error-driven learning is the only source of change in attention. There is ample evidence in the unsupervised learning literature documenting thematic (family resemblance) sorting of items, a result that indicates mere observation of sets of related features is sufficient to activate prior knowledge (e.g., Kaplan & Murphy, 1999; Medin, Wattenmaker, & Hampson, 1987; Spalding & Murphy, 1996). In supervised learning as well, during error-free trials, sequential observation of thematically coherent category members may be sufficient to activate semantic representations that lead classifiers to allocate more attention to the related dimensions on future trials. Alternatively, attention change might be more dynamic. For example, Kaplan and Murphy (2000) reported suggestive evidence that, after subjects noticed themes that allowed errorless classification performance, they attempted to relate unrelated features to the themes, a result that indicates attention to the unrelated features in the absence of error. Indeed, their subjects' intact learning of unrelated features (as compared to the feature learning in the no-knowledge control condition) suggests continued or redirection of attention to the unrelated features. Likewise, Bott, Hoffman, and Murphy (2007) found that a pre-established feature-category association that allowed perfect classification did not prevent learning of additional features (i.e., they observed no "blocking"), implying a redirection of attention to stimulus dimensions without error. Thus, our third question concerns whether attention shifts occur only in response to incorrect trials or whether they also occur after learners have solved the learning problem, that is, in the absence of error.

To address these three questions, we report an eyetracking study of supervised learning of two categories each with a subset of the features related to a common theme. In cognitive research, eyetracking has been proved to be an effective tool to study on-line attention (e.g., Ferreira & Clifton, 1986; Grant & Spivey, 2003; Griffin & Bock, 2000; Haider & Frensch, 1999; Just & Carpenter, 1984; Lee & Anderson, 2001; Rayner, 1998), and in recent years has been successfully applied to studying selective attention in category learning tasks (Blair, Watson, & Meier, 2009a; Blair, Watson, Walshe, &

Maj, 2009b; Rehder & Hoffman, 2005ab; Rehder, Colner, & Hoffman, 2009; Watson & Blair, 2008; also see Kruschke, Kappenman, & Hetrick, 2005). We now use eyetracking to study how attention is affected by prior knowledge.

Overview of Experiments

Two novel categories of ants labeled “Dax” and “Kez” were constructed from six binary dimensions. To equate features’ category and cue validity, we used a one-away structure including the prototype for each category (Table 1). Whereas most previous studies of thematic category learning have used linguistic feature descriptions, we used spatially separated pictorial features suitable for eyetracking. Figure 1 illustrates two example category prototypes. In each category, four *related features* were associated with an environmental theme by describing them as useful in either a cold, tundra-like environment or a hot, desert-like environment. The other two *neutral features* were unrelated to these themes. Table 2 presents example feature descriptions for the prototypes in Figure 1, where antenna, mouth, forearm, and foot are theme-related, and tail and wing are neutral. Participants acquired this knowledge before category training by studying, one at a time and in a random order, each of the 12 features and their functions (without any information regarding which category they belonged to).

Note that to prevent the themes from being too obvious, neither the word “tundra” nor “desert” were explicitly mentioned in the descriptions. Instead, our intent was that each description only indirectly suggested those themes by virtue of their usefulness in such environments (similar to the materials used by Murphy, Heit, and others in past research). We have little doubt that there exist knowledge manipulations so strong that they result in learners immediately noticing the themes. To take an extreme example, we could have added the phrase “useful for surviving on the tundra” or “useful for surviving in the desert” to each related feature description, virtually ensuring that subjects would realize that there were two types of features. However, this would have meant that subjects would have, in effect, pre-learned the categories—all that would have been left was for them to learn which group were referred to as “Daxes” and which as “Kezes.” Although this is learning of a sort, it is not what is usually taken as category learning in which subjects must learn not only the category labels but also which features tend to co-occur with one another. Also recall that one of our goals, articulated in our second question, was to

determine whether the effects of knowledge *can* emerge during the course of learning—making the themes too blatant would have made this impossible. These materials allow for the possibility that the influence of the themes will appear gradually over the course of learning.

In Experiment 1, we conducted a non-eyetracking study to establish whether these new materials would induce the standard behavioral effects of prior knowledge on category learning. First, to confirm that such knowledge induces faster learning overall, we compared learning in the *related condition* in which knowledge was present to an *unrelated condition* in which all six dimensions were neutral with respect to the themes. Second, in the related condition we compared learning of the related dimensions to the neutral ones to confirm that the former were learned better than the latter (as in, e.g., Bott et al., 2007). Another purpose of Experiment 1 was to test whether subjects could learn the “prior knowledge” we provided as part of the experimental session. Unlike previous research on thematic category learning that used textual descriptions for features (e.g., “drives on glaciers”), our participants are required to first associate each pictorial feature to its functions (a technique also used by Krascum & Andrews, 1998). Participants' performance on learning those feature-function associations will evaluate the feasibility of this technique.

In Experiment 2, we conducted an eyetracking study to address the main questions surrounding attention and prior knowledge. Because the purpose of Experiment 1's unrelated condition was to confirm that our new materials would induce faster learning, we tested only the related condition in Experiment 2.

Experiment 1

Method

Materials. Dax and Kez categories were constructed from six binary dimensions: antenna, mouth, forearm, foot, tail, and wing. Table 1 presents category structure in which the Dax and Kez prototypes are 111111 (Figure 1A) and 000000 (Figure 1B), respectively. Four feature assignments to Dax/Kez prototypes were used to balance features to categories: 111111/000000, 101010/010101, 010101/101010, and 000000/111111. This counterbalancing resulted in each feature being paired with each other and the tundra or desert themes an equal number of times. In the related condition, Daxes were related to the tundra theme and Kezes to the desert theme. To relate Dax and Kez categories to the themes, four of the

six dimensions were accompanied with theme-related descriptions and the remaining two had neutral descriptions (see Table 2 for an example). To balance dimensions' screen location (e.g., top vs. bottom) or type (e.g., head vs. tail), the neutral dimensions were either tail and foot, wing and mouth, or forearm and antenna, with the remainder theme-related. In the unrelated condition, all dimensions were neutral. The Appendix presents the three types of descriptions (tundra, desert, or neutral) for each of the 12 features in Figure 1. The two experimental conditions (related vs. unrelated), four assignments of features to categories, and three assignments of related/neutral dimensions resulted in 24 experimental cells.

Participants. Thirty New York University undergraduates participated for course credit. They were randomly assigned to the 24 cells with the constraint of at least one person in one cell. This resulted in 14 and 16 participants in the related and unrelated conditions, respectively.

Procedure. The experiment consisted of three phases: knowledge acquisition, category learning, and a single-feature test. In knowledge acquisition, participants studied, one at a time, a total of 12 features, six from each category. Each screen displayed an ant with one visible feature and the other five features hidden behind gray rectangles (see top of Figure 2, for an example). Below the ant were descriptions of the visible feature. Importantly, at this point, no information regarding to which category the feature is typical was provided. Participants studied the 12 features on their own pace by navigating 12 screens with left/right arrow keys. The bottom of each screen displayed its number (1-12), and the presentation order of the feature was randomized for each participant.

To ensure learning, participants were required to take a multiple-choice test followed by a recall test. Both tests consisted of 12 questions, one for each feature. In the multiple-choice test, a question presented an ant with one visible feature, and participants chose one of the four alternatives below the ant (Figure 2). The alternatives included the correct answer—the description presented during the previous phase—and three incorrect answers. For instance, suppose "a" (the tundra description for the forearm feature, "1") is the correct answer to the question in Figure 2. The distractors included desert and neutral descriptions for the same forearm feature "1" and the tundra description of the other forearm feature "0". The order of the four alternatives was randomized for each question, and the order of the questions was randomized for each subject. Immediate feedback was provided for each question, and after the test, total

number of errors was also provided. When any error occurred during the test, participants were returned to the initial screens for additional study and then retook the test that presented only the questions they missed. This process repeated until all questions were answered correctly.

The purpose of the recall test was to ensure that participants could not only recognize but also recall the feature descriptions, because otherwise there would not be a knowledge effect during category learning. The recall test was the same as the multiple-choice test except that participants verbally described each feature instead of making a choice. The experimenter provided feedback for each question, and after the test, total number of errors was also provided. In particular, any error during the recall test obligated the participant to restart the knowledge acquisition phase all over again including initial learning, multiple-choice, and recall tests. This process repeated until participants answered all recall test questions correctly in a row. The knowledge acquisition phase took on average about 12 minutes.

The category learning phase began with two practice trials followed by the training blocks that randomly presented 14 exemplars, seven from each category (Table 1). Each trial began with a center cross (+) appearing for 1.8 s followed by presentation of an exemplar. Participants classified the exemplar as either a Dax or Kez by pressing "z" or "/" keys, respectively. Immediate feedback was provided in words below the exemplar ("Correct" or "Wrong") and the exemplar remained visible for 3.8 s after the response. For the practice trials, features were replaced with simple patterns and geometric shapes, and one trial displayed positive and the other trial displayed negative feedback. The learning ended after two consecutive errorless blocks or after the 15th block. Participants were informed of how close they were to this goal after each block.

Finally, a single-feature test followed category learning. Each trial presented an ant displaying one visible feature (as in top of Figure 2), and participants classified the feature as in training, that is, by pressing "z" for Dax and "/" for Kez. No feedback was provided. After each choice, participants rated confidence in their decision by positioning a slider on a scale whose left and right ends were labeled "Very Uncertain" and "Very Certain." The slider could be set to 21 distinct positions, and responses were scaled to range from 0 to 100. The order of the 12 features was randomized for each participant. The whole experiment took about 50 minutes.

Results

Because there were no effects of the counterbalancing factors in any of the following analyses, the results are presented collapsed over these factors.

Participants were very accurate in the knowledge acquisition phase. No participants made more than a total of 7 errors; 22 participants committed zero errors. Related ($M = .97$) and unrelated (.98) participants were equally accurate ($t < 1$), suggest that the thematic and neutral descriptions were equally easy to learn.

In category learning, 12 of 14 related participants and 14 of 16 unrelated participants reached the learning criterion of two consecutive errorless blocks. The 12 related learners reached the criterion in fewer blocks ($M = 5.5$) than the 14 unrelated learners (8.57), $t(24) = 2.85$, $p < .01$, while committing fewer total errors (8.67 vs. 19.07), $t(24) = 3.29$, $p < .01$. These results replicate previous research demonstrating faster learning of thematic categories (e.g., Murphy & Allopenna, 1994; Rehder & Ross, 2001).

The single-feature test results (Table 3) provide insight into which features participants learned and used for classification. Given that the themes helped category learning, we suspected that related learners would exhibit better learning of the related dimensions as a result of using these to distinguish the categories. Consistent with this expectation, the related learners showed greater accuracy on the related dimensions ($M = .89$) than the neutral ones (.71), $t(11) = 1.79$, $p = .05$ (one-tailed) and than the neutral dimensions in the unrelated condition (.76), $t(24) = 1.85$, $p = .07$. Although learning of the neutral dimensions was numerically lower in the related versus unrelated condition, this difference was not significant, $t < 1$, consistent with the result that prior knowledge helps learning without hurting learning of unrelated features (Kaplan & Murphy, 2000).

To obtain a more sensitive measure, *signed confidence ratings* were computed in which the ratings in each trial were set to the rating [0–100] when they were correct and negated when they were incorrect [–100–0]. More positive signed rating reflect more accurate and confident responding; zero reflects chance responding. Table 3 indicates that related learners' mean signed ratings were greater than 0 for both dimension types, p 's $< .01$. More importantly and consistent with the accuracy measure, the

ratings were greater for the related dimensions (67.1) than for the neutral ones (37.0), $t(11) = 1.82, p < .05$, and than for the neutral dimensions in the unrelated condition (43.8), $t(24) = 2.32, p < .05$. Once again, ratings on the neutral dimensions did not differ between the two groups, $t < 1$.

Analysis of training errors. In addition to the single-feature test, the pattern of errors during training can provide insight into which dimensions participants used at different points during training. For example, only committing errors on the pair of items that exhibit exception features on a particular dimension (e.g., on 011111 and 100000, which have exceptions on dimension 1) during a series of trials suggests that the learner is responding on the basis of a simple rule on that dimension (or at least giving greater weight to that dimension). Accordingly, in the related condition, we classified training items into those with exception features on related dimensions (R-exception items), on neutral dimensions (N-exception items), or those with no exception features (Prototypes). The relative probability of committing an error on R-exception versus N-exception items over blocks will reveal at what point during training participants' greater learning of the related dimensions' feature-category associations emerged. In particular, it indicates whether the effect of prior knowledge was immediate (i.e., occurred in the first training block) or whether it arose largely in later blocks after the learners had experience with exemplars.

Figure 3 presents the related learners' average error probabilities over blocks for each item type. So that each subject contributes to each data point, in constructing Figure 3 we assumed that participants would have continued error-free performance up through the 15th block had they continued classifying even after reaching the learning criterion of two errorless blocks (i.e., each subjects' final block performance was "padded out" for the full 15 blocks). The figure indicates that in block 1 the related learners actually had a *lower* error rate on the R-exception items (.219) than the N-exception items (.354). In other words, this group did not exhibit better learning of the related dimensions in the first 14 training trials. This conclusion is further supported by the lower block 1 error rate on R-exception items as compared to exception items in the unrelated condition in which all dimensions were neutral (.288). Note that despite their numerically smaller error rate, the R-exception items did not differ significantly from the N-exception items, $t(11) = 1.42, p = .09$, or the exception items of the unrelated group, $t(24) = 1.16, p > .20$.

In most training blocks after the first, up until errors are eliminated entirely, learners' error rate on the R-exception items was larger than on the N-exception items, an expected result given their greater learning of the related dimensions exhibited on the final single-feature test. After block 1, the average total number of errors made on the 8 R-exception items ($M = 3.9$) was about four times greater than on the 4 N-exception items (1.0). These results suggest that learners' use of the related dimensions did not emerge immediately but rather slowly after experience with category exemplars (and the receipt of error feedback). (Experiment 2 will present eye movement data that provides direct support for this interpretation.) Note that, as expected, Figure 3 indicates lower error rates on the Prototype items than either type of exception items.

Discussion

Experiment 1 replicated the standard results in thematic category learning. First, learning occurred in fewer blocks with fewer total errors in the related as compared to the unrelated condition. Second, single-feature tests showed related learners' better performance on related as compared to neutral dimensions. Importantly, these results obtained despite that the so-called “prior” knowledge was in fact presented to subjects as part of the experimental session. Together, these results confirm that our novel materials designed for eyetracking induce the standard behavioral effects of knowledge on learning.

Recall that while we intended that our materials would induce knowledge effects, it was also our hope that the themes would not be too obvious, because in such a case category learning comes trivial (i.e., learners only need to associate two response keys to the two themes). In fact, our results showed that the themes were not obvious to most subjects. Not only did 2 of 14 related participants fail to learn the categories, the remaining 12 subjects required between 2 and 10 blocks to learn ($M = 5.5$, $SD = 2.32$), suggesting that the categories were not pre-learned during knowledge acquisition. This interpretation is further supported by the analysis of training errors in which learning of the related versus neutral dimensions did not differ in the first block of training.

Experiment 2

Having determined that our new materials and training procedure induce the standard effects of knowledge on category learning, in Experiment 2 we set out to answer our main questions. Does

knowledge induce any change to what is attended (e.g., more attention to related vs. neutral dimensions)? Does that effect arise in response to observing examples of category members? And, is that change in attention mediated by error feedback? To answer these questions, we replicated Experiment 1's related condition using an eyetracker. Because the main questions involved comparison of related versus neutral dimensions, the unrelated group, having only neutral dimensions, was not tested in this experiment.

Method

Materials. The materials were the same as in Experiment 1.

Participants. Twenty-four New York University undergraduates participated for \$10. All had normal vision with corrective lenses or better. They were randomly assigned in equal numbers to one of the four assignments of features to categories and to one of three assignments of related/neutral dimensions.

Procedure. The procedure was the same as in Experiment 1, with a few additional steps for eyetracking during category learning phase. Participants were first fitted and calibrated to the eyetracker. Each trial began with a drift correction in which participants fixated on a center target allowing the eyetracker to recalibrate (compensate for small movements of the eyetracker on the subject's head). We used a *gaze-contingent display* such that a feature was fully visible when it was fixated but blurred when it was not, making it less likely for participants to use peripheral vision to obtain stimulus information. Gaze-contingent displays thus help ensure that eye movements are a reliable indicator of which information participants extract from the display. Eye movements were recorded from the left eye. After each classification response, auditory feedback indicated whether they were correct (a chime) or incorrect (a ding). The exemplar remained visible on the monitor for 4 s after the response. Following two practice trials, 14 exemplars were randomly presented in each block. The whole experiment lasted about an hour.

Eyetracking dependent variables. The eyetracking software yields, for each trial, a stream of fixations and their corresponding x-y screen locations and durations. We defined six circular *areas of interest* (AOIs) that encompass the features displayed on the screen. All fixations outside the AOIs were discarded, as were those that occurred after classification response. Using the remaining fixations, we computed four measures on each classification trial.

The first is the *number of dimensions observed* in each trial. To compute this measure, we counted a dimension "observed" if that dimension is observed at least once in each trial. Thus, it ranges [0–4] for the related dimensions and [0–2] for the neutral dimensions. The second, *fixation probability* ranging [0–1], is obtained by dividing the number of dimensions observed by 4 and 2 for the related and neutral dimensions, respectively. This measure equates number of each dimension types and thus indicates the probability that a related or neutral dimension is fixated in a trial. The third, *proportion fixation number* ranging [0–1], is computed by taking the number of fixations to the related dimensions and dividing it by the total number of fixations to all dimensions. Similarly, *proportion fixation time* is computed by taking the fixation time to the related dimensions and dividing it by total time fixating all dimensions. The fourth, *relative priority score* ranging [0–1], captures the relative ordering of fixations to the related versus neutral dimensions. To compute this measure, we first weighted fixations in each trial according to the terms in arithmetic sequence, $\{n, n-1, \dots, 1\}$, of n ordered fixations such that the first fixation was given a weight of n , the second fixation was given a weight of $n-1$, and the last was given a weight of 1. The relative priority score was then obtained by dividing the sum of weights to the related dimensions by the sum of total weights. Thus, a greater relative priority score indicates that the related dimensions were fixated earlier in a trial than the neutral dimensions.

Results

Basic learning results. Once again, participants were very accurate on the multiple-choice and recall tests during knowledge acquisition. Average test accuracy was .97; no participant made more than a total of 7 errors, and 12 committed zero errors. During training, 20 of 24 participants reached the learning criterion of two consecutive errorless blocks on an average of 6.5 blocks (as compared to 5.5 in Experiment 1) while committing on an average 10.60 total number of errors (8.67 in Experiment 1). The nonlearners committed on average 59.50 total errors.

Single-feature test. Table 3 presents single-feature test results. Consistent with Experiment 1, learners exhibited greater accuracy on the related dimensions (.91) than on the neutral ones (.70), $t(19) = 3.31, p < .01$. Once again, the signed confidence ratings were greater than 0 for both dimension types, p 's $< .01$, and greater for the related dimensions (73.6) than for the neutral ones (29.1), $t(19) = 5.75, p < .001$.

Analysis of training errors. Figure 4 presents the error rates on the different types of training items in each block. Recall that the error rates on the R-exception and N-exception items provide information regarding at what point participants' begin to exhibit greater learning of the corresponding dimension. Unlike in Experiment 1, in block 1 the error rate on R-exception items (.231) was numerically larger than on N-exception items (.175). This difference did not reach significance, however, $t(19) = 1.31$, $p = .10$, replicating Experiment 1's results that the related features were not pre-learned during knowledge acquisition and that the related learners do not show better learning of those dimensions during initial period of training. After block 1, learners made an average of 5.9 total errors on the 8 R-exception items as compared to 1.6 on the 4 N-exception items (compare with 3.9 and 1.0 in Experiment 1). Once again, use of the related dimensions emerged only after learners observed a number of category members (and committed a number of classification errors).

Eye fixations. The primary goal of Experiment 2 is to investigate how thematic coherence affects attention allocation during category learning. We first present the number of related and unrelated dimensions observed in each block averaged over the 20 learners (Figure 5A). For constructing this and the following graphs, we assumed that learners' eye movements would have been identical to those in their last block had they continued classifying for the full 15 blocks (and thus every subject contributes to each data point in the figure). Figure 5A shows that learners initially observed about 2.9 of the 4 related dimensions on each trial and gradually increased their fixations to those dimensions over the course of training. In contrast, they initially observed about 1.3 of the 2 neutral dimensions and gradually decreased those fixations during training. To equate the different number of related and neutral dimensions, Figure 5B presents the probabilities of fixating each dimension type. The figure indicates that the learners fixated the two types of dimensions with about equal probability at the start of training but became more (less) likely to fixate the related (neutral) dimensions. By the end of learning, the probability of fixating a related feature was .86 as compared to .37 for the neutral features.

A 2 x 2 within-subjects ANOVA was conducted on the fixation probabilities in Figure 5B with dimension type (related vs. neutral) and block (first vs. last) as factors. There was a main effect of dimension type, $F(1, 19) = 20.294$, $MSE = .079$, $p < .001$, confirming the greater chance of fixating

related dimensions. There was no main effect of block ($p > .10$), but a significant interaction between dimension type and block, $F(1, 19)=25.904$, $MSE = .033$, $p < .001$, confirmed the increase (decrease) in fixating the related (neutral) dimensions. T-tests revealed that learners were more likely to fixate the related dimensions than the neutral ones in all blocks, p 's $< .03$, except block 1, $p > .09$. Note that even the small (and nonsignificant) difference in block 1 might have resulted from fixations in the later trials of the block. Consistent with this conjecture, the probabilities of fixating related and neutral dimensions during the first half of block 1 (i.e., the first seven trials) were .71 and .70, respectively, $t < 1$. This pattern of eye movements corroborates the conclusion, reached on the basis of training errors, that learners have no preference for attending the related dimensions at the start of training.

These results are further supported by the more sensitive eyetracking measures—proportion fixation number and time—presented in Figure 5C. Because there were four related and two neutral dimensions, a value of .67 ($= 4/6$) reflects a bias toward neither dimensions. The figure shows that both proportions start off slightly greater than .67 and then shift in favor of the related dimensions. T-tests comparing the first and last blocks confirmed increases in both proportions, p 's $< .001$. In addition, both proportions were significantly greater than .67 in all blocks, p 's $< .02$, except blocks 1 and 2, p 's $> .05$; in the first half of block 1 (trials 1–7) these proportions were .66 and .67, respectively. These results are consistent with the fixation probabilities in Figure 5B demonstrating that the learners had no preference for fixating the related dimensions at the start of training. Instead, the effect of knowledge on attention emerged gradually with the observation of category members (and the receipt of error feedback).

Finally, the relative priority score presented Figure 5C indicates that, within a trial, learners also had no preference for fixating the related dimensions before the neutral dimensions at the start of training. The priority score starts just above .67 (indicating that the two dimension types had about equal priority) and then gradually increases in favor of the related dimensions over blocks. The priority score was reliably larger in the last block as compared to the first, $t(19) = 3.75$, $p < .01$, and was greater than .67 in all blocks, p 's $< .05$, except the first, $p > .05$; the difference between the relative priority score in the first half of block 1 (.70) and .67 did not approach significance, $t < 1$.

Backward learning curves. The fixation analyses indicated learners' gradual shift in attention to

the related dimensions during the course of training. We also asked how that shift relates to reduction in classification error. One possibility is that learners could have shifted attention only after they learned to distinguish the categories (i.e., after the last classification error), because in the presence of error they might have felt it necessary to consider the neutral dimensions. For example, participants in Rehder and Hoffman (2005a) stopped fixating nondiagnostic dimensions only after errors were largely eliminated, suggesting the above possibility. Alternatively, in the learning of thematic categories, learners might begin to attend more (less) to the related (neutral) dimensions before the last error, which then allows them to solve the classification problem using thematic knowledge.

To relate error reduction to change in attention, we created backward learning curves (Figure 6) by translating each subjects' trial numbers so that their last error occurred on trial 0. Because the primary interest was the relationship between knowledge use and error reduction, we included only 14 of the 20 learners (i.e., whom we dub the *knowledge users*) whose single-feature test results and eye movements both showed evidence of knowledge use.¹ Figure 6 presents their averaged backward learning curves for a number of dependent variables. In the graphs, we included only the 60 trials (about 4 blocks) before the last error and 28 trials (the 2 blocks of learning criterion) after the last error. Because of the relatively smaller number of participants, each data point presents performance averaged over a subblock of four trials.

Figure 6A presents the probability of error averaged over the knowledge users (the jump in error rate at subblock 0 obtains because that data point always includes the last error at trial 0). Of greater interest are Figures 6B and 6C, which present the number of dimensions fixated and fixation probabilities, respectively. First, consider the eye fixations that occurred before the last error (i.e., in the figures, on negative blocks). Both figures indicate a shift in attention consisting of a sharp increase in fixations to the related dimensions (and perhaps a small decrease in fixations to the neutral dimensions) that begins from about three blocks before the last error. T-tests on the fixation probabilities confirmed that the related dimensions were fixated with greater probability than the neutral dimensions starting from the data point indicated by an arrow (Figure 6C), p 's < .05. Both the proportion fixation time and number measures (Figure 6D) were also significantly greater than .67 from the same data point indicated by an arrow, p 's

< .01. These results establish that the knowledge users began to direct their attention to the related dimensions well before the last error.

Next, consider the eye fixations that occurred after the last error (i.e., on positive blocks). Figures 6B, 6C, and 6D all indicate continued shift in attention after the last error, that is, despite the *absence* of error feedback. After the last error, fixation probabilities for the related dimensions rose from .80 to .91, and those for the neutral dimensions dropped from .50 to .27.² A 2 x 7 within-subjects ANOVA was conducted on the fixation probabilities (Figure 6C) with dimension type (related vs. neutral) and subblock (1 to 7) as factors. There was a main effect of dimension type, $F(1, 13) = 40.617$, $MSE = .281$, $p < .001$, confirming the greater chance of fixating related dimensions. There was no main effect of subblock ($F < 1$), but a significant interaction between dimension type and subblock, $F(6, 78) = 5.796$, $MSE = .020$, $p < .001$, confirmed the increase (decrease) in fixating the related (neutral) dimensions. Considering the two types of dimensions separately, fixation probabilities increased from the first two positive subblocks to the last two ($p = .10$) for the related dimensions and decreased for the neutral dimensions ($p = .09$). The two proportions presented in Figure 6D also showed a reliable increase during the positive subblocks, p 's < .05.

For completeness, Figure 7 presents the backward learning curves for the 6 of the 20 learners (i.e., the *knowledge nonusers*) whose single-feature test results and eye movements neither showed evidence of knowledge use (see footnote 1). Figures 7B and 7C show that these subjects gradually fixated more dimensions but that the fixation probabilities to the two dimension types are virtually indistinguishable. The proportion fixation number and time (Figure 7D) fluctuates around .67, indicating the absence of a knowledge effect on attention for these subjects.

Individual differences. We also asked whether the pattern of eye fixations exhibited in Figure 6 was manifested consistently by all knowledge users or whether there was substantial variability. In fact, we identified six knowledge users whose eye movements were similar to one another but distinct from the group average in Figure 6. These subjects (whose backward learning curves are presented in Figure 8) were distinct because they (a) fixated all six dimensions at the start of training, (b) showed at best only a small and irregular preference for fixating the related dimensions before committing their last error, and

(c) learned the categories very quickly. On average, these six subjects made an average of only 3.0 total errors, the last one of which occurred on trial 6.7—two of these subjects made only two total errors and two made only one. Although we do not wish to over-interpret the data from such a small number of subjects (e.g., the data points corresponding to negative subblocks are based on only three or fewer subjects), we think it likely that these individuals recognized the category themes before the start of training, that is, during the knowledge acquisition phase of the experiment, and thus only needed to learn which label went with which theme (whether the tundra ants were labeled “Dax” and the desert ants “Kez” or vice versa). Nevertheless, just like the rest of the knowledge users, these subjects showed a strong preference for fixating the related versus neutral dimensions by the end of learning. Note that these subjects are responsible for the increase and then decrease in fixations to the neutral dimension in Figure 6C that occurs around subblock 0.

Relating eye movements during training to test performance. Finally, we related eye movements during training to the single-feature test results. We hypothesized that participants would exhibit better learning of the dimensions as a function of how often they were fixated during training. Accordingly, we submitted each learner's mean signed confidence ratings on each dimension to a regression analysis in which the predictor was the proportion of all fixations during training that a dimension received. The average weight assigned to this predictor (i.e., slope; $M = 298.0$, $SD = 308.0$) was significantly greater than 0, $t(19) = 4.33$, $p < .001$, and indicates that for each .10 increase in proportion fixation number, the signed confidence rating increased by 29.8. In other words, fixating a dimension more often meant that it was learned better. Indeed, that the mean intercept in this analysis ($M = 9.1$, $SD = 56.5$) did not differ from 0, $t < 1$, reveals the expected result that no learning of a dimension occurred if it was never fixated. Similar results were obtained using proportion fixation time as a predictor.

We also asked whether the better learning of related dimensions can be explained solely by the greater attention they received. Learners' signed confidence ratings were predicted from (a) proportion fixation number and (b) whether the dimension was related or neutral. In this analysis, both proportion fixation number, $t(19) = 2.66$, $p < .05$, and dimension type, $t(19) = 2.45$, $p < .05$, were significant predictors; the signed confidence rating increased by 20.9 for each .10 increase in proportion fixation

number and by 28.8 for related versus neutral dimensions. Qualitatively similar results obtained when the predictors were proportion fixation time and dimension type. In other words, the better learning of related dimensions was partly, but not fully, mediated by the greater number of fixations they received. We'll return to these results in the General Discussion.

Discussion

Experiment 2 answers our three questions regarding the effects of knowledge on attention. First, eye fixations showed that prior knowledge indeed affects what category information is attended, as learners devoted more attention to related dimensions than neutral ones: By the end of learning, the probability of fixating a related feature was .86 as compared to .37 for the neutral features. Second, learners generally showed no initial tendency to fixate related features more than neutral ones but then gradually shifted attention to related dimensions during the course of training. Third, this shift continued after the classification problem was solved, that is, in the absence of error feedback: Whereas knowledge users were 30% more likely to fixate a related feature dimension than a neutral one when they committed their last error (fixation probabilities of .80 and .50, respectively), that difference grew to over 63% (.91 and .27) by their final trial. Finally, eye fixations during training were a significant predictor of feature learning, although above and beyond fixations there was still an effect of whether the dimension was related or neutral.

General Discussion

This article has addressed how prior knowledge—that relates features of categories—influences attention to category features. As reviewed earlier, prior research has documented large effects of such knowledge on how people learn and reason with categories, and many investigators have considered the possibility that these effects are often mediated by changes in attention. Such proposals are perhaps unsurprising given that attention weights are the main “free parameters” of the field’s standard models of category learning—thus, explaining knowledge effects in terms of attention weights potentially allows those models to be applied successfully to yet another body of empirical findings without change to their basic assumptions. Yet, in the absence of any direct evidence of attentional changes, such proposals have remained speculative.

In the following sections we discuss the answers to the three questions we posed regarding the effect of knowledge on attention and their implications for current models of knowledge-based category learning. In the final section we relate knowledge's effect on attention to other effects it has on category learning.

An Effect of Prior Knowledge on Attention

Using eyetracking, we first sought to answer the most basic question, namely, whether knowledge induces any change to what is attended. The answer is that it does. Recall that we argued that knowledge might exert its effects not through attention but solely through how category information is encoded. The memory literature is rife with examples of how items can be recalled more readily when they are encoded with respect to existing knowledge, and better memory for category features can't but help promote successful classification. And, the presence of knowledge might allow classification to become an act of inference in which people reason from observed features to category membership. But, contra these accounts, in Experiment 2 a preference to fixate related versus neutral features began to emerge in the second block of training; by the end of training, learners were over twice as likely to fixate related features as compared to neutral ones. To our knowledge, this is the first direct confirmation of the frequent proposal that prior knowledge directs attention to knowledge-relevant information (e.g., Heit & Bott, 2000; Kruschke, 1993; Murphy & Medin, 1985; Murphy & Allopenna, 1994; Pazzani, 1991; cf. Kaplan & Murphy, 2000).

This pattern of attention induced by prior knowledge has important implications for current models of knowledge-based learning. On the positive side, in past simulations of supervised category learning, both KRES and Baywatch correctly predicted the faster learning in the presence of knowledge in the related condition as compared to unrelated condition observed in Experiment 1. And, both models have predicted the better learning of related features versus neutral ones when category features consisted of both related and neutral dimensions, a result that we also obtained in both Experiments 1 and 2. However, the poorer learning of neutral features is predicted not because of the reduced attention allocated to neutral features, but rather because of various forms of cue competition that arise from error-driven learning. For example, in KRES (Rehder & Murphy, 2003; Harris & Rehder, 2006) knowledge-

related features tend to increase each other's activation which in turn accelerates the increase in connection strengths between them and the category label. Because this also results in error being reduced more rapidly, it slows the increase in the connection strengths between the neutral features and the category label. In Baywatch (Heit & Bott, 2000), related features are learned faster because they are additionally connected to the category label via common prior concept units that accelerate their learning at the cost of the neutral features. These processes used to explain the better learning of related versus neutral features are analogous to those used by the Rescorla-Wagner learning rule to account for the classic phenomenon of *overshadowing* in the animal learning literature (Kamin, 1969; Rescorla & Wagner, 1972).

However, it is well known that many standard effects of cue competition can arise through not only from the dynamics of error-driven learning but also attentional mechanisms (Kruschke, 2001, 2003; Kruschke & Blair, 2000; Mackintosh, 1975; Sutherland & Mackintosh, 1971). For example, Kruschke et al. (2005) found that the learning of a cue-outcome relationship was slowed when the cue had been irrelevant during a previous learning stage, and eyetracking confirmed that this effect arose in part because learners failed to fixate the cue (i.e., they exhibited *learned inattention*). The present results suggest such attentional effects on learning hold in knowledge-based category learning as well: Because learners attend the neutral features less often, they will be learned less well than the related ones—at the extreme, learning of a feature will cease entirely once it stops being fixated. Thus, neutral features were at a double disadvantage in learning, suffering from both the effects of cue competition we have described and the fewer attentional resources they receive. (Whether neutral features are at disadvantage in absolute terms, that is, relative to situations where prior knowledge is absent entirely, is an important question that we return to later.)

The second implication that our eye movement results have for models concerns how items ended up being categorized at the end of training. That neutral features are learned less well of course means that they are contributing less to learners' accurate classification performance. But on top of that, at the end of training the neutral features were fixated less often than the related ones. In other words, the neutral features were at a double disadvantage in classification as well—they provided a relatively weak

source of evidence for category membership that was largely ignored anyway. As a consequence, items ended up being classified largely on the basis of the knowledge-relevant information.

Models like Baywatch and KRES, in contrast, assume that information about all features enter the network on every trial throughout training—in effect, they assume that related and neutral features are attended equally. But ignoring that neutral features receive fewer attentional resources means that in past simulations these models have mistakenly attributed the poorer learning of those features solely to effects of cue competition. And, doing so means that they have mistakenly attributed neutral features' limited influence on final classification performance solely to their poorer learning. In other words, in the absence of mechanisms that direct attentional resources toward knowledge irrelevant information and away from irrelevant information, Baywatch and KRES mischaracterize the effects of prior knowledge on how features are learned and their ultimate role in classification performance.

Knowledge Selection and Construction in Response to Observed Category Members

The second question we asked is whether the effect of knowledge on attention can emerge as a result of observing category members. The answer is that it can. Recall that there are good reasons to expect that the impact of knowledge will be greatest initially and then decrease with increasing experience with category members, at least under some circumstances. This account is appealing as prior knowledge is that which learners *bring to* the learning task, as compared to empirical observations that come later. As mentioned, Heit (1995) found that predictions (e.g., of whether a shy person would avoid parties) were initially influenced by only prior knowledge but later came to be more dominated by empirical observations. And, Pazzani (1991) found that subjects learned to predict an outcome (whether a balloon would inflate) in only a few trials, apparently because prior knowledge directed them to rely on features they knew were causally related to the outcome (stretching the balloon beforehand helps, and adults are usually more successful than children). But, contra this account, most Experiment 2 learners tended to allocate eye fixations to knowledge-relevant features only after exposure to multiple category members.

It is important to note that the studies of Heit and Pazzani differ from the present one in a crucial way, namely, that the outcomes being predicted (attending parties, a balloon inflating) were already

familiar to subjects and so served as a ready cue to what prior knowledge was likely to be relevant (see Wisniewski & Medin, 1994; Wisniewski, 1995 for other examples of the effect of such “top-down” knowledge). But although category labels are sometimes familiar, the labels of most new categories are themselves new. The labels “Blackberry” or “iPod” (or “credit default swap”) were initially as opaque to you as “Kez” was to our subjects. In such cases, prior knowledge must enter via a different route, namely, through the semantic associates of features of category members. But because these items usually have many features, and each of those features have many associations, determining which semantic representations are relevant to the current learning problem is unlikely to occur immediately.

Heit and Bott (2000) have labeled the process by which observations activate relevant semantic representations as a “knowledge selection,” and, like us, emphasized that many observations may be required before relevant knowledge is identified. For example, although during training our subjects may have tried to make use of feature descriptions with phrases such as “slippery ground,” “low temperature,” and “hard soil,” it may not have been immediately obvious what those concepts had in common or how they might be related to each other. However, repeated presentation of the related features (and repeated recall of the feature descriptions) eventually allowed them to triangulate onto what they had in common: that the ground was slippery because it was *icy* (rather than merely wet), that the soil was hard because it was *frozen* (rather than just highly compacted), and both of these things were true because the temperature was *below freezing* (not just the “low” 40° of a chilly autumn afternoon). And of course learners only directed attention toward theme-relevant features after realizing that the ant was adapted to a cold, icy environment.

Models like Baywatch and KRES, in contrast, assume that knowledge is in place from the start of training rather than being constructed dynamically in response to observed category members. But although these models account for a variety of knowledge-based effects, they are at odds with the present eyetracking data. An example serves to illustrate. Baywatch was used to model an experiment reported by Heit and Bott (2000) in which subjects learned to distinguish “Doe” buildings that had features common to churches from “Lee” buildings that had features common to office buildings. As in the current experiments, they found that an effect of knowledge on learning was absent initially and then increased as

training progressed. Baywatch accounted for these results by incorporating prior knowledge in the form of network nodes that represented prior churches and office buildings and that had existing connections with features typical of those concepts (e.g., “steeply angled roof” for churches, “flat roof” for office buildings). Even though this knowledge was built into the model beforehand and so was present from the start of training, it had no initial effect on classification because the strengths of the connections between it and the category labels were initially zero (i.e., the model still had to learn that the church-like buildings were labeled “Doe” and the office-like buildings were labeled “Lee”). But although this account explains the absence of any effect of knowledge on *accuracy* in the first learning block, it fails to explain the absence of any first block effect on *eye movements* in our Experiment 2: If the tundra and desert themes were active from the start, we would have expected a preference to fixate feature dimensions related to that knowledge. But as we have seen, the effect of knowledge on eye movements was absent initially and then *grew* as training proceeded, suggesting that the effect of knowledge on attention slowly emerged as a function of exemplar observations. Thus, although we think that Baywatch captures much about knowledge-based category learning, its assumption about built-in knowledge oversimplifies the process by which attention gradually shifts in response to observed category members. A similar critique applies to KRES that also assumes that knowledge is in place from the start of training.

Because discovery of a common theme is instead a stochastic process that occurs when category features activate related semantic associations predicts that the point of theme discovery will vary substantially from study to study depending on the stimulus materials and learning procedure and, within the same study, from subject to subject, and this is what has been found. For example, whereas Heit and Bott (2000) found no first-block learning advantage for related features during the learning of church and office buildings, they did when subjects learned to distinguish a tractor-like type of vehicle from one with many features of race cars. Similarly, Kaplan and Murphy (2000) found better learning of related features after one learning block consisting of only 12 trials—a striking result given that the 12 relevant features were each presented once (one per training exemplar). One difference between these studies and ours of course concerns the knowledge used. Whereas they relied upon semantic representations their subjects possessed before they walked into the experiment, we instructed our subjects on knowledge as part of the

experiment. Because this “prior” knowledge was probably not encoded as strongly as real-world semantic representations, it is likely that it was retrieved less readily (i.e., it was less “available”) during the category learning phase.

Moreover, we have reported the substantial variability in how our own subjects used the prior knowledge we provided. On the one hand, because one group of subjects (Figure 7) exhibited neither enhanced attention to (nor learning of) related versus neutral features, they apparently never noticed the themes common to each categories’ features. This result is perhaps not too surprising because, in creating the materials, we intended to make feature descriptions only weakly suggestive of the themes. But on the other hand, another group of subjects (Figure 8) made few errors and ended up fixating virtually every related feature on every trial (and rarely fixated the neutral ones)—these six individuals apparently noticed and made use of the themes quite early.

We suggest that these differences arise because the linkages between observed features and semantic representations vary greatly across domains and individuals. Theme discovery will occur quickly if a category’s features have many highly available semantic associates in common and few that are not common, increasing the chances of the simultaneous activation of related representations that is required for a theme to be noticed (or constructed). It will occur slowly if features have many available associates but few in common such that simultaneous activation of related representations is unlikely.

The (Non)Necessary Role of Error in Attending Knowledge-Relevant Information

The third question we asked is whether error feedback is required to mediate shifts in attention to knowledge-relevant information. The answer is that it is not. Recall that we noted that all current accounts of how attention changes during learning are based on error. For example, ALCOVE predicts gradual shifts in attention to stimulus dimensions that reduce error (Kruschke, 1992). Hypothesis testing models also assume that attention shifts between dimensions when classification errors result in the rejection of old rules (Nosofsky et al. 1994; also see Kruschke & Johansen, 1999). Although these models generally do not specify how attention shifts might be influenced by prior knowledge, one might imagine that error is necessary to abandon a simpler learning strategy (e.g., testing one-dimensional rules) and start a search of semantic memory for relevant knowledge (or to change what knowledge is being used). But, contra this

account, eye movements in Experiment 2 showed that attention continued to shift to the related dimensions even after subjects learned to classify all items correctly: Whereas knowledge users were about 30% more likely to fixate a related feature dimension than a neutral one when they committed their last error (80% vs. 50%), by their final trial that difference was greater than 60% (91% vs. 27%). Error is not a necessary condition for knowledge-induced changes in attention.

We propose two possible explanations for attention shifts in the absence of error. The first is the processes of theme discovery we have described, that is, through the simultaneous activation of semantic representations common to several category features. We don't assume that this process is independent of error, of course. As mentioned, error might trigger the abandonment of a current strategy and thus a search through semantic memory for useful representations. However, in our experiments merely observing features of to-be-classified stimuli may have been sufficient for learners to recall the feature descriptions we provided which in turn activated related representations, enabling the discovery of the tundra and desert themes. Moreover, the relatedness of the four knowledge-related dimensions need not have occurred at the same time—learners may have noticed the thematic relationship between only two or three of these dimensions before committing their final error and discovered the others afterwards.

Although attention shifted in the absence of error, it is important to note that our learners were still receiving feedback, and it is possible that this *positive* feedback helped promote the discovery of the themes. On the other hand, the extensive literature documenting knowledge effects in *unsupervised* category learning suggests that the discovery of category themes can occur in the absence of any sort of feedback. When subjects are simply asked to sort simultaneously-presented items into groups (receiving no feedback) in the absence of prior knowledge, the well-known finding is that they usually do so on the basis of a single dimension (e.g., green items go into one pile, white ones into another). But when prior knowledge is available, they produce theme-based sorts instead, that is, items with many features of one theme (e.g., jungle vehicles) go into one group while those with features of another (arctic vehicles) go into another (Kaplan & Murphy, 1999; Medin et al., 1987; Spalding & Murphy, 1996). In other words, merely observing the items and their features was sufficient to activate the underlying themes. That Experiment 2 learners shifted attention to theme-related dimensions after reaching criterion suggests that

spontaneous theme construction can also occur during classification learning, that is, even when items are presented sequentially rather than simultaneously.

A second reason that attention might shift without error is that it is likely that our cognitive systems are trying to not only increase *accuracy* but also decrease *response time*—all else being equal, a faster categorization response is more adaptive than a slower one. Indeed, the response times of Experiment 2's 14 knowledge users decreased steadily from 6.8 s at the point of their last error to 3.9 s at the end of training. One way that latency can be decreased is by gathering less information in preparation of a decision, and of course to maintain accuracy lower quality sources of information should be discarded before higher-quality ones. For example, our analysis of errors on different types of training items in both Experiments 1 and 2 indicated that, before committing their last error, subjects had already learned the related features better than the neutral ones. But once errors were eliminated entirely, the need for speed led our learners to recognize that fewer dimensions were needed for accurate performance—in our category structure only 3 of 6 dimensions were required to classify all training items perfectly. And, given a choice, the poorly learned neutral dimensions were the first to go (see Nelson & Cottrell, 2007, for one computational implementation of this idea).

The desire to increase speed in addition to reducing error should lead to attention shifts in situations not involving prior knowledge, and in fact this has been found. For example, Rehder and Hoffman (2005a) found that learners first discovered a one-dimensional rule that distinguished two categories but then only restricted attention (eye movements) to that dimension several (error-free) trials later. Moreover, although this result and the present experiments show that attention can shift when learners are receiving only positive feedback, other studies have found shifts in the absence of any feedback whatsoever. For example, Blair et al. (2009a) found that learners continued to optimize attention even after a criterion of 24 correct trials was reached and feedback stopped altogether. Of course, attentional shifts in the absence of error pose problems for all category learning models that tie attentional learning to error-driven mechanisms (e.g., Kruschke, 1992).

These possibilities suggest that knowledge-induced attention shifts can be both a *cause* and an *effect* of learning. On the one hand, as Pazzani (1991) proposed, prior knowledge can direct attention to

information needed for learning. In our experiments as well, subjects began to favor related dimensions before their last error, suggesting that doing so helped them solve the learning problem. But attention shifts can also reflect learning that has already occurred, as when less valuable sources of information are bypassed in order to respond more rapidly.

Attention versus Encoding, Inference, and Interpretation in Knowledge-Based Category Learning

Finally, it is important to consider to what extent the effect of prior knowledge on category learning can be understood as being mediated by its influence on attention. Recall that in Experiment 2 we asked whether feature learning could be predicted by eye fixations alone but found instead that related features were learned better than neutral ones even controlling for fixations, indicating that knowledge exerts its effects through means other than attention. We now consider three possibilities, namely, encoding, inference, and feature interpretation.

First, there is ample evidence that the sort of encoding processes enabled by prior knowledge we have mentioned. Recall that although we found a learning advantage for related over neutral features in the presence of prior knowledge, in Experiment 1 we found that the neutral features were learned *no worse than* neutral features in the unrelated control condition. These results replicate those by Kaplan and Murphy (2000) who, using a somewhat different category structure, also found no evidence that neutral features were learned worse in the presence of knowledge. They noted that these results are surprising given the standard error-driven learning accounts we have reviewed that predict that faster learning of some features must come at the expense of slower learning of others. And of course they are doubly surprising in light of the new findings reported here that neutral features are also attended less often, putting them at a double disadvantage relative to related features. We believe that the answer to this apparent paradox lies in how knowledge affects not just how features are attended but also how they are processed and thus encoded. For example, Kaplan and Murphy also found evidence that learners attempted to assimilate the supposedly neutral features to the categories' themes (e.g., posttests revealed that after learning subjects were more likely to think that an arctic vehicle would have a previously knowledge-irrelevant feature, e.g., "license plate on front," if it was associated with arctic vehicles during training). We think it likely that similar processes also occurred in our experiments. For example, whereas

in Table 2 only the antenna, mouth, forearm, and foot of Daxes are related to the tundra theme, subjects may have also tried to construct some reason why the tail and wings of Daxes (intended by us to be “neutral”) were in fact useful in such environments; even if unsuccessful, the mere attempt to relate these features to an existing knowledge structure may have made them more recallable (Heit, Bott, & Briggs, 2004). In other words, when knowledge is present, learning is not a zero-sum game. Instead, it provides the semantic relations and structures that promote the effective encoding of many sources of information, even those that are only peripherally related.

Second, we have also mentioned how in many cases prior knowledge allows classification to become an act of inference. For example, in Heit and Bott’s (2000) study, subjects’ mental representation of Doe buildings probably included the fact that they were “church like,” suggesting that during classification they used observed features to infer the concept “church” and from that the concept label “Doe” (this inferential process is explicit in their Baywatch model). Consistent with this interpretation, Heit and Bott found that learners classified a feature as “Doe” even if was never observed during training so long as their prior knowledge indicated that it was typical of churches. (Subjects were at chance on never-presented features that were associated with neither churches nor office buildings.) In our own study as well, we have little doubt that had we tested subjects on features related to the arctic and desert themes that were not presented during training they would have correctly inferred that they were related to the “Dax” and “Kez” categories, respectively. In fact, in our single-feature test, although this inference may not necessarily have helped subjects’ accuracy (as in e.g., Hoffman, Harris, & Murphy, 2008), it might have changed their subsequent confidence ratings by the virtue of being more certain when the features were related versus neutral. Other studies provide evidence of the inferences in service of classification that knowledge supports. For example, Rehder and Kim (in press) found direct evidence of how classifiers can use of knowledge (in their case, causal knowledge) to infer from observed features the presence of unobserved properties that determine category membership. And, Rehder and Ross (2001) found, like Bott and Heit, that people can correctly classify items with features never seen during training on the basis of semantic knowledge relating them to the category. Finally, recall Murphy and Medin’s (1985) well-known example of classifying a party-goer who jumps into a pool as *drunk*—one reasons

from aberrant behavior to its underlying cause even if one has never before observed a swimming drunk. Clearly, classification performance on novel items cannot be solely explained in terms of how those items were attended and encoded during training.

Finally, prior knowledge can also influence how stimuli are interpreted in the first place. For example, like the studies of Pazzani (1991) and Heit (1995) we have reviewed, Wisniewski and Medin (1994) used familiar category labels but, unlike those studies, used ambiguous pictorial stimuli. They found that the category labels influenced how the pictures were interpreted. For example, in the same picture a character was interpreted either as “dancing” when subjects were told the pictures were drawn by “creative” versus “noncreative kids” (dancing was taken to be a sign of creativity) or as “climbing in a playground” when told they were drawn by “city” versus “farm kids” (playgrounds are in cities but not on farms). For present purposes, the important point is that there is no reason to think that these differences arose from differences in *attention* between the two conditions—the character was attended equally, but interpreted differently.

In summary, although we think that attention is an important vehicle by which knowledge influences category learning, knowledge also exerts its influence through other means, including how stimulus items are encoded, the inferential processes it supports, and how features are interpreted.

Summary

Using eyetracking we found that (a) knowledge indeed changes what features are attended, with knowledge-relevant features being fixated more often than irrelevant ones, (b) this effect was not due to an initial bias to attend relevant dimensions but rather emerged as a result of observing category members, and (c) this effect grew even after a learning criterion was reached, that is, in the absence of error feedback. We argue that models of knowledge-based category learning will remain incomplete until they include mechanisms by which prior knowledge is selected dynamically in response to observed category members and which then directs attention to knowledge-relevant dimensions and away from irrelevant ones.

References

- Blair, M., Watson, M. R., & Meier, K., M. (2009a). Errors, efficiency, and the interplay between attention and category learning. *Manuscript submitted for publication.*
- Blair, M., Watson, M. R., Walshe, R. C., & Maj, F. (2009b). Extremely selective attention: Eyetracking studies of dynamic attention allocation to stimulus features in categorization. *Manuscript submitted for publication.*
- Bott, L., Hoffman, A. B., & Murphy, G. L. (2007). Blocking in category learning. *Journal of Experimental Psychology: General, 136*, 685-699.
- Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorical word lists. *Journal of Verbal Learning and Verbal Behavior, 8*, 323-343.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268-294.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348-368.
- Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thoughts. *Psychological Science, 14*, 462-466.
- Griffin, Z., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*, 274-279.
- Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 172-190.
- Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory and Language, 34*, 686-708.
- Harris, H. D., & Rehder, B. (2006). Modeling category learning with exemplars and prior knowledge. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1440-1445). Mahwah, NJ: Erlbaum.

- Heit, E. (1994). Models of the effects of prior knowledge on category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1264-1282.
- Heit, E. (1995). Belief revision in models of category learning. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 176-181). Mahwah, NJ: Erlbaum.
- Heit, E. (1998). Influences of prior knowledge on selective weighting of category members. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 712-731.
- Heit, E., & Bott, L. (2000). Knowledge selection in category learning. In D. L. Medin (Ed.), *Psychology of learning and motivation* (pp. 163-199). San Diego: Academic Press.
- Heit, E., & Bott, L. B., J. (2004). Modeling the effects of prior knowledge on learning incongruent features of category members. *Journal of experimental psychology. Learning, memory, and cognition*, *30*, 1065-1081.
- Hoffman, A. B., Harris, H. D., & Murphy, G. L. (2008). Prior knowledge enhances the category dimensionality effect. *Memory & Cognition*, *36*, 256-270.
- Just, M. A., & Carpenter, P. A. (1984). Using eye fixations to study reading comprehension. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 151-182). Hillsdale, NJ: Erlbaum.
- Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In R. M. Church & B. A. Campbell (Eds.), *Punishment and aversive behavior* (pp. 279-296). New York: Appleton-Century Crofts.
- Kaplan, A. S., & Murphy, G. L. (1999). The acquisition of category structure in unsupervised learning. *Memory & Cognition*, *27*, 699-712.
- Kaplan, A. S., & Murphy, G. L. (2000). Category learning with minimal prior knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 829-846.
- Keil, F. C. (1981). Constraints on knowledge and cognitive development. *Psychological Review*, *88*, 197-226.

- Krascum, R. M. & Andrews, S. (1998) The effects of theories on children's acquisition of family resemblance categories. *Child Development*, 69, 333-346.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22-44.
- Kruschke, J. K. (1993). Three principles for models of category learning. In G. V. Nakamura, R. Taraban, & D. L. Medin (Eds.), *Categorization by humans and machines: The psychology of learning and motivation* (Vol. 29, pp. 57-90). San Diego: Academic Press.
- Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, 45, 812-863.
- Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science*, 12, 171-175.
- Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*, 7, 636-645.
- Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1083-1119.
- Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 830-845.
- Lee, F. J., & Anderson, J. R. (2001). Does learning a complex task have to be complex?: A study in learning decomposition. *Cognitive Psychology*, 42, 267-316.
- Macintosh, N. (1975). A theory of attention: variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276-298.
- Maddox, W. T. (2002). Learning and attention in multidimensional identification and categorization: Separating low-level perceptual processes and high-level decisional processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 99-115.
- Maddox, W. T., Ashby, F. G., & Waldron, E. M. (2002). Multiple attentional systems in perceptual categorization. *Memory & Cognition*, 30, 325-339.

- Maddox, W. T., & Dodd, J. L. (2003). Separating perceptual and decisional attention processes in the identification and categorization of integral-dimension stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 467-480.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207-238.
- Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, *19*, 242-279.
- Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 904-919.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*, 289-316.
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
- Nelson, J. D., & Cottrell, G. W. (2007). Probabilistic model of eye movements in concept formation. *Neurocomputing: An International Journal*, *70*, 2256-2272.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39-57.
- Nosofsky, R. M. (1992). Exemplars, prototypes, and similarity rules. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (pp. 149-167). Hillsdale, NJ: Erlbaum.
- Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*, 53-79.
- Pazzani, M. (1991). The influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *17*, 416-432.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372-422.

- Rehder, B., Colner, R. M., & Hoffman, A. B. (2009). Feature inference learning and eyetracking. *Journal of Memory and Language, 60*, 394-419.
- Rehder, B. & Hoffman, A.B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology, 51*, 1-41.
- Rehder, B., & Hoffman, A. B. (2005b). Thirty-something categorization results explained: Selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 811-829.
- Rehder, B., & Murphy, G. L. (2003). A Knowledge-Resonance (KRES) model of category learning. *Psychonomic Bulletin & Review, 10*, 759-784.
- Rehder, B. & Ross, B. H. (2001). Abstract coherent concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1261-1275.
- Rehder, B. & Kim, S. (in press). Classification as diagnostic reasoning. *Memory & Cognition*.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. R. Prokasy (Eds.), *Classical conditioning II*. New York: Appleton-Century-Crofts.
- Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75*, (13, 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.
- Spalding, T. L. & Murphy, G. L. (1996). Effects of background knowledge on category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 525-538.
- Stein, N. L., & Bransford, J. D. (1979). Constraints on effective elaboration: Effects of precision and subject generation. *Journal of Verbal Learning and Verbal Behavior, 18*, 769-772.
- Sutherland, N. S., & Mackintosh, N. J. (1971). Mechanisms of animal discrimination learning. New York: Academic Press.

Watson, M., & Blair, M. (2008). *Attentional allocation during feedback: Eyetracking adventures on the other side of the response*. Paper presented at the Proceedings of the 28th Annual Conference of the Cognitive Science Conference.

Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 449-468.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science*, *18*, 221-281.

APPENDIX

Materials for Experiments 1 and 2

The features and their associated knowledge are presented. There were 12 features, two features for each of the six dimensions. For each feature, three types of knowledge (i.e., Tundra, Desert, & Neutral) were invented resulting in 36 descriptions in total.

Antenna (0)



- 0_t: The ants need to photosynthesize amino acids to sustain their life. Because the daytime is short, they use this fan type of antennae to maximize the surface area exposed to the sunlight.
- 0_d: Because the air is hot and dry, the ants are vulnerable to dehydration. To maintain hydration, the ants use this fan type of antennae to absorb water vapor from the air.
- 0_n: The ants use sunlight and moonlight to orient themselves. This fan-like antennae allows to absorb enough light for purpose of orientation.

Antenna (1)



- 1_t: Because the temperature is very low, parts of ants' eyes (e.g., cornea, iris, pupil) often freeze and the ants become blind. When that happens, this thread type of flexible antennae is used to detect close objects.
- 1_d: Because the temperature is very high, the ants need to dissipate excess body heat. This thread type of antennae promotes heat dissipation.
- 1_n: The ants become blind at night and use this thread type of antennae to detect close objects.

Foot (0)



- 0_t: Because the ground surface is extremely cold, the ants conserve body heat by switching the toe that comes into contact with the ground in each step.
- 0_d: Because the ground surface is extremely hot, the ants switch the toe that comes into contact with the ground in each step to avoid burning.
- 0_n: The ants communicate through chemicals called pheromones. Each of these three toes releases a unique chemical to convey different messages.

Foot (1)



- 1_t: Because the ground surface is slippery, the ants need to have wide feet to maintain their footing.
- 1_d: Because of the sandy soil, the ants have wide feet that prevent them from sinking below the surface.
- 1_n: The ants protect themselves from enemies approaching from behind by kicking with these sharp protrusions.

Forearm (0)



- 0_t: Because the ground is slippery in the ants' environment, the forearm with thorn-like protrusions helps the ants to move without slipping.
- 0_d: Because the ants' prey (e.g., fleas) hide in sand, the ants use this type of forearm to sweep the sand and detect the prey.
- 0_n: When the ants engage in a fight, they use this saw-like forearm to tear the enemy apart.

Forearm (1)



- 1_t: Because of frequent blizzards, the ants need to anchor themselves during high winds. This type of forearm allows the ant to hold its position.
- 1_d: Because of strong direct sunlight, the ants dig deep into the ground in order to be cool when they rest. This type of forearm helps the ants to do the job easier and faster.
- 1_n: The ants sometimes plunder other ants' colony of eggs. This hook-type of forearm is useful in digging in search of the eggs.

Mouth (0)



- 0_t: Because sources of food are frozen and tough, the ants mash and grind them using the upper and lower parts of the mouth before swallowing.
- 0_d: Because sources of food are covered with sand, they need to be cleared before swallowing. The inner surface of the ants' mouth has short but stiff hairs that filter out these impurities.
- 0_n: The ants are herbivorous. This long mouth is used to grind tough fibroid materials in plants before they are swallowed.

Mouth (1)



- 1_t: Because the ground is frozen, the ants need to cut and break tough soil in search of their food. This type of mouth with sharp incisors serves this function.
- 1_d: Because the air is dry and the sunlight is strong, food dries out quickly. The ants hold their food in the cavity of their mouth on the way to their nest so that the food does not become dry.
- 1_n: The ants need to transport food to their colony. This mouth allows them to hold the food in the cavity of their mouth until they arrive at their colony.

Tail (0)



- 0_t: Because water tends to exist in a frozen state, the ants acquire water by collecting dew drops with this trumpet-shaped tail early in every morning.
- 0_d: Because the air is dry and water is scarce, the ants need to collect water whenever possible. This trumpet-shaped tail is used to collect rain during the rare rainstorm.
- 0_n: The ants lay a large number of eggs at a time. This trumpet-shaped tail allows the ants to deliver a large number of eggs.

Tail (1)



- 1_i: Because water exists in a frozen state on the surface, the ants have to find melted water beneath the surface. This type of tail probes into places where melted water might exist.
- 1_d: Because the air is dry and water is scarce, the ants store extra water in the humps of their tail.
- 1_n: The ants feed proteins stored in the humps to their larvae using the sharp nozzle in the end of tail.

Wing (0)



- 0_i: Because of low temperature, the wing ends become tattered. The gray area in the wings promotes cell growth by which the damaged wing ends are replaced with new ones.
- 0_d: Because of high temperature, the wing ends become tattered. The gray area in the wings promotes cell growth by which the damaged wing ends are replaced with new ones.
- 0_n: The ants have red spots in the wing ends. The color becomes brighter in the mating season by the hormones produced in the gray area.

Wing (1)



- 1_i: Because hailstones hit the ants' wings and tear them apart, the ends of their wings have thick veins that protect them from being split.
- 1_d: Because sandstorms tear the ants' wings apart, the ends of their wings have thick veins that protect them from being split.
- 1_n: While flying, the ants control their rapid changes in direction by adjusting the fore- and rear-flaps in each wing.

Author Note

ShinWoo Kim and Bob Rehder, Department of Psychology, New York University.

Correspondence concerning this article should be sent to ShinWoo Kim, Department of Psychology, 6 Washington Place, New York, NY 10003 (E-mail: shinwoo.kim@nyu.edu).

This material is based upon work supported by the National Science Foundation under Grant No. 0545298. We thank Gregory L. Murphy and Harlan D. Harris for their comments on an earlier version of this manuscript.

Footnotes

¹ Note that each of these subjects showed strong evidence of knowledge use in both single-feature test results and eye fixations. In fact, we failed to identify any subject who showed evidence of knowledge use in one result but not in the other (e.g., better learning of related dimensions but equal fixation to the related and neutral dimensions). The other 6 subjects (i.e., the *knowledge nonusers*) showed evidence of knowledge use in neither single-feature test results nor eye fixations.

In addition, these knowledge users exhibited greater fixation probabilities to the related dimensions (.77) than the neutral ones (.63) in block 1, $t(13) = 3.86, p < .01$. However, this difference disappeared considering only the first 7 trials (the initial half) of block 1 (.76 vs.71), $t(13) = 1.27, p = .23$. These results suggest rejection of preselection hypothesis (of related dimensions), but show somewhat early effect of knowledge on attention that later grows as a function of exemplar observations.

² Note the small increase and then decrease in the fixation probabilities for the neutral dimensions in the subblock [-1-4] of Figure 6C. This arises because of a relatively homogenous subset of the knowledge users (discussed in the following paragraph) who learned the categories only after a small number of trials.

Table 1

Abstract structure for the Dax and Kez categories.

Exemplars	Dimensions					
	Tail	Foot	Wing	Mouth	Forearm	Antenna
Dax						
D0	1	1	1	1	1	1
D1	1	1	1	1	1	0
D2	1	1	1	1	0	1
D3	1	1	1	0	1	1
D4	1	1	0	1	1	1
D5	1	0	1	1	1	1
D6	0	1	1	1	1	1
Kez						
K0	0	0	0	0	0	0
K1	0	0	0	0	0	1
K2	0	0	0	0	1	0
K3	0	0	0	1	0	0
K4	0	0	1	0	0	0
K5	0	1	0	0	0	0
K6	1	0	0	0	0	0

Table 2

Example feature descriptions for the Dax and Kez prototypes in Figure 1.

Dimension	Dax [Tundra Theme]	Kez [Desert Theme]
Related		
Antenna	Because the temperature is very low, parts of ants' eyes (e.g., cornea, iris, pupil) often freeze and the ants become blind. When that happens, this thread type of flexible antennae is used to detect close objects.	Because the air is hot and dry, the ants are vulnerable to dehydration. To maintain hydration, the ants use this fan type of antennae to absorb water vapor from the air.
Mouth	Because the ground is frozen, the ants need to cut and break tough soil in search of their food. This type of mouth with sharp incisors serves this function.	Because sources of food are covered with sand, they need to be cleared before swallowing. The inner surface of the ants' mouth has short but stiff hairs that filter out these impurities.
Forearm	Because of frequent blizzards, the ants need to anchor themselves during high winds. This type of forearm allows the ant to hold its position.	Because the ants' preys (e.g., fleas) hide in sand, the ants use this type of forearm to sweep the sand and detect the prey.
Foot	Because the ground surface is slippery, the ants need to have wide feet to maintain their footing.	Because the ground surface is extremely hot, the ants switch the toe that comes into contact with the ground in each step to avoid burning.
Neutral		
Tail	The ants feed proteins stored in the humps to their larvae using the sharp nozzle in the end of tail.	The ants lay a large number of eggs at a time. This trumpet-shaped tail allows the ants to deliver a large number of eggs.
Wings	While flying, the ants control their rapid changes in direction by adjusting the fore- and rear-flaps in each wing.	The ants have red spots in the wing ends. The color becomes brighter in the mating season by the hormones produced in the gray area.

Table 3

Single-feature test results from Experiments 1 and 2 (learners only).

	Related Condition		Unrelated Condition
	Related Dimensions	Neutral Dimensions	Neutral Dimensions
Experiment 1			
Accuracy	.89	.71	.76
Signed confidence rating	67.1	37.0	43.8
Experiment 2			
Accuracy	.91	.70	
Signed confidence rating	73.6	29.1	

Figure 1

An example of the two categories' prototypes.

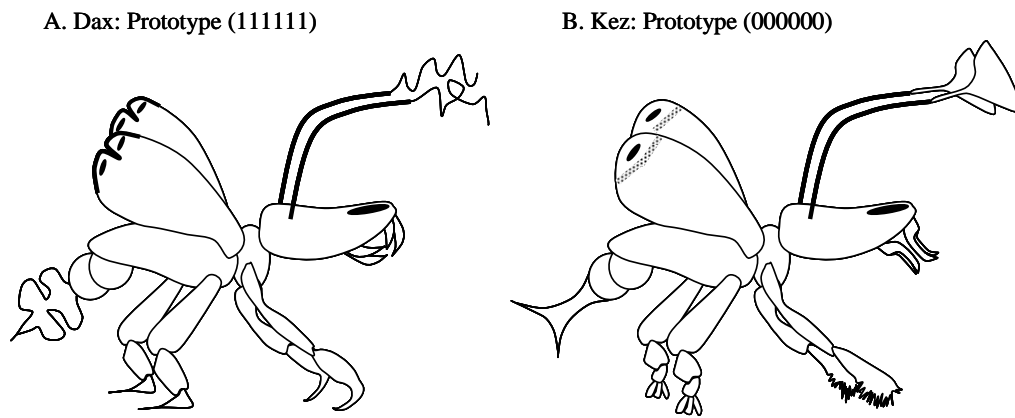
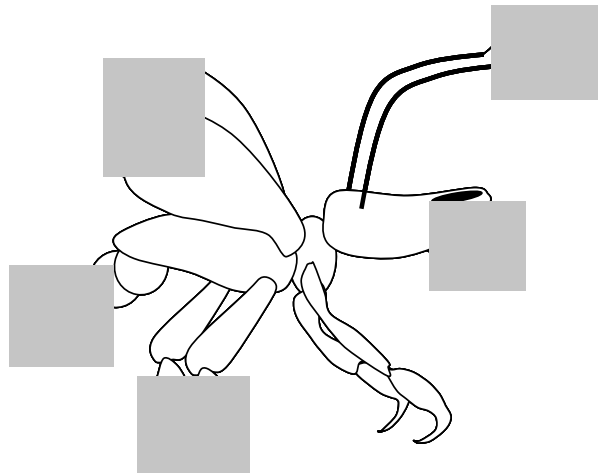


Figure 2

An example of a multiple-choice question.



Q. Which of the followings is correct?

- a. Because of frequent blizzards, the ants need to anchor themselves during high winds. This type of forearm allows the ant to hold its position. (1_v)
- b. Because of strong direct sunlight, the ants dig deep into the ground in order to be cool when they rest. This type of forearm helps the ants to do the job easier and faster. (1_d)
- c. The ants sometimes plunder other ants' colony of eggs. This hook-type of forearm is useful in digging in search of the eggs. (1_n)
- d. Because the ground is slippery in the ants' environment, the forearm with thorn-like protrusions helps the ants to move without slipping. (0_i)

Question 4 of 12

Note. Information within parentheses was not presented to participants.

Figure 3

Average error probabilities for the three types of training items in the related condition of Experiment 1.

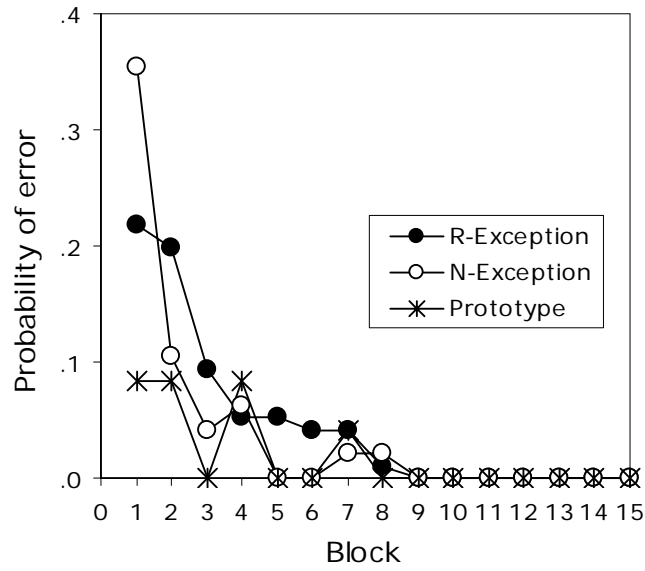


Figure 4

Average error probabilities for the three types of training items in Experiment 2.

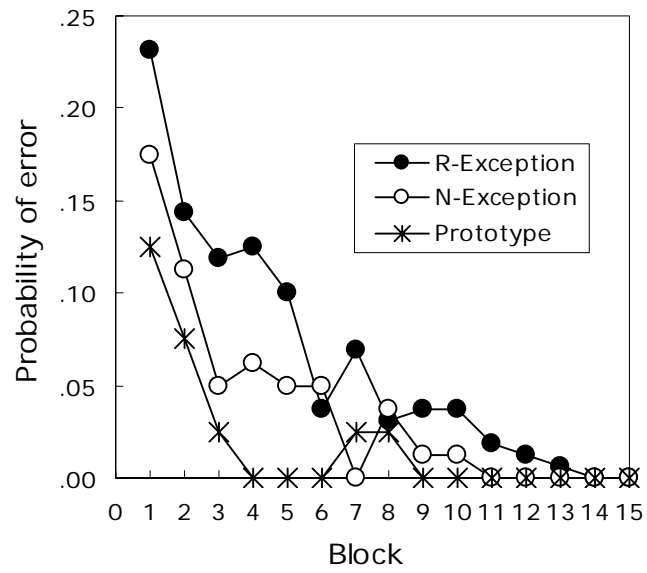


Figure 5

Eye-fixation data from Experiment 2. (A) Number of related and neutral dimensions fixated in each block. (B) Probability of fixation to the two dimension types in each block. (C) Proportion of fixation number/time and relative priority in each block.

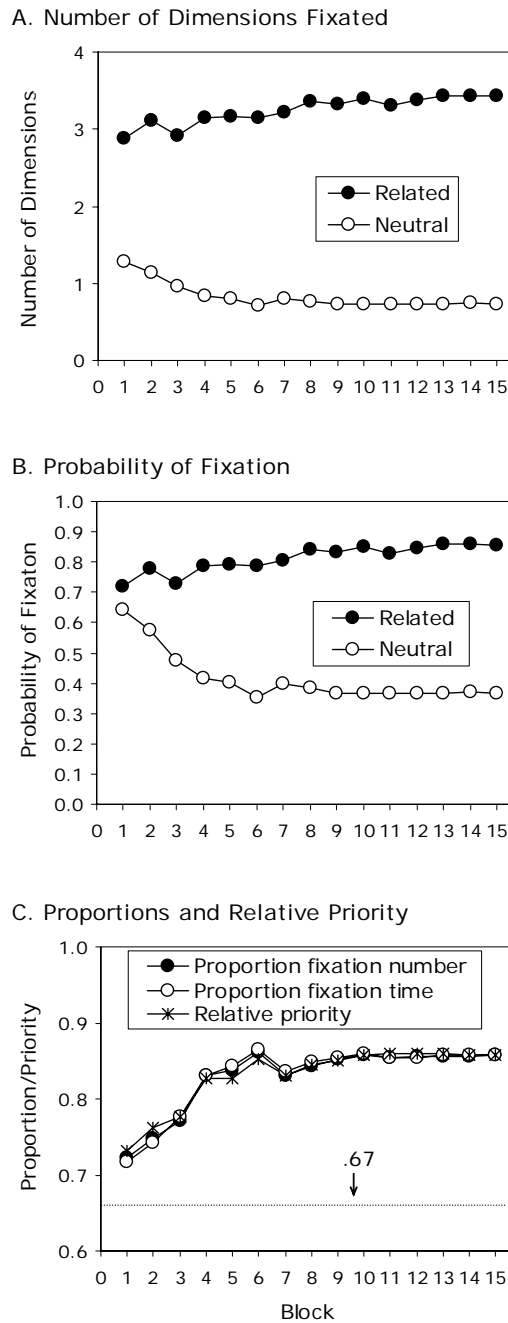


Figure 6

Backward learning curves of the knowledge users (n = 14). Each data point (subblock) presents performance averaged over four trials.

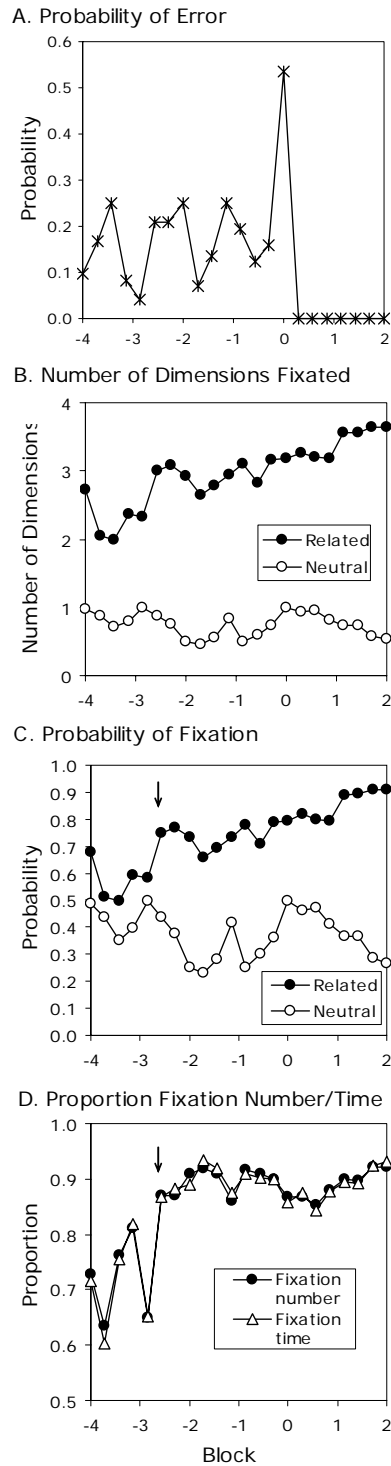


Figure 7

Backward learning curves of the knowledge nonusers (n = 6). Each data point (subblock) presents performance averaged over four trials.

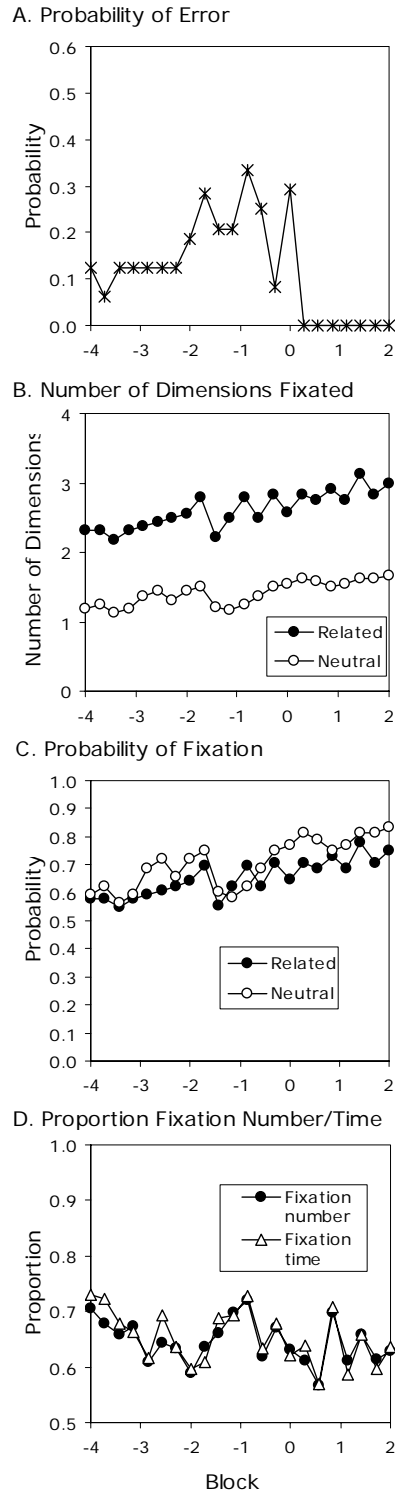


Figure 8

Backward learning curves of a subset of the knowledge users (n = 6). Each data point (subblock) presents performance averaged over four trials.

