



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Feature inference learning and eyetracking

Bob Rehder^{a,*}, Robert M. Colner^a, Aaron B. Hoffman^b^a Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, United States^b Department of Psychology, University of Texas at Austin, Austin, TX, United States

ARTICLE INFO

Article history:

Received 30 July 2008

Revision received 9 December 2008

Available online 6 February 2009

Keywords:

Category learning

Category representation

Eyetracking

ABSTRACT

Besides traditional supervised classification learning, people can learn categories by inferring the missing features of category members. It has been proposed that feature inference learning promotes learning a category's internal structure (e.g., its typical features and interfeature correlations) whereas classification promotes the learning of diagnostic information. We tracked learners' eye movements and found in Experiment 1 that inference learners indeed fixated features that were unnecessary for inferring the missing feature, behavior consistent with acquiring the categories' internal structure. However, Experiments 3 and 4 showed that fixations were generally limited to features that needed to be predicted on *future* trials. We conclude that inference learning induces both supervised and unsupervised learning of category-to-feature associations rather than a general motivation to learn the internal structure of categories.

© 2008 Elsevier Inc. All rights reserved.

When people classify objects, describe concepts verbally, engage in problem solving, or infer missing information, they must access their conceptual knowledge. As a result, the question of how people acquire concepts has been a critical part of understanding how people experience the world and how they interact with it in appropriate ways.

Researchers have developed sophisticated formal theories that explain certain aspects of concept acquisition. These theories are largely based on the study of what has come to be known as *supervised classification learning*—a task that dominates experimental research in this area (Solomon, Median, & Lynch, 1999). In a supervised classification learning experiment, subjects are presented with items whose category membership is unknown; they classify each item and then receive immediate feedback. However, an emerging literature has expanded the range of learning tasks that can be used to inform our models of concept acquisition. By studying different tasks we can understand other aspects of concept acquisition, including the interplay between how categorical knowledge is used and the concept acquired (Brooks, 1978; Chin-Parker &

Ross, 2002; Ross, 2000; Yamauchi, Love, & Markman, 2002; Yamauchi & Markman, 1998, 2000a). For example, investigators have compared supervised classification learning with *feature inference learning* in which learners are presented with an item whose category membership is already identified and asked to infer one of its unknown features. That is, rather than predicting a missing category label on the basis of features, feature inference learners predict a missing feature on the basis of the category label (and perhaps other features). One reason that classification and feature inference learning have been compared is that the two tasks can be equated in a number of ways (Markman & Ross, 2003). Indeed, classification and certain forms of an inference task can be shown to be formally equivalent (Anderson, 1991).

Research comparing classification and feature inference learning has revealed that the two tasks result in the acquisition of different sorts of information about concepts. Whereas it has been established that classification promotes learning the most diagnostic features for determining category membership (Medin, Wattenmaker, & Hampson, 1987; Rehder & Hoffman, 2005a, 2005b; Shepard, Hovland, & Jenkins, 1961; Tversky, 1977), there is evidence that feature inference fosters learning additional category information. For example, Chin-Parker and Ross

* Corresponding author. Fax: +1 212 995 4349.

E-mail address: bob.rehder@nyu.edu (B. Rehder).

(2004) manipulated the diagnosticity and prototypicality of feature dimensions and found that categorization learners were only sensitive to diagnostic features whereas inference learners were also sensitive to nondiagnostic but prototypical features (also see Anderson, Ross, & Chin-Parker, 2002; Lassaline & Murphy, 1996; Sakamoto & Love, 2006; Yamauchi & Markman, 2000b). The task also affects the rate of learning different category structures. Fewer trials are required to learn linearly separable (family-resemblance) category structures via inference than via classification (Yamauchi & Markman, 1998). However, when a comparable non-linearly separable category structure was tested, classification had the learning advantage (Yamauchi et al., 2002).

Differences in how category information is acquired across classification and inference tasks were initially explained in terms of exemplars and prototypes. For example, Yamauchi and Markman (1998) argued that inference learners form representations consistent with prototype models because they seem to extract family-resemblance information such as typical features. In contrast, by focusing on diagnostic information, classification encourages representations consistent with learning rules and exceptions (perhaps via exemplar memorization). However, the prototype interpretation has been challenged by arguments noting the differences between the classification and inference tasks. This debate is worth discussing in detail.

Yamauchi and Markman (1998, Experiment 1) contrasted classification and inference learning with a family resemblance category structure, with four exemplars per category (see Table 1). Each item consisted of a label and four binary feature dimensions. Note that category A and B members were derived from the prototypes 0000 and 1111, respectively, but each member had an *exception feature* from the opposite category's prototype. Participants either classified the eight exemplars into two categories or they predicted a feature missing from every exemplar. To keep the classification and inference tasks as closely matched as possible, inference learners were not presented with *exception feature trials* in which the to-be-predicted feature was from the opposite category. For example, they were never presented with the category A item 000x and asked to predict (on the basis of A1 in Table 1) a '1' for the unknown value *x* on dimension 4. Instead, they were only given *typical feature trials* in which they predicted the category's typical feature (e.g., a '0' for item Ax001). (Presenting only typical feature trials makes the inference task parallel to the classification task in which the predicted category la-

bel is always "typical.") Following learning, all participants completed a transfer test in which participants made inferences on both types of features. They were asked to infer typical features (just as they had during training), as well as exception features (e.g., they predicted *x* in item A000x). Participants were told to respond based on the categories they had just learned and did not receive feedback.

Yamauchi and Markman observed that inference participants reached the learning criterion in fewer blocks than classification participants. Perhaps this should not come as a surprise, because whereas classification required integrating information across multiple feature dimensions, none of which were perfect predictors alone, the inference learners could use the category label as a perfect predictor of missing features.

A second important result concerned how people responded to the exception feature trials during test. Again, strict adherence to exemplars (see Table 1) requires one to predict a value typical of the opposite category (e.g., predict *x* = 1 for A000x). In contrast, responding on the basis of the category prototype means responding with a typical feature. In fact, Yamauchi and Markman found that inference learners responded with the category's typical feature far more often than did the classification learners, suggesting that classification learners more often based inferences on the training exemplars whereas inference learners based theirs on the prototype. This result, coupled with formal model fits (although see Johansen & Kruschke, 2005), led Yamauchi and Markman to conclude that inference learners represent prototypes and classification learners represent exemplars.

Subsequent investigations have expanded on this conclusion by demonstrating that the inference task fosters acquisition of category information other than just a prototype, at least if a prototype is construed as a single, most representative category member. For example, Chin-Parker and Ross (2002) demonstrated that inference learning results in greater sensitivity to within-category correlations than classification learning, even when those correlations are not necessary for correct classification (also see Anderson & Fincham, 1996). The inference task also facilitates the acquisition of abstract commonalities shared by category members. For example, Rehder and Ross (2001) found that, as compared to classification, inference learners more readily learned categories defined by interfeature semantic relationships common to multiple category members (also see Erickson, Chin-Parker, & Ross, 2005; Yamauchi & Markman, 2000b). These findings are important because they suggest that inference learning promotes acquisition of not only a prototype but also of the category's internal structure (including interfeature correlations and semantic relations) more generally. Together with those cited earlier, these results led Markman and Ross (2003) to suggest that inference learners are "trying to find aspects of categories that facilitate the prediction of missing features" leading them to "pay particular attention to relationships among category members and often compare category members" (p. 598). They propose, "because [inference] learners are focused on the target category...the task may be viewed (from the learner's perspective) as figuring out what the category's members are like, that is, the inter-

Table 1
Yamauchi and Markman (1998) category structure.

Exemplar	Category label	D1	D2	D3	D4
A1	A	0	0	0	1
A2	A	0	0	1	0
A3	A	0	1	0	0
A4	A	1	0	0	0
B1	B	1	1	1	0
B2	B	1	1	0	1
B3	B	1	0	1	1
B4	B	0	1	1	1

nal structure of the category” (p. 598). We'll refer to the proposal that inference learning induces in learners a goal to acquire the categories' internal structure as the *category-centered learning hypothesis* (CCL).

There are, however, alternative interpretations of some of these data. For example, Johansen and Kruschke (2005) proposed that, rather than prototypes, inference learners in Yamauchi and Markman (1998) and Chin-Parker and Ross (2004) acquired category-to-feature rules instead. This *set-of-rules model* is viable because inference learners were never presented with exception-feature trials during training. As a result, they could succeed simply by learning associations (rules) relating the categories' labels to their typical features. The classification learners, in contrast, were forced to either learn an imperfect rule with exceptions or memorize exemplars. Of course, at first glance CLL and the set-of-rules model may seem to be equivalent, because they both predict that in many circumstances (e.g., those in Chin-Parker & Ross, 2004; Yamauchi & Markman, 1998) people will infer typical values for missing features. However, rather than invariably encoding the category's prototype, the set-of-rules model predicts that which rules are learned depends on the inferences made during training, inferences that need not be restricted to typical features. For example, Johansen and Kruschke compared a condition in which learners only made typical feature inferences (just as in Yamauchi & Markman) with one in which they only inferred exception features. Whereas at test participants in the former condition inferred typical features, those in the latter inferred exception features, that is, they responded on the basis of category-to-feature associations required during training rather than the categories' prototype (see Nilsson & Olsson, 2005; Sweller, Hayes, & Newell, 2006 for related evidence). Note that the set-of-rules model does not explain the learning of within-category correlations (Chin-Parker & Ross, 2002) or other sorts of abstract commonalities (Erickson et al., 2005; Rehder & Ross, 2001)—points we return to in the General Discussion.

The goal of our experiments is to distinguish between CCL and the alternative category-to-feature rule hypothesis using an eyetracker. The gathering of eye movement data as another dependent variable is useful because the two hypotheses make distinct predictions regarding the allocation of attention during inference learning. Recall that, according to the CCL hypothesis, inference learners' motivation to learn the internal structure of categories—to learn what categories are “like”—leads to the comparison of category members and the learning of interfeature correlations and other abstract commonalities (Markman & Ross, 2003). These processes require that, during a feature inference trial, learners attend to many if not most of the item's features. For example, given the inference problem Ax001, CCL predicts that while predicting the missing value (x) on dimension 1 learners will frequently attend to not only the category label (A) but also to the other features on dimensions 2–4 (001). It does so because learning interfeature correlations requires attending to at least one of these non-missing features (enabling the encoding of the correlation between it and the predicted feature); comparing the item with other category members

remembered from previous trials (enabling the extraction of commonalities across category members) requires attending to most or all of those features. The following experiments track eye movements in order to test for the presence of this pattern of attention allocation during feature inference learning.

In contrast, the most straightforward prediction derivable from the rule-based hypothesis is that inference learners will fixate only the category label as they attempt to learn the correct rule between it and the to-be-predicted feature dimension. On this account, for the inference problem Ax001 learners will attend to the category label but not dimensions 2–4, because those dimensions are irrelevant to the rule relating the category label and the first dimension. Of course, it should be clear that this prediction rests on some assumptions, namely that learners: (a) are motivated to perform as efficiently as possible, and thus will attend only to information needed to make the inference (i.e., the antecedent of the rule, the category label) and (b) do not have some other reason to attend to the non-queried feature dimensions (e.g., dimensions 2–4 in the inference problem Ax001). Regarding the first assumption at least, previous research has shown that rule-based learners can in fact limit their eye movements to only rule-relevant information. For example, Rehder and Hoffman (2005a) had participants perform a supervised classification task (the Shepard et al., 1961, Type I problem) in which one dimension was perfectly predictive of category membership, that is, it could be solved with a one-dimensional rule. They found that eye movements were quickly allocated exclusively to the single diagnostic dimension, indicating that subjects can readily attend to only that information needed to solve a learning problem via a rule. Regarding the second assumption, later in this article we will consider reasons why feature inference learners might fixate non-queried feature dimensions even when they are learning category-to-feature rules.

In Experiment 1, we start by replicating the classification and inference conditions of the seminal Yamauchi and Markman (1998) study using an eyetracker. Note that eyetracking has proven to be an effective tool in many areas of research, reflecting the close relationship between gaze and cognitive processing (see Liversedge & Findlay, 2000; Rayner, 1998 for reviews). Although it is well known that attention can dissociate from eye gaze under certain circumstances (Posner, 1980), in many cases changes in attention are immediately followed by the corresponding eye movements (e.g., Kowler, Anderson, Doshier, & Blaser, 1995). Indeed, Shepherd, Findlay, and Hockey (1986) have demonstrated that although attending without making corresponding eye movements is possible, it is not possible to make an eye movement without shifting attention, and there is evidence that attention and eye movements are tightly coupled for all but the simplest stimuli (Deubel & Schneider, 1996). Especially relevant to the present research is that eyetracking has been used successfully in learning studies (e.g., Kruschke, Kappenman, & Hetrick, 2005) including those involving category learning (Rehder & Hoffman, 2005a; Rehder & Hoffman, 2005b). Note that to further strengthen the relationship between eye gaze and cognitive processing, the following experiments go beyond

these previous learning studies by using *gaze contingent displays* in which stimulus information (i.e., a feature) is only displayed on the screen when the learner fixates that screen position, a technique that rules out the use of peripheral vision.

Experiment 1's use of eyetracking will thus provide new evidence regarding the competing attentional claims of the CCL and the set-of-rules model. With these initial results in hand, Experiments 2–4 will focus exclusively on inference learning by using eyetracking to test a number of additional hypotheses regarding what information such learners use to successfully complete the feature inference task. To foreshadow the results, we will argue that our eyetracking results support *neither* CLL nor the set-of-rules model. Instead, we will offer a new account of the feature inference task—the *anticipatory learning account*—the postulates that inference learners are not generally motivated to learn what categories are “like” but rather attend to that category information that they think will be needed in the future, and by so doing incidentally learn many additional aspects of a category's structure.

Experiment 1

Method

Participants

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

Materials

The category structure was identical to that used by Yamauchi and Markman (1998) (Table 1). The stimuli were designed to facilitate the recording of eye movements with dimensions separated in space. The dimensions included the category label (“A” or “B”) and four feature dimensions with pairs of feature values that were pretested for nearly equal discrimination times. The category label and features were equidistant from the center of the display. Five lines originated at the center of the display and terminated at each dimension's screen location. Examples of category A and B prototypes are presented in Fig. 1. The SensoMotoric Instruments (Berlin, Germany) Eyelink I system was used to record eye movements.

Design

Participants were randomly assigned in equal numbers to the inference learning and classification learning conditions. They were also randomly assigned to one of five layouts of physical stimulus dimensions to screen locations, so that category labels and features appeared an approximately equal number of times in each of the five screen positions. For a given subject the position of the category label and features remained fixed throughout the experiment.

Procedure

The design and procedure replicated the classification and inference conditions of Yamauchi and Markman (1998). (We omitted their mixed condition that interleaved both classification and inference learning trials.) Before

each trial, participants fixated a dot in the center of the display while a drift correction was performed to compensate for shifts in the position of the eyetracker on the participant's head. Immediately following the drift correction the stimulus appeared. We used a gaze-contingent display such that a feature only became visible when it was fixated. To minimize the chance of that the subject would detect any change in the display in their peripheral vision, when not fixated the screen location displayed that dimension's two possible feature values superimposed on one another. The gaze-contingent display eliminates the possibility that learners can use their peripheral vision to obtain stimulus information, thereby ensuring that the eye movements are a reliable indicator of what information learners extract from the display. The feature values were superimposed in order to reduce the chance that subjects would notice the change in the display in their peripheral vision. Note that participants were told that the display was operating in a way that prevented use of their peripheral vision and were instructed to inform the experimenter if the object's features ever became ambiguous. In this case the trial would be restarted after recalibrating the eyetracker.

The experiment had a learning phase followed by a test phase. In the learning phase all participants responded until they reached a learning criterion of 90% accuracy on three consecutive blocks, or until they completed 30 blocks. Participants in the inference learning condition were presented with stimulus items with an intact category label but with one missing (queried) feature. An example inference trial is presented in Fig. 2A. A dashed line that extended from the fixation point to the queried location obviated the need for learners to fixate stimulus dimensions in order to determine which dimension was being queried. The queried location displayed the two possible values for that dimension presented side by side, with their left–right positions randomized on each trial. For classification learners, the two possible category labels were presented side by side, also in a random left–right position (Fig. 2B). A dashed line also indicated the location of the category label, even though that location remained constant throughout the experiment for each classification participant.

Participants used the left or right arrow key to select the correct response. Following a response, the correct feature or category label would appear in the queried location. A tone signaled a correct or incorrect response. The completed stimulus item remained on the screen for 3 s after feedback.

Following Yamauchi and Markman (1998), in the classification condition each exemplar in Table 1 was presented once per block. Regarding the inference condition, note that 32 unique inference trials can be derived from those exemplars (one for each dimension for each exemplars). However, because the eight exception feature trials were not presented during training, the remaining 24 typical feature trials were each presented once over three blocks subject to the constraint that each dimension was queried once for each category in an eight-trial block. In both conditions the presentation order of learning trials was randomized within each block.

The test phase was identical for all participants and included classification trials followed by inference trials. Participants first made 10 classification judgments on the

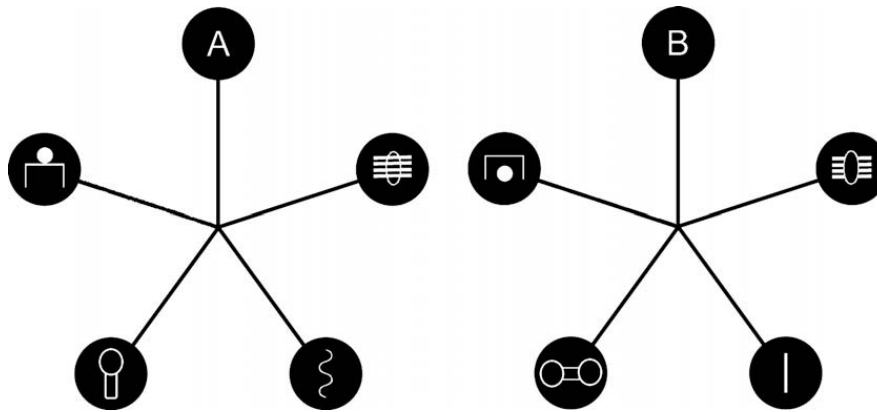


Fig. 1. Examples of category prototypes. Top features (“A” and “B”) are category labels.

eight exemplars from the learning phase and the two category prototypes. They then made 32 feature inferences in which they inferred every feature of every exemplar (including both typical and exception features). The presentation order of trials was randomized within each test phase. Feedback was not provided during the test phase.

Eyetracking dependent variables

Eye movement data were analyzed by defining five circular areas of interest (AOIs) around the physical locations of the five dimensions on the display. Fixations to locations outside of the AOIs were discarded before computing dependent measures. *Probability of fixation* was calculated as a binary measure of whether a dimension’s AOI was fixated at least once on a trial. Averaging this binary variable over multiple trials yields the probability of a dimension being observed during those trials. In addition, the duration of each fixation was recorded and the *proportion fixation time* was calculated as the total fixation time on a dimension divided by the total fixation time to all dimensions.

Results

We analyzed data from the 38 participants (19 per condition) who reached the learning criterion. Replicating the result from Yamauchi and Markman, inference participants

required fewer learning blocks to reach criterion ($M = 7.9$, $SE = 0.92$) than classification participants ($M = 13.2$, $SE = 1.79$), $t(36) = -2.64$, $p < 0.01$.

We next examined participants’ accuracy during the test phase. As expected, performance on the classification test was better when it matched the training task. Classification participants were significantly more accurate ($M = 0.98$) than inference participants ($M = 0.77$) when classifying the eight exemplars from the training phase, $t(36) = -6.55$, $p < 0.001$. Classification accuracy on the two novel prototypes was essentially perfect in both the classification ($M = 0.97$) and inference ($M = 1.0$) conditions, $t < 1$.

Fig. 3 plots the results of the inference test for the classification and inference conditions as a function of whether the test trial was a typical feature trial or an exception feature trial for both Yamauchi and Markman (1998) (Fig. 3A) and the present Experiment 1 (Fig. 3B). On typical feature trials our inference learners were slightly more accurate ($M = 0.93$) than classification learners ($M = 0.90$) although, unlike Yamauchi and Markman, this difference was not significant. Other studies have also found no difference in the tendency of inference and classification learners to infer typical features on typical feature trials (e.g., Sweller et al., 2006).

More importantly, the results for the exception feature trials replicated those of Yamauchi and Markman’s. Recall that on these trials responding in a manner faithful to the exemplars in Table 1 requires inferring a feature typical

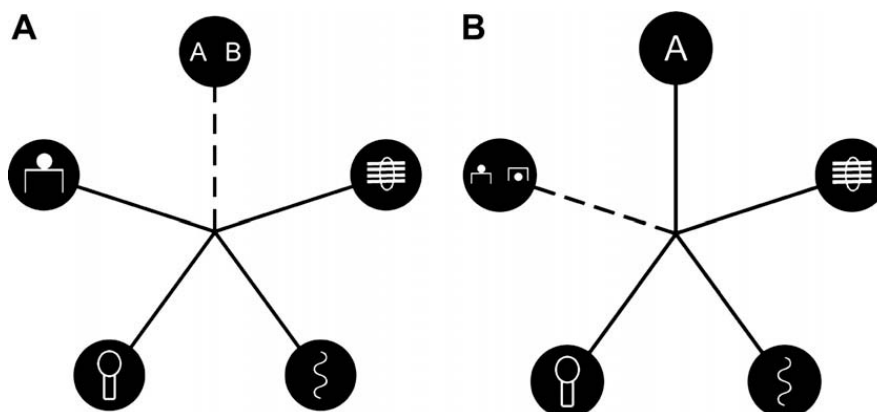
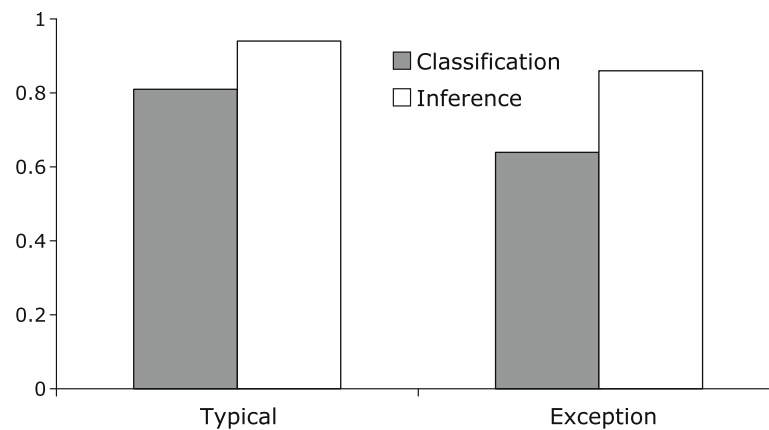


Fig. 2. (A) Example inference trial. (B) Example classification trial. Note that use of a gaze-contingent display meant that only the currently fixated feature was visible.

A. Yamauchi & Markman (1998)



B. Experiment 1

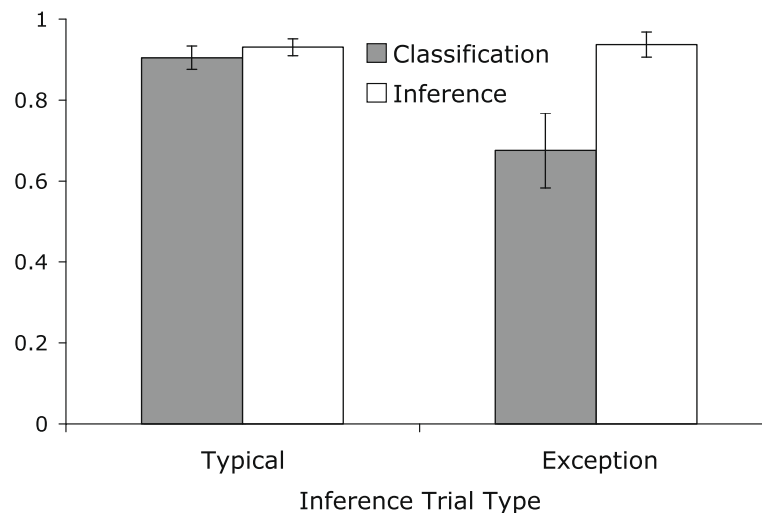


Fig. 3. Inference test results from Experiment 1. (A) Yamauchi and Markman (1998). (B) Experiment 1. Error bars are standard errors of the mean.

of the opposite category. Instead, on exception feature trials inference participants inferred a prototype-consistent feature at the same rate ($M = 0.93$) as they did on typical feature trials. In contrast, participants in the classification condition were far less likely to infer a prototype-consistent feature ($M = 0.67$).

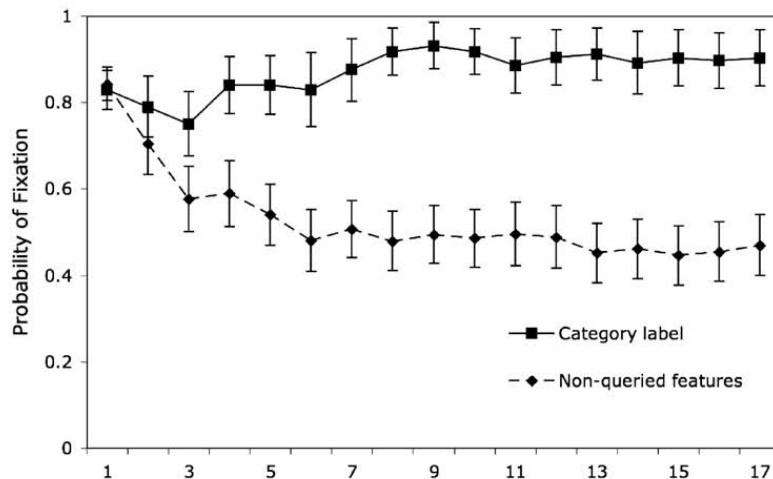
A 2×2 ANOVA with learning condition (categorization vs. inference) as a between-subject factor and trial type (typical vs. exception) as a within-subject factor revealed main effects of condition, $F(1,36) = 7.86$, $MSE = 0.050$, $p < 0.001$, trial type, $F(1,36) = 4.48$, $MSE = .052$, $p < 0.05$, and an interaction between the two, $F(1,36) = 5.05$, $MSE = 0.052$, $p < 0.05$. An analysis of just the exception feature trials confirmed the lower number of prototype-consistent responses in the classification condition than in the inference condition, $t(36) = 2.68$, $p < 0.01$.

We next examined eye movements to understand more about the inference learning process. Were subjects learning a set of category-to-feature rules or were they attempting to acquire the internal structure of the categories? Our first analysis examined the probability of fixation averaged over participants. Fig. 4A plots the probability of observing the cat-

egory label and the average probability of observing a non-queried feature dimension over the course of learning. In this figure, fixations to the queried feature dimension and parts of the screen outside the AOIs have been excluded. For purposes of constructing Fig. 4A, we assumed that those participants who completed training before block 17 (the number of blocks required by the slowest learner) would have continued fixating as they had during their last actual training block.

Fig. 4A shows that the probability that an inference learner fixated a non-queried feature dimension was 0.84 in the first block or, equivalently, they fixated about 2.5 of the three non-queried dimensions on average. Although there was a reduction in the probability of fixating non-queried dimensions during learning (to about 0.48 by the sixth block), it never approached zero even at the end of learning. Note that the probability of observing the category label started off high (0.83 in the first block) and remained high throughout (>0.90 at the end of learning), a result confirming that the category label is the primary basis that inference learners use to predict missing features (Gelman & Markman, 1986; Johansen & Kruschke, 2005; Yamauchi, Kohn, & Yu, 2007; Yamauchi & Markman, 2000a).

A. Probability of fixation



B. Proportion fixation time

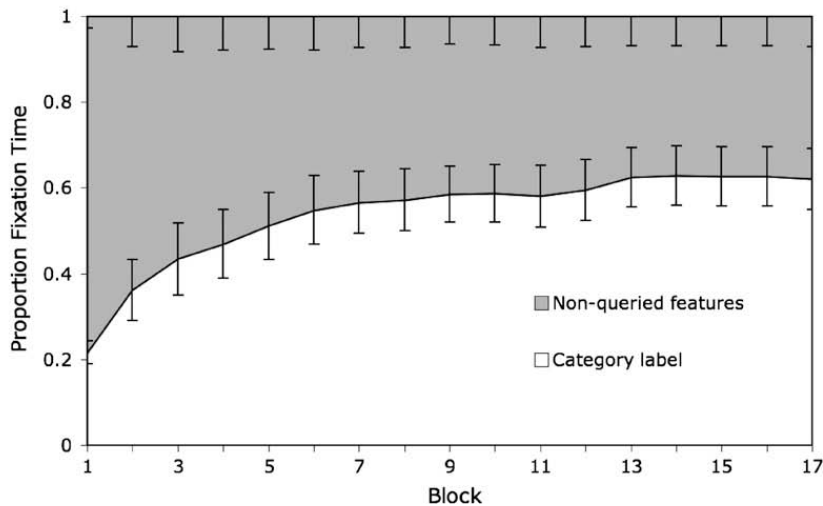


Fig. 4. Eye movements during inference learning in Experiment 1. (A) Probability of fixation. (B) Proportion fixation time. Error bars are standard errors of the mean.

The same qualitative result is seen in Fig. 4B which presents the proportion of fixation time to the category label and the three non-queried features (combined) over the course of learning. (Again, fixations to the queried dimension and irrelevant parts of the screen have been excluded.) On the first block inference learners spent 78% of their time fixating the three non-queried feature dimensions and 22% fixating the category label. Therefore, in fact, they spent a greater proportion of time fixating the average non-queried dimension ($78/3 = 26\%$) than the category label itself. Although proportion fixation time to the non-queried feature dimensions decreased during the course of learning, it remained substantial (40%) even at the end of learning.

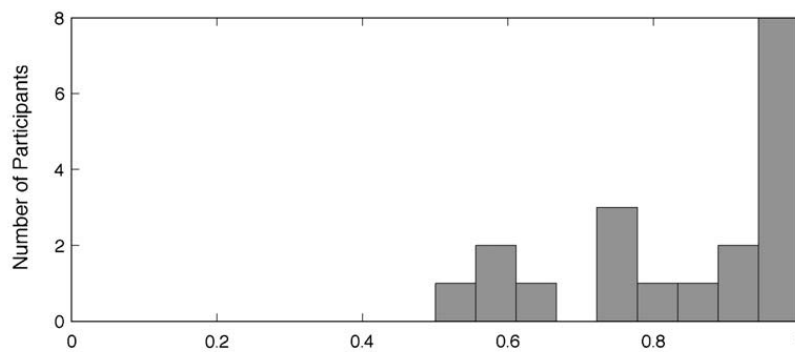
These eyetracking results are important for two reasons. First, they reflect a pattern of attention allocation unlike what has been observed in a supervised classification task that can be solved with a one-dimensional rule. In such tasks, attention is eventually optimized exclusively to the perfectly diagnostic dimension (Rehder & Hoffman, 2005a). Here, participants continued to look at unnecessary dimensions throughout learning. Second, the observed

pattern of attention allocation is unlike what was predicted from the simple rule account. Indeed, learners attended in a way consistent with the notion that they were learning the internal structure of the categories and not (just) learning category-to-feature rules.

A closer analysis of eye movements at the beginning and end of learning revealed substantial differences among participants. The histograms in Fig. 5 present the average probability of fixating a non-queried feature dimension for each individual participant at the beginning and end of learning. Fig. 5A shows that on the first block the majority of participants are fixating most non-queried features. The large majority fixated all three non-queried dimensions on every trial, all but three fixated at least two, and none fixated fewer than one. This pattern of fixations is consistent with the view that most participants are attempting to learn within-category information rather than just category-to-feature rules.

On the last block of learning fixations exhibited a bimodal distribution (Fig. 5B). One cluster of participants averaged about two fixations to non-queried dimensions,

A. First block



B. Last block

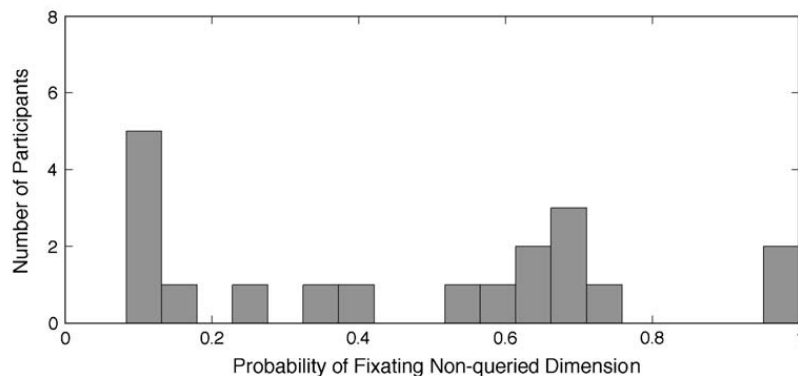


Fig. 5. Histograms of fixations to nonqueried dimensions in Experiment 1's inference test. (A) First block. (B) Last block.

consistent with the possibility that they were still attempting to learn within-category information even after they had reached the learning criterion. The second cluster of participants made very few fixations to non-queried dimensions and thus at the end of training are relying almost entirely on the category label to infer the missing feature. Finally, Fig. 5B shows there were also two subjects that fixated all three non-queried dimensions at the end of learning. Interestingly, these two subjects also never fixated the category label, indicating that they solved the inference problem by integrating the information from the three non-queried dimensions. Examination of Table 1 indicates that one can succeed on valid feature trials by predicting a '0' value for the queried dimension if the three non-queried dimensions have 2 out of 3 '0's and a '1' value if the other three dimensions have 2 out of 3 '1's. Note that although the results in Fig. 4B thus indicate substantial variability in participants' use of non-queried feature dimensions, it remains the case that even at the end of learning most participants were fixating most of those dimensions—a result that the category-to-feature rule account has difficulty explaining.

For completeness, the classification condition's eye tracking results are presented in Fig. 6. Fig. 6 presents the probability of fixating the category label and the average probability of fixating the four feature dimensions as a function of learning block for participants that reached the learning criterion. The figure reveals that learners fixated the (queried) category label on virtually every trial, an

unsurprising result given that it was necessary to examine that screen location to determine whether to respond with the left or right arrow key. In addition, Fig. 6 indicates that participants' probability of fixating a feature dimension was in excess of 0.90 or, equivalently, they fixated over 3½ of the four feature dimensions on average throughout training. The fixations early in learning replicate previous research showing that classification learners begin by fixating most dimensions (Rehder & Hoffman, 2005a, 2005b). Fixations to most dimensions at the end of learning is unsurprising in light of the requirements of the classification task. Unlike inference learners, the classification condition did not have a perfect predictor (i.e. the category label). Instead, diagnostic information from a minimum of three features was required to predict the correct category label.

One final aspect of the eyetracking data is worth noting. In this section we have chosen to report fixations to feature dimensions in a manner that is insensitive to the values displayed on those dimensions. For example, for the inference trial Ax001, feature dimensions 2–3 display typical features ('0's) whereas the fourth displays an exception feature (a '1'). For the classification trial x1000 (which should be classified as an A) dimensions 2–4 display typical features whereas the first displays an exception feature. On both sorts of trials, it is possible that learners devote either more or less attention to the exception features as compared to the typical features. However, because these differences were not directly pertinent to the theoretical

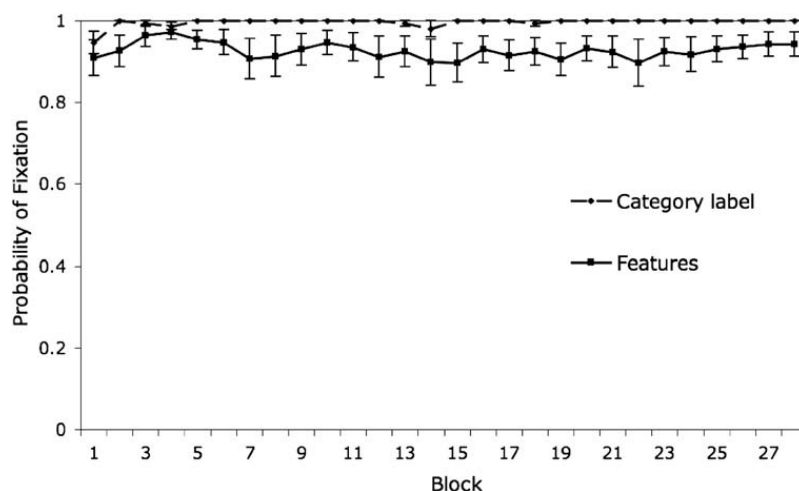


Fig. 6. Eye movement data during classification learning in Experiment 1. Error bars are standard errors of the mean.

issues addressed in this article, we report these data, for all four experiments, in [Appendix A](#).

Discussion

The purpose of Experiment 1 was to differentiate between CCL and the alternative category-to-feature rule hypothesis. If inference learners were in fact acquiring rules, then previous research using eyetracking to investigate the acquisition of a rule (Rehder & Hoffman, 2005a) suggested that they would focus attention on only that information needed to infer the missing feature, namely, the category label itself. In fact, however, the majority of subjects fixated features in addition to the category label throughout learning—behavior consistent with learning the internal structure of the category, including typical features, interfeature correlations, and other abstract commonalities among category members.

Of course, by the end of learning about a third of the inference participants were only fixating the category label and perhaps those participants could appropriately be called “rule learners.” However, the eyetracking data indicated that this group, just like the rest of the participants, was fixating most stimulus dimensions at the start of learning. Perhaps even these participants started off trying to learn what the categories were like but decided by the end of training that they already knew all the “internal structure” there was to learn and thus ceased fixating the non-queried feature dimensions. But regardless of whether one interprets these participants as rule learners or not, it is clear that most inference participants were fixating most feature dimensions during most of learning, behavior that provides tentative evidence against the category-to-feature rule account and for the category-centered learning view.

Experiment 2

Although the results of Experiment 1 thus provide some initial evidence that inference learners were attempting to learn the internal structure of the category

ries, before concluding in favor of CCL it is necessary to consider alternative explanations for the fixations to non-queried feature dimensions. Conceivably, Experiment 1's inference learners were acquiring and applying rules but had other reasons to attend to the non-queried dimensions. For example, one simple possibility is that subjects found the stimuli intrinsically interesting or salient. Note that our prediction for the category-to-feature rule account that inference learners would only fixate the category label was based on another study (Rehder & Hoffman, 2005a) that found few fixations to information irrelevant to a feature-to-category-label rule. However, the stimuli used in that study differed from those in Experiment 1 by having only three dimensions and familiar alphanumeric symbols as features. The stimuli used in Experiment 1 may have been viewed as more complex and interesting by comparison. Thus, to directly confirm that learners would limit attention (eye movements) to only rule-relevant information with the stimuli used in Experiment 1, in Experiment 2 we presented subjects with a classification task that involved discovering a one-dimension rule. If the fixations to other feature dimensions in Experiment 1 were due to the salience of the stimuli then we should also observe many fixations to irrelevant feature dimensions in Experiment 2. In Experiment 3 we will consider yet other reasons for the fixations to the non-queried feature dimensions in Experiment 1.

Method

Participants

A total of 22 New York University undergraduates participated in the experiment for course credit. Two subjects did not complete the experiment in the allotted time and were excluded from the analysis.

Materials

The abstract category structure is depicted in [Table 2](#). Sixteen exemplars were equally divided into two categories. The first dimension (D1) was perfectly correlated with category membership. The remaining three dimensions

Table 2
Category structure tested in Experiment 2.

Exemplar	Category label	D1	D2	D3	D4
A1	A	0	0	0	0
A2	A	0	0	0	1
A3	A	0	0	1	0
A4	A	0	0	1	1
A5	A	0	1	0	0
A6	A	0	1	0	1
A7	A	0	1	1	0
A8	A	0	1	1	1
B1	B	1	0	0	0
B2	B	1	0	0	1
B3	B	1	0	1	0
B4	B	1	0	1	1
B5	B	1	1	0	0
B6	B	1	1	0	1
B7	B	1	1	1	0
B8	B	1	1	1	1

were uncorrelated with the category label. The abstract category structure was instantiated with the same stimuli used in Experiment 1.

Design

Participants were assigned randomly in equal numbers to two counterbalancing factors. As in Experiment 1, a five-level position factor determined the mapping of physical stimulus dimensions to screen locations. In addition, a four-level factor determined which physical dimension was perfectly predictive of category membership.

Procedure

The experimental procedure was similar to the classification condition in Experiment 1. On each trial a dashed line terminated at the location in which both category labels (“A” or “B”) were displayed. Which label appeared on the left and which on the right varied randomly over trials. After each classification judgment participants received auditory feedback and were presented with the complete exemplar including its category label for 3 s.

Each block consisted of the presentation of all 16 exemplars depicted in Table 2. The presentation order of trials was randomized within each block. The experiment ended after a participant completed two consecutive blocks above 93% accuracy or 16 total blocks. Eye movements were recorded throughout the experiment.

Results

The 20 participants who reached the learning criterion did so on average in about four blocks ($M = 4.3$, $SE = 0.36$). Eye movement data was analyzed in the same way as in Experiment 1. AOIs were defined around the location of each feature dimension and the category label and the number and duration of fixations inside each AOI was recorded. The binary observation and proportion fixation time measures were computed as in Experiment 1.

Fig. 7A plots the probability of observing the single diagnostic dimension and the average probability of observing an irrelevant dimension over the course of learn-

ing. Very early on (i.e. by the second block), fixations to the diagnostic and irrelevant dimensions start to diverge. On the last two blocks before the learning criterion was reached, learners were observing the diagnostic dimension on nearly every trial. In contrast, the chance of observing an irrelevant dimension dropped to 0.16 by the end of learning. Fig. 7B, which presents the proportion of time fixating the two types of dimensions, tells a similar story. The eye movements suggest that whereas attention was spread evenly over the four dimensions early in learning, by the end of learning participants allocated nearly 90% of their attention on the single relevant dimension. (Remember that fixations to the queried dimension—in this case the category label—are not included in Fig. 7B.)

An analysis of individual participants revealed that even the group average of 0.16 overestimates the probability that the typical participant fixated irrelevant dimensions at the end of learning. The histogram of fixation probabilities to the irrelevant dimensions on the last block (Fig. 8) reveals two outliers who observed every dimension on every trial. In fact, the remaining 18 subjects were almost completely ignoring the irrelevant dimensions by the time they reached the learning criterion. For these 18 participants, the probability of observing an irrelevant dimension was less than 0.07 by the end of training.

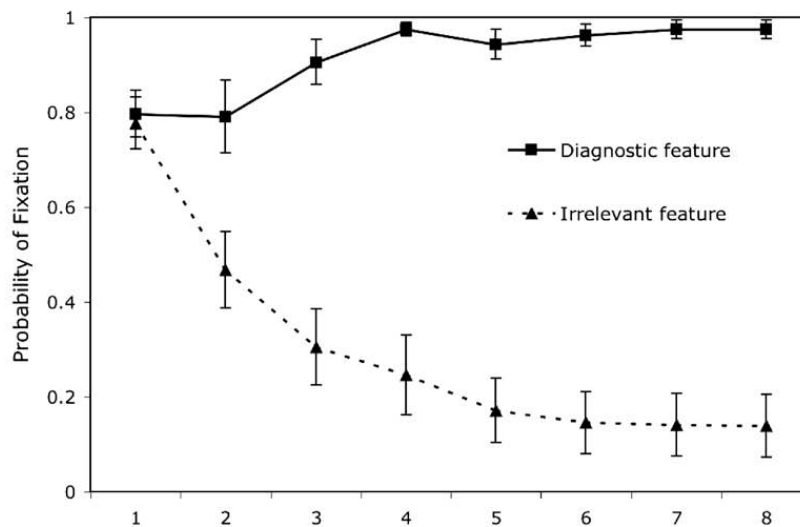
Discussion

The purpose of Experiment 2 was to confirm whether learners would limit their eye movements to only the information relevant to a one-dimensional rule with the same stimuli used in Experiment 1. In fact, by the end of learning fixations to irrelevant feature dimensions were virtually absent for the vast majority of learners, suggesting that the fixations to multiple feature dimensions in Experiment 1 were not due to the intrinsic salience of the stimuli. Note that the results of Experiment 2 replicated those of Rehder and Hoffman (2005a) who used different stimuli and a different category structure. They also found that classification learners began fixating most stimulus dimensions but then quickly learned to attend exclusively to the single perfectly diagnostic dimension.

Experiment 3

Experiment 1 provided evidence that inference learners distribute attention among multiple feature dimensions and Experiment 2 ruled out the possibility that those fixations were due to the intrinsic salience of the stimuli. These results would seem to support the CCL hypothesis, the claim that inference training motivates learners to acquire categories' internal structure. However, there are two other potential explanations of the fixations to the non-queried feature dimensions found in Experiment 1. One is that participants realized they would need to predict those dimensions on later trials. That is, subjects might have been engaged not only in supervised learning of the predicted feature, but also in unsupervised learning of those features that were not being queried on that trial. According to this *anticipatory learning hypothesis*, subjects

A. Probability of fixation



B. Proportion fixation time

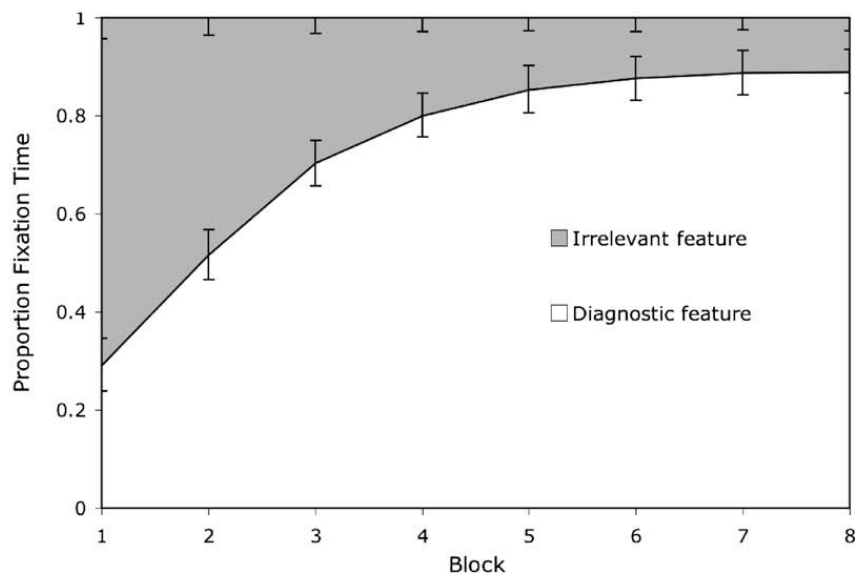


Fig. 7. Eye movements during classification learning in Experiment 2. (A) Probability of fixation. (B) Proportion fixation time. Error bars are standard errors of the mean.

were involved in a more complex learning strategy than predicted by the category-to-feature rule account, as they attempted to learn multiple category-to-feature associations during each trial. Nevertheless, because the primary goal of this strategy is (just like the category-to-feature rule account) to do well at the experimental task, it differs from CCL, that assumes a more general motivation to acquire the internal structure of the categories.

A second possibility is that while the category label was perfectly predictive of each to-be-predicted feature, the three non-queried features, taken together, could also predict those features. Recall that, based on the exemplars in Table 1, one should predict a '0' value for a dimension when there are '0's on two of the other three dimensions and a '1' when there are '1's on two of the three dimensions. Indeed, two Experiment 1 participants

reached criterion without fixating the category label at all, indicating that they used just this strategy. Note that this predictive validity of the non-queried features makes the inference task in Experiment 1 formally different from the classification task in Experiment 2 in which the irrelevant dimensions carried no information about the correct category label. This raises the possibility that even some Experiment 1 subjects who fixated the category label may have also fixated the non-queried features because they provided a second basis for predicting the missing feature. Like the anticipatory learning hypothesis, this *redundant predictor hypothesis* suggests that fixations to non-queried dimensions reflected subjects' desire to optimize performance on the assigned task, in this case by using all sources of relevant information to predict features.

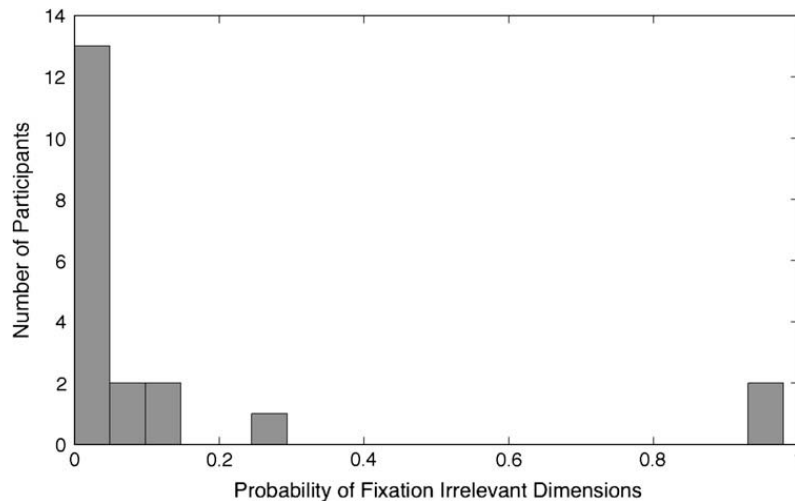


Fig. 8. Histograms of fixations to irrelevant dimensions in Experiment 2.

The purpose of Experiment 3 was to discriminate between the anticipatory learning hypothesis and CCL (and the redundant predictor hypothesis). Following Anderson et al. (2002, Experiment 3), inference learners were trained on the category structure in Table 1 but made inferences on only two of the four feature dimensions. The other two dimensions were presented on each training trial but were never queried throughout training. Participants were then presented with the same tests as in Experiment 1. According to the anticipatory learning hypothesis, on each trial inference learners should focus on the two dimensions that are queried during training (the one being queried on that trial and the one that will be queried on a future trial) and fixations to the two never-queried dimensions should be rare. In contrast, according to CCL inference learners' motivation to learn the internal structure of categories should cause them to fixate most or all of the dimensions on most training trials. This result will also obtain if the other dimensions are being used as a redundant predictor. Classification and inference tests followed training, as in Experiment 1.

Method

Participants and materials

A total of 33 New York University undergraduates participated for course credit. The abstract category structure and its physical instantiation were identical to Experiment 1.

Design

Participants were randomly assigned to one of five configurations of the physical locations of the item dimensions and to one of six possible pairs of feature dimensions to serve as the never-queried features. The classification condition tested in Experiment 1 was omitted in Experiment 3.

Procedure

The procedure was identical to Experiment 1, except that only two of the four possible feature dimensions were queried during the learning phase; thus, only 12 of the 24

typical feature trials presented in Experiment 1 were presented here. Each 8-trial learning block was generated by sampling from those 12 trials subject to the constraint that each of the 8 exemplars in Table 1 was presented once. As in Experiment 1, no exception feature trials were presented.

Results

Only the 30 participants who reached the learning criterion were included in the following analyses. These participants reached criterion in an average of 6.1 blocks ($SE = .13$) as compared to 7.9 blocks in Experiment 1's inference condition. The faster learning in Experiment 3 may be attributable to participants making inferences on only two dimensions versus four. In contrast, performance on the classification test that followed training was slightly worse in Experiment 3 (mean accuracy of 0.73, $SE = 0.06$) than in Experiment 1 (0.77). Conceivably, this lower accuracy was a result of the poorer learning of the never-queried dimensions, and indeed performance on the feature inference test confirms this conjecture. Fig. 9, which present inference test performance as a function of whether the dimension was sometimes- or never-queried and whether it was a typical or exception feature trial, reveals that participants made prototype-consistent responses far more often on sometimes-queried dimensions (92%) as compared to never-queried dimensions (63%). That is, participants exhibited good learning of the typical features on those dimensions that were queried but not those that weren't. Note that, just as in Experiment 1, inference learners predicted prototype consistent features even on exception feature trials even though strict adherence to the exemplars in Table 1 requires predicting the feature from the opposite category on those trials.

A 2×2 repeated measures ANOVA of the inference test results with dimension (sometimes- or never-queried) and trial type (typical or exception) as factors revealed a main effect of dimension, $F(1, 29) = 40.6$, $MSE = 2.21$, $p < 0.0001$, reflecting the greater number of prototype-consistent responses on the sometimes-queried dimensions. Neverthe-

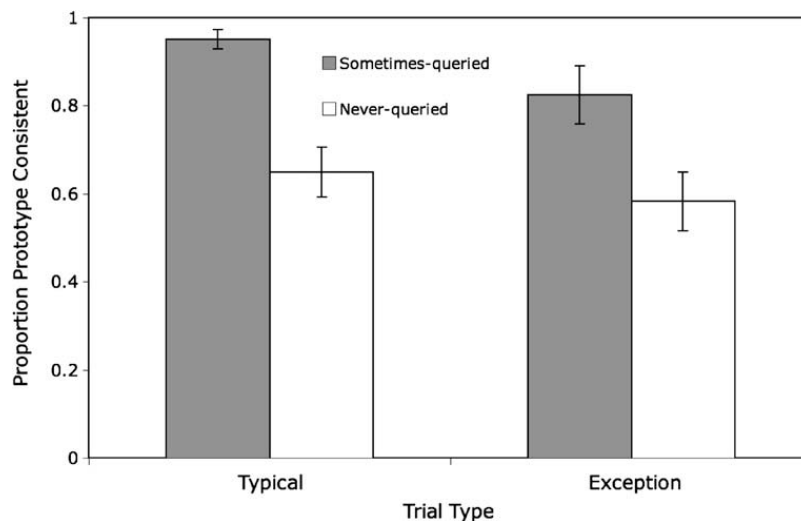


Fig. 9. Inference test results from Experiment 3. Error bars are standard errors of the mean.

less, accuracy on the never-queried features was significantly better than chance, $t(29) = 3.35$, $p < .05$. Somewhat surprisingly, the effect of trial type was significant, $F(1, 29) = 5.58$, $MSE = 0.274$, $p < 0.05$, indicating that prototype-consistent responses were somewhat less likely on exception feature trials than typical feature trials (0.70 vs. 0.80), indicating that, unlike Experiment 1, inference learners acquired some configural (exemplar-based) knowledge about the category structure. In this analysis the interaction was not significant, $F(1, 29) = 1.20$, $MSE = 0.027$, $p > 0.30$.

The key data of course are the eye movements. Did learners' desire to acquire within-category information lead them to fixate most or all feature dimensions or will anticipatory learning lead them to only focus on those dimensions that are sometimes queried during training? The probability of observing sometimes- versus never-queried dimensions is plotted in Fig. 10A. In fact, as early as the first block of learning, participants showed signs of ignoring the never-queried dimensions. At the end of training, inference learners' probability of fixating never-queried dimensions was 0.13 whereas their probability of fixating the sometimes-queried dimension was over four times greater, 0.60. Note that the probability of fixating the never-queried dimensions at the end of training in Experiment 3 was about the same as fixating the irrelevant dimensions in Experiment 2 (0.16). Finally, the histograms in Fig. 11 indicate that even the rare fixations to the never-queried dimensions at the end of training were concentrated in a few participants. Whereas most learners showed substantial probability of fixating the two never-queried dimensions at the start of training (Fig. 11A), at the end (Fig. 11B) only three subjects were fixating those with dimensions with probability greater than 0.5; two-thirds were virtually ignored the never-queried dimensions. Note that the three outliers in Fig. 11B consistently fixated the category label in addition to the two never-queried dimensions, making them unlike the two outliers in Experiment 1 (Fig. 5B) that ignored the category label and used the other feature dimensions to predict the missing one.

The proportion fixation times in Fig. 10B tell a similar story. (As in Experiment 1, fixations to the queried dimension and to screen locations outside the AOIs are omitted.) Whereas fixation times to the never- versus sometimes-queried dimensions are close to equal in the first block, by the end of training, learners were fixating the sometimes-queried dimension 26% of the time and the two never-queried dimensions 8% of the time (4% each). Of course, at the end of training learners also spent over 60% of the time fixating the category label, confirming that the category label was the primary basis for feature inference. Nevertheless, that learners were also spending significant time fixating the sometimes-queried dimension supports the conclusion that they were anticipating that dimension would be queried on an upcoming trial, and so they were trying to learn about it.

Our finding that learners devoted substantial attention to the sometimes-queried dimension but little to the never-queried ones is important for two reasons. First, it indicates that participants had little interest in learning the categories' internal structure above and beyond predicting those dimensions they are being queried on. Second, it speaks against the possibility that the other feature dimensions were being used as a redundant predictor because successfully predicting the missing feature requires attending to all three of those dimensions.¹

Finally, although the eyetracking results presented thus far suggest that subjects had little interest in learning the categories' internal structure, we considered one additional source of evidence—participants' eye movements *after* they responded, while they received feedback.

¹ Note that, due to our method for generating training blocks (see Experiment 3's Procedure section), the sometimes-queried feature dimension considered alone provides no evidence for the missing feature. Specifically, if D_i and D_j are the two queried dimensions, then during training $P(D_i = 1 | D_i = 1) = P(D_i = 1 | \sim D_j = 0) = .50$, indicating that the value on one queried dimension provided no evidence for the other dimension. In other words, although learners generally fixated the sometimes-queried dimension, they didn't use the value they found there to help predict the missing feature.

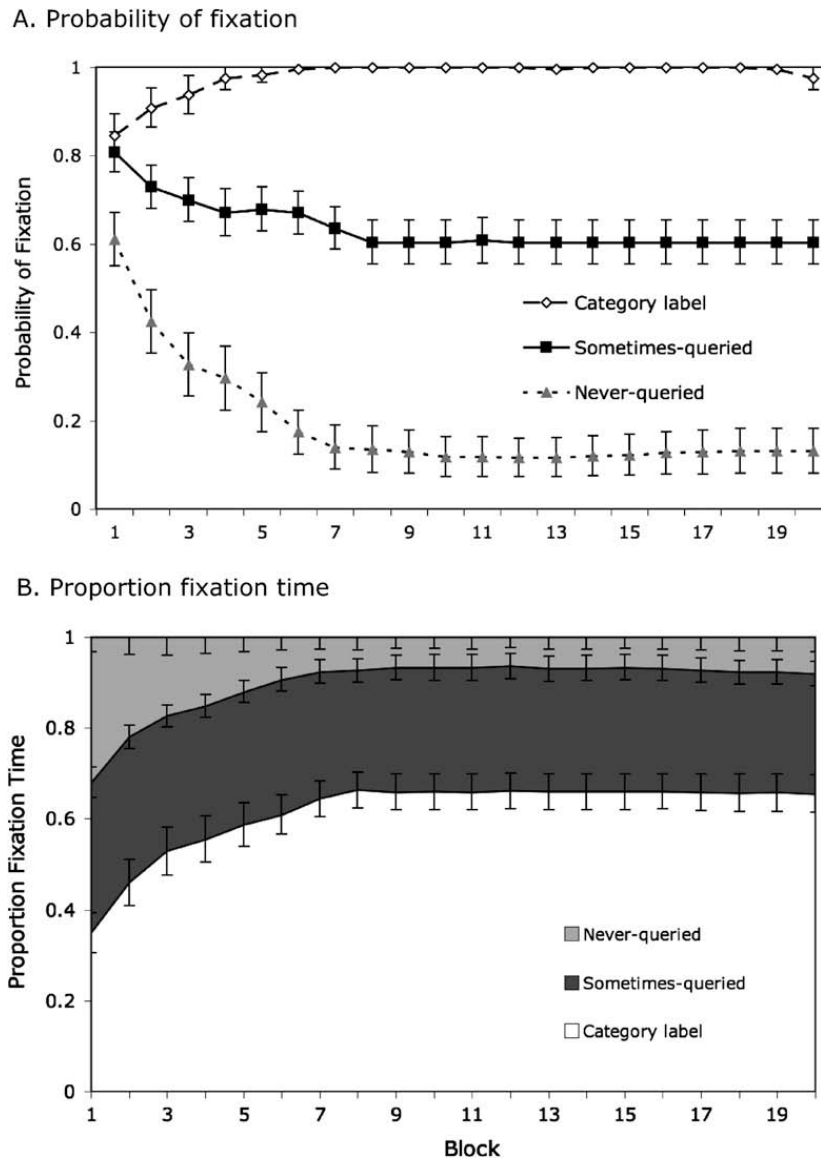


Fig. 10. Eye movements before feedback during inference learning in Experiment 3. (A) Probability of fixation. (B) Proportion fixation time. Error bars are standard errors of the mean.

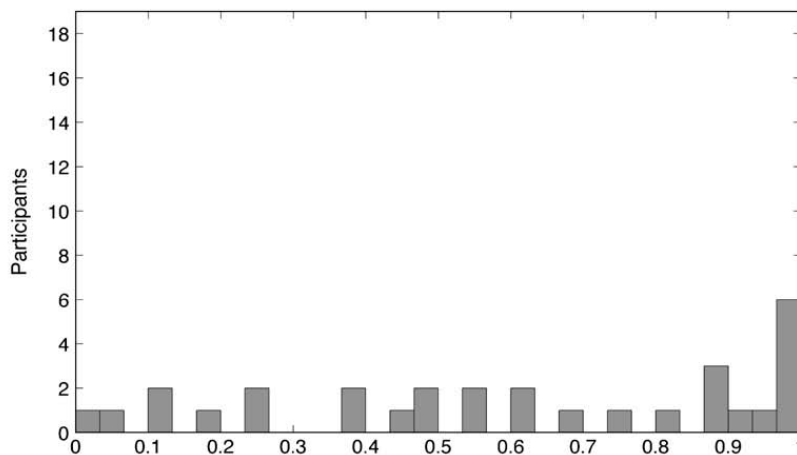
Perhaps participants studied the never-queried dimensions in the three seconds in which the stimulus remained on the screen after they predicted the missing feature. Fig. 12A and B present the fixation probabilities and proportion fixation times, respectively, after participants made their response. Unlike Fig. 10, Fig. 12 includes fixations to the queried dimension as it displayed the correct feature during feedback. First note the large probability of observing and the time they spent fixating the queried dimension, an unsurprising result since at this point that dimension always displayed the correct answer. More importantly, Fig. 12A shows that at the end of learning after feedback rose to 0.27 (from 0.13 before feedback). And, Fig. 12B shows that learners are spending about the same amount of time fixating the sometimes- and never-queried dimensions, consistent with the possibility that they were attempting to acquire information about the categories' internal structure after they responded. Of course, rather than these after-feedback fixations reflecting learning, it

is possible that they merely represent learners' random eye movements as they waited for the mandatory 3 s for the next trial to begin. And these after-feedback fixations to the never-queried dimensions don't change the fact that the learning of those dimensions was poor at best. But the results in Fig. 12 raise the possibility that learning, meager though it was, may have occurred after participants predicted the missing feature. We test this possibility more directly in Experiment 4.

Discussion

Experiment 3 was designed to determine whether the fixations to non-queried feature dimensions in Experiment 1 were better explained by CCL or by anticipatory learning. In fact, Experiment 3 showed many more fixations to the sometimes-queried dimension than the never-queried ones, suggesting that the fixations to non-queried features in Experiment 1 were made for the purpose of making future inferences. Consistent with this interpretation, Exper-

A. First block



B. Last block

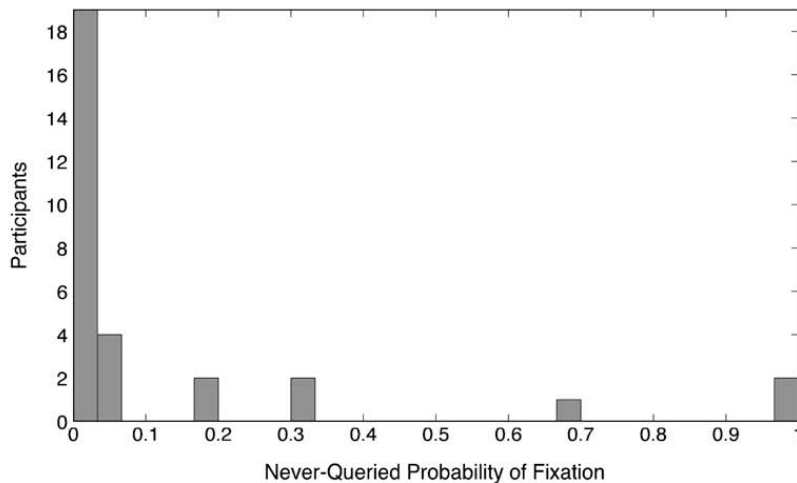


Fig. 11. Histograms of fixations to never-queried dimensions in Experiment 3. (A) First block. (B) Last block.

Experiment 3's inference learners made far fewer prototype-consistent responses on never-queried dimensions than sometimes-queried ones, reflecting the poorer learning of the former. Note that Anderson et al. (2002) and Sakamoto and Love (2006) also found reduced learning of never-queried dimensions (albeit using somewhat different types of transfer tests than Experiment 3).

Nevertheless, although Experiment 3 found reduced learning of the never-queried dimensions, it also found (along with Anderson et al. and Sakamoto and Love) that responding on those dimensions differed from chance, indicating that participants learned at least something about the categories above and beyond what was required by the experimental task. Our eye-tracking data raised the possibility that learning may have occurred at the beginning of training or after participants predicted the missing feature and received feedback, periods when participants were frequently fixating the never-queried dimensions. In Experiment 4 we consider these and other possibilities regarding when the learning of those dimensions occurred.

The absence of fixations to the never-queried dimensions at the end of training also speaks against the possibility that participants were using the non-queried feature

dimensions as a redundant predictor for the queried dimension, because one must fixate all three of the non-queried dimensions in order to infer the correct value for the missing feature.

Experiment 4

One central goal of this article was to use eye tracking to assess the category-learning centered hypothesis, the claim that feature inference learning motivates people to learn the internal structure of categories. Whereas Experiment 1 revealed, in support of CCL, large numbers of fixations to feature dimensions other than the one being queried, Experiment 3 suggests that those fixations were largely due to anticipatory learning rather than subjects' general desire to learn what the categories were like. Despite these results, however, it is possible to defend CCL by noting that the goal of learning the internal structure of categories may not be as important as the one made explicit in the experimental instructions (in Experiment 3, to infer two of the four feature dimensions correctly). Said differently, although feature inference learning may induce

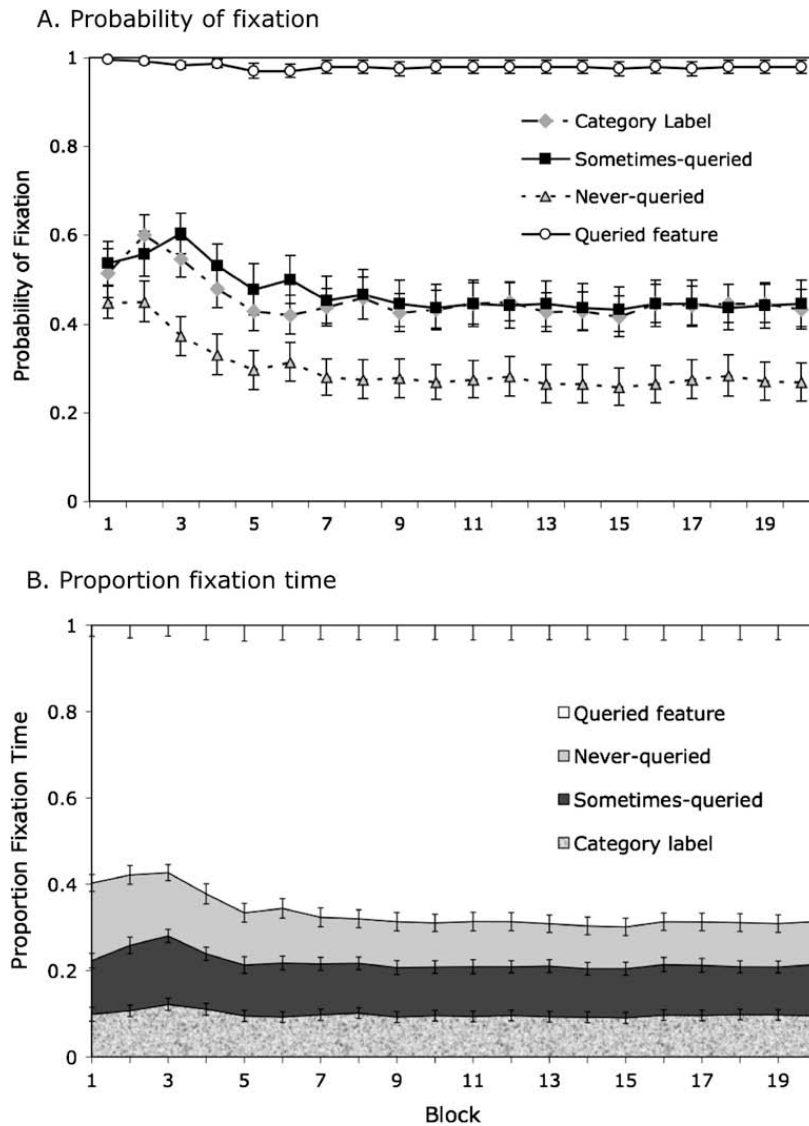


Fig. 12. Eye movements after feedback during inference learning in Experiment 3. (A) Probability of fixation. (B) Proportion fixation time. Error bars are standard errors of the mean.

a goal to learn categories' internal structure, it may be unrealistic to expect that goal to compete on even footing with our participants' more pressing need to complete the experiment with the lowest possible expenditure of time and effort (Markman & Ross, 2003; Payne, Bettman, & Johnson, 1990). In this light, it is notable that there was evidence that Experiment 3's participants were attempting to learn something about the never-queried dimensions. First, although learners were over four times as likely to fixate sometimes- versus never-queried dimensions at the end of learning, fixation times to those two types of dimensions were nearly equal during feedback. And, fixations to the never-queried dimensions were also frequent at the beginning of learning, suggesting that participants may have started off trying to learn the categories' internal structure but that this goal was supplanted by the shorter-term goal of completing the experiment quickly. Finally of course, participants made prototype-consistent responses for the never-queried dimensions with greater than chance probability. These results can be taken together as evi-

dence that participants were devoting at least some cognitive resources to acquiring the categories' internal structure, even aspects of that structure that were not needed to succeed at the experimental task.

Unfortunately, however, each of the sources of evidence for even this weakened version of CCL has an alternative interpretation. For example, although there were numerous after-feedback fixations to the never-queried dimensions, as mentioned those fixations might have just reflected the random eye movements that occurred before the next trial began. On this account, any after-feedback learning of the never-queried dimensions occurred *incidentally*, that is, in the absence of any explicit attempt to acquire the categories' internal structure. Second, Experiment 3 participants frequently fixated the never-queried dimensions at the beginning of learning, but those fixations may reflect uncertainty about which dimension would be queried. On this account, the early fixations to the never-queried dimensions (and thus perhaps the partial learning of those dimensions) were also a result of

anticipatory learning as subjects assumed they would eventually be queried on those dimensions. Subjects ceased fixating those dimensions once they realized they were never queried. Finally, even the above-chance performance on the never-queried dimensions does not necessarily reflect learning that occurred during training, because participants might have acquired that information during the test phase itself. Recall that during the inference test subjects made inferences on all dimensions, including those never queried during training. But once participants realized they would be tested on new dimensions, they may have tried to learn those dimensions on the subsequent test trials. Such learning was possible because each inference test trial presented a whole item, that is, one with values on all dimensions except one. On this account, the learning of the never-queried dimensions was also a result of anticipatory learning, but learning that commenced at the beginning of the test phase rather than during training itself.

These observations raise the possibility that even the reduced fixations to (and learning of) the never-queried dimensions in Experiment 3 might overestimate subjects' commitment to learn the categories' internal structure. To test this possibility, three changes to Experiment 3's experimental procedure were introduced in Experiment 4. First, the duration of the after-feedback stimulus display was placed under control of the participant by having the space bar trigger the start of the next trial. If after-feedback fixations to the never queried dimensions in Experiment 3 reflected only random eye movements while waiting for the next trial, such fixations will be reduced or eliminated in Experiment 4. Second, unlike Experiment 3, in Experiment 4 subjects were told at the beginning of the experiment which two dimensions would be queried. If early fixations to the never queried dimensions in Experiment 3 merely reflected participants' initial ignorance regarding which dimension would be queried, such fixations will also be reduced or eliminated in Experiment 4. Finally, the test phase began with a new single feature classification test in which participants were presented with a single feature and asked which category it belonged to. Participants could not learn about the other feature dimensions because no values were displayed on those dimensions. If the learning of the never-queried dimensions in Experiment 3 occurred at the beginning of test rather than training, such learning will be eliminated in Experiment 4 as measured by the single feature test.

Methods

Participants and materials

A total of 35 New York University undergraduates participated for course credit. The abstract category structure and stimuli were identical to those of Experiments 1 and 3.

Design

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

Procedure

The training inference task was identical to that in Experiment 3 with two exceptions. First, additional instructions defined which two feature dimensions would be queried during the experiment. A mock item with random feature values was presented on the screen with the two dimensions that would be queried circled in red. The participant was informed that "you will only be asked to make judgments about the two circled features, never about the non-circled features." After calibrating the eye-tracker and immediately before the first trial participants were asked to point to the feature dimensions that would be tested on a mock item. After responding the queried features were once again circled in red. The second difference with Experiment 3 was that participants could terminate the after-feedback stimulus presentation and proceed to the next trial by hitting the space bar.

After the learning criterion was reached, participants were presented with three blocks of single-feature classification trials. Each of the eight category features were presented once in each block and participants were asked to respond with the category most strongly associated with that feature. The order of trials within block was randomized. Following the single feature classification test participants were presented with the same whole item classification and inference tests used in Experiments 1 and 3.

Results

Only the 30 participants who reached the learning criterion were included in the following analyses. These participants reached criterion in an average of 4.6 blocks ($SE = 0.33$) as compared to 6.4 blocks in Experiment 3. This faster learning in Experiment 4 may be attributable to informing participants which dimensions would be queried. Interestingly, faster learning occurred despite the fact that subjects devoted less time to studying training items after feedback was provided. Whereas the after-feedback stimulus display was fixed at 3 s in the previous experiments, Experiment 4 learners viewed the stimulus an average of 1.1 s before hitting the space bar to begin the next trial.

The critical test in Experiment 4 concerned performance on the new single feature classification test. Would participants exhibit any learning of the never-queried dimensions since they were told which features would be queried, the duration of the feedback period was self-paced, and there was no possibility of learning during the test itself? Not only were participants far more accurate on the sometime-queried features ($M = 0.98$, $SE = 0.01$) as compared to the never-queried ones, ($M = 0.55$, $SE = 0.03$), $t(29) = 15.7$, $p < 0.0001$, accuracy on the never queried features was not significantly better than chance, $t(29) = 1.43$, $p = 0.16$. That is, participants exhibited no learning of the categories' typical features on the never-queried dimensions.

On the whole-item classification test, participants achieved an accuracy of (0.70, $SE = 0.03$) similar to that in Experiment 3 (0.73). For the feature inference test, Fig. 13 presents the proportion of prototype consistent responses for sometimes- and never-queried features and for

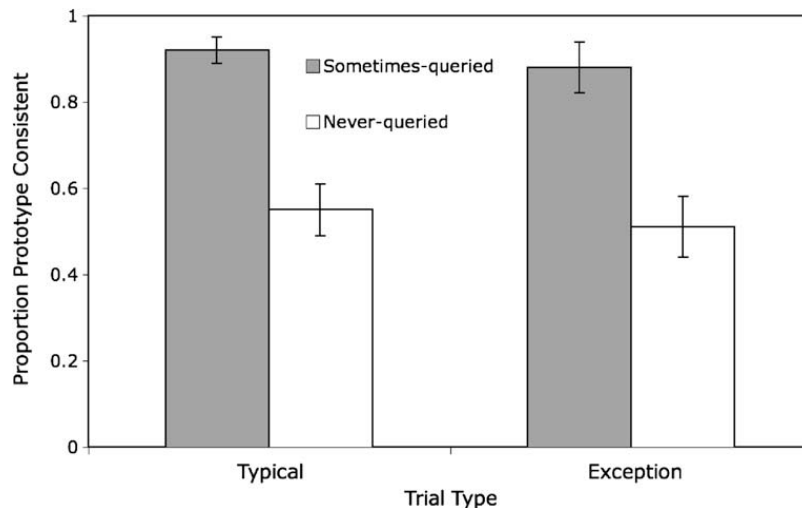


Fig. 13. Inference test results from Experiment 4. Error bars are standard errors of the mean.

typical and exception feature trials. The figure indicates that, as in Experiment 3, participants made far more prototype-consistent responses on the sometimes-queried dimensions (0.91) than the never-queried ones (0.54) reflecting the weaker learning of the latter. Notably, prototype-consistent responses on the never-queried dimensions were even less common than in Experiment 3 (0.64), a result consistent with the idea that informing participants which features would be queried and the self-paced feedback period reduced learning of those dimensions further still. A 2×2 repeated measures ANOVA of the inference test results with dimension (sometimes- or never-queried) and trial type (typical or exception) as factors revealed a main effect of dimension, $F(1,29) = 46.1$, $MSE = 4.00$, $p < 0.0001$. Performance on the never-queried dimensions did not differ from chance, $t < 1$. Unlike Experiment 3, Fig. 13 learners were no more likely to infer a typical feature on a typical feature trial (0.73) than an exception feature trial (0.69), $F < 1$, reflecting a failure to learn about the never-queried dimensions.

Eye movements were analyzed to assess the impact of telling participants which dimensions would be queried. If participants were simply learning in anticipation of future trials then fixations to the never-queried features should be reduced relative to Experiment 3. Examination of Fig. 14A which presents the probability of fixating never- and sometimes-queried dimensions and the category label confirms this conjecture, as the probability of observing a never-queried dimension during the first learning block (0.33) was about half of what it was in Experiment 3 (0.61). The reduction of fixations to the never-queried dimensions relative to Experiment 3 continued throughout learning (e.g., fixation probabilities in the last block of 0.08 vs. 0.13). In addition, Fig. 14B indicates that the proportion of time fixating the two never-queried dimensions is less in Experiment 4 than in Experiment 3 during both the first block (0.14 vs. 0.32) and the last (0.03 vs. 0.08).

These observations are confirmed by the histograms in Fig. 15 which show that, in comparison to Experiment 3, most learners ignored the never-queried dimensions dur-

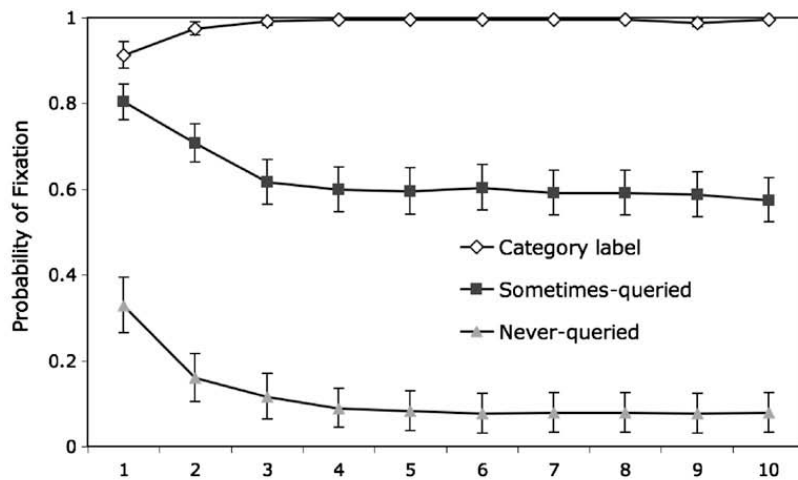
ing both the first (Fig. 15A) and last (Fig. 15B) blocks of learning. Indeed, Fig. 15B indicates that the small average probability of fixating the never-queried dimensions at the end of learning was almost entirely due to two participants who fixated those dimensions on every trial. Interestingly, those two participants also scored well (average accuracy of 0.83) on the single feature tests of the never-queried dimensions. If data from these two participants are excluded, the probability of fixating the never-queried dimensions in the last block drops from 0.08 to 0.03 and accuracy on the single-feature test for those dimensions drops from 0.55 to 0.52. Note that these outliers were also fixating the category label on every trial, making them unlike the Experiment 1 outliers that ignored the category label and used the other feature dimensions to predict the missing one. Of course, that the remaining participants were virtually ignoring the never-queried dimensions before they made their response rules out the possibility that they were using the other three feature dimensions as a redundant predictor.

Finally, an analysis of eye movements after feedback in Fig. 16 also shows a reduction in fixations to never-queried dimension relative to Experiment 3 both during the first block (probability of fixation 0.27 vs. 0.45; proportion fixation time 0.07 vs. 0.18) and the last (probability of fixation 0.05 vs. 0.27; proportion fixation time 0.03 vs. 0.10). Recall that in Experiment 4 participants were free to end the after-feedback stimulus presentation at any time. This suggests that many after-feedback fixations to never-queried features in Experiment 3 were an artifact of the fixed feedback duration of 3 s—they had to look at something.

Discussion

The purpose of Experiment 4 was to determine whether the learning of the never-queried dimensions in Experiment 3, slight though it was, was due to inference learners' desire to acquire the internal structure of the categories or to secondary aspects of the experimental procedure. In fact, no learning of those dimensions was exhibited in Experiment 4, a result that contradicts the possibility that participants were motivated to learn within-category

A. Probability of fixation



B. Proportion fixation time

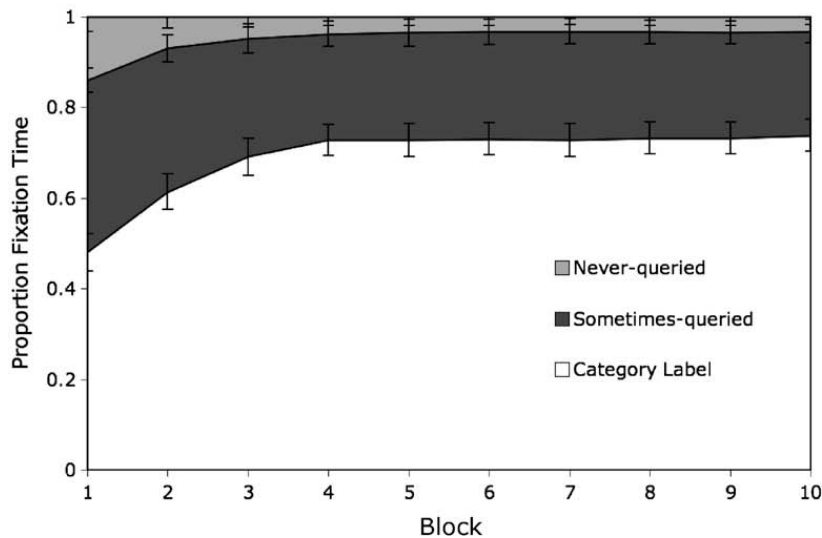


Fig. 14. Eye movements before feedback during inference learning in Experiment 4. (A) Probability of fixation. (B) Proportion fixation time. Error bars are standard errors of the mean.

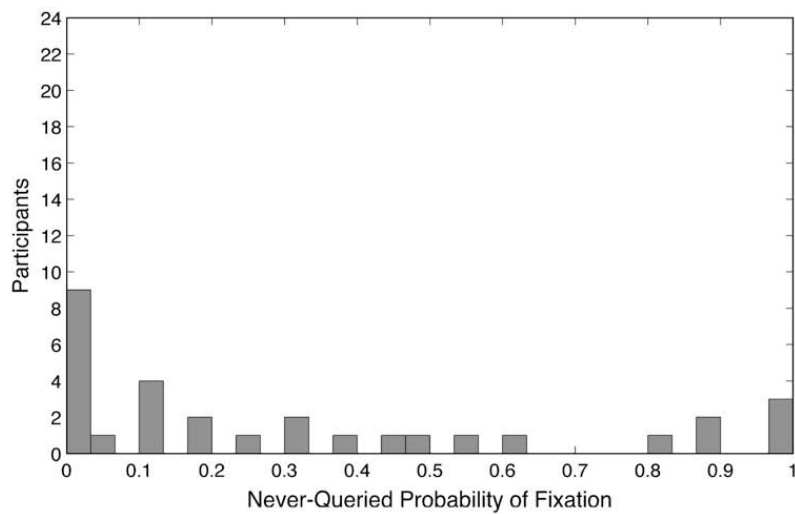
information above and beyond that needed to succeed at the experimental task.

Recall that although Experiment 3's inference learners were more likely to fixate sometimes-queried dimensions than never-queried ones throughout training, fixation probabilities for the never-queried dimensions were about .60 at the beginning of training. In contrast, this probability was cut in half in Experiment 4 by telling participants which dimensions would be queried. We interpret this finding as indicating that Experiment 3 participants were initially unaware of which feature dimensions would be queried and thus fixated all of them in anticipation of the future inferences they would be required to make. But they quickly learned which dimensions were never queried and stopped fixating them. In addition, whereas Experiment 3 learners were about as likely to fixate never-queried dimensions as sometimes-queried ones after feedback was received, after-feedback fixations to the never-queried dimensions were rare in Experiment 4. We interpret this as indicating that the after-feedback fixations in Experiment

3 were not due to participants deferring their learning of the never-queried dimensions to after they predicted the missing feature. Rather, they were due to the random fixations that occurred while the participant was waiting for the next trial to begin.

The results of the various tests provide some information about which of the procedural changes introduced in Experiment 4 affected learning. The lower rate of prototype-consistent responding on the never-queried dimensions in Experiment 4's inference test as compared to Experiment 3's suggests that some learning of never-queried dimensions in the Experiment 3 occurred as a result of either incidental learning that occurred after feedback or anticipatory learning that occurred at the beginning of the experiment (before participants knew which dimensions would be queried). And, that accuracy on the inference test was (slightly) greater than accuracy on the single feature test hints that some learning in Experiment 3 may have been from anticipatory learning at the beginning of the test phase itself. But regardless of which of

A. First block



B. Last block

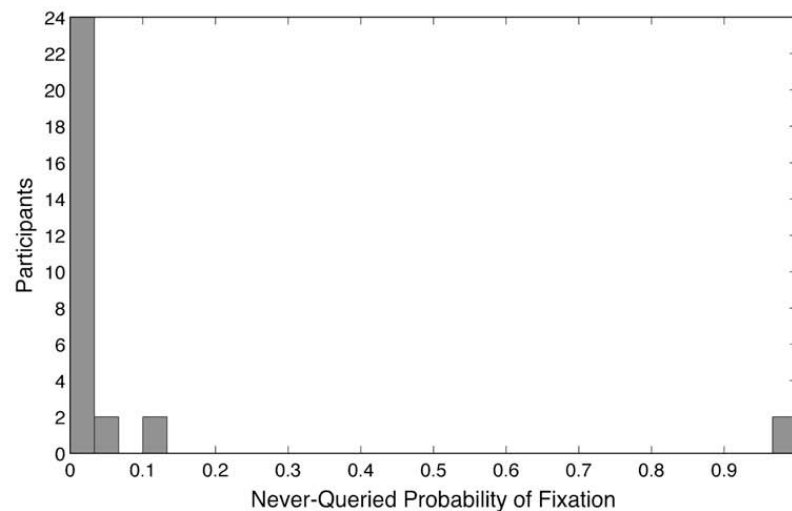


Fig. 15. Histograms of fixations to never-queried dimensions in Experiment 4. (A) First block. (B) Last block.

these procedural changes were responsible for these differences, the important point is that when the task demands that led participants to attend to the never-queried dimensions were eliminated, Experiment 4 participants exhibited no learning of those dimensions.

General discussion

The goal of this research was to investigate the mental processes involved in feature inference learning. Our specific aim was to discriminate between the category-centered learning hypothesis that states that inference learning promotes the goal of acquiring categories' internal structure and the alternative view that it induces the learning of category-to-feature rules. In fact, we found that neither hypothesis explained our results entirely. In the first two sections below we consider the implications our results have for these alternative views of inference learning. We then discuss our own proposal regarding the presence of anticipatory learning

during the feature inference task. We close by discussing the conditions that promote learning what a category is "like" and the role of category labels in feature inference.

Implications for category-centered learning

To review our main results, first recall that Markman and Ross (2003) proposed that predicting features motivates people to discover what a category is like. On the basis of this proposal we predicted that inference learners would attend to most stimulus dimensions on most training trials in order to learn the category's typical features, potential correlations among those features, and other abstract commonalities shared by category members. Consistent with this view, Experiment 1 found that inference learners fixated many feature dimensions in addition to the one being queried, and did so despite the fact that the category label could serve as the basis for perfect category-to-feature (inference) rules. Then, using a classifica-

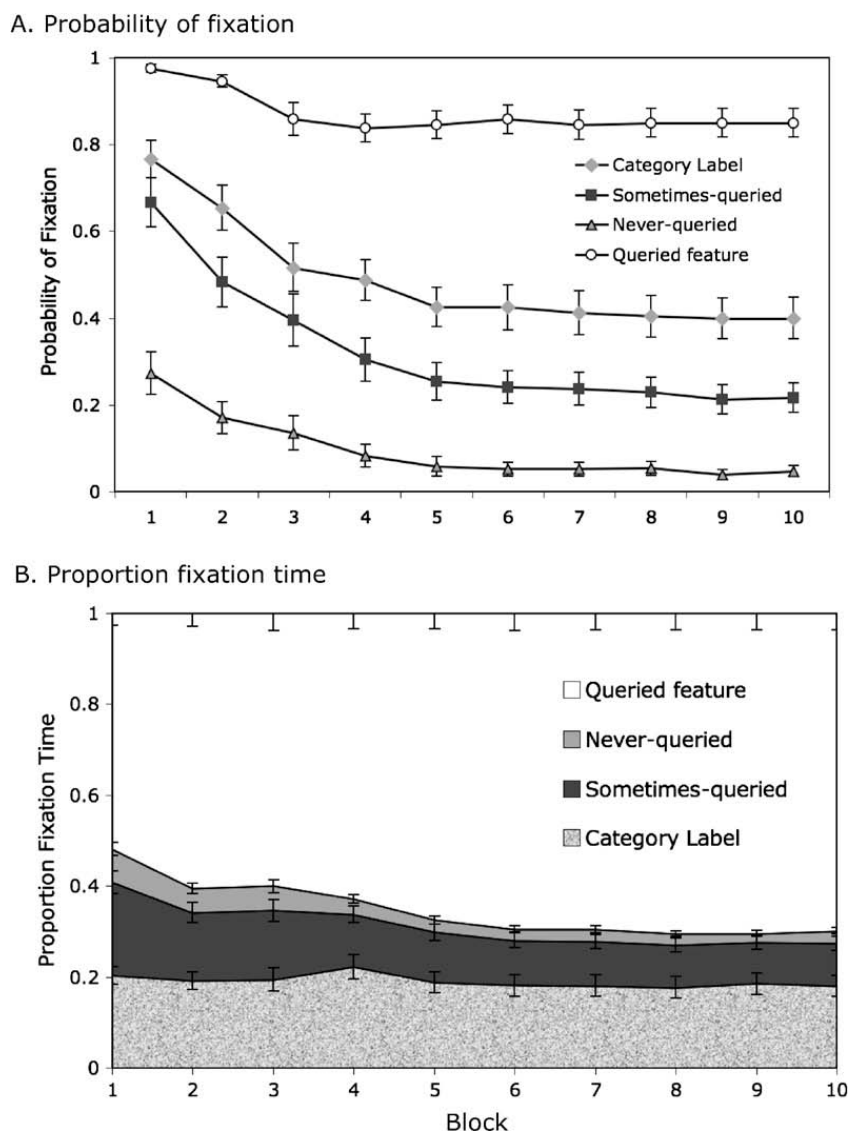


Fig. 16. Eye movements after feedback during inference learning in Experiment 4. (A) Probability of fixation. (B) Proportion fixation time.

tion task, Experiment 2 further demonstrated that the fixations to non-queried feature dimensions in Experiment 1 were not from the intrinsic salience of the stimuli. Taken together, Experiments 1 and 2 provided initial evidence that inference learning indeed promotes that acquisition of categories' internal structure.

However, in Experiments 3 and 4, we tested the CCL prediction that people should continue to fixate features that are never queried during the experimental task (so they can learn as much as possible about the categories' internal structure) but, contra this prediction, inference learners were far more likely to fixate sometimes-queried dimensions than the never-queried ones throughout training. In addition, they were also far more likely to predict prototype-consistent features for the sometimes-queried dimensions than for the never-queried ones, suggesting that participants learned little about the never-queried dimensions. These effects were especially pronounced when (a) participants were told beforehand which dimensions would be queried and (b) by making the after-feedback period self-paced in Experiment 4; indeed

inferences on the never-queried dimensions during test did not differ from chance. Recall that Experiments 3 and 4 also ruled out the possibility that the fixations to non-queried dimensions in Experiment 1 arose because those dimensions provided a secondary basis (redundant with the category label itself) on which to predict the missing feature.

We interpret these results as indicating that inference learners are not generally motivated to learn the internal structure of categories. Instead, they are motivated to do well on their assigned task. To accomplish this, we suggest that inference learners are engaged in both supervised and unsupervised learning. They fixate information (the category label) that is used to predict the feature that is missing on that trial and then use the feedback they receive to strengthen that category-to-feature association (supervised learning). But in addition, they also fixate feature dimensions they believe will be queried on future trials in order to strengthen the associations between them and the category label (unsupervised learning). Thus, although inference learners attended to information that

was unnecessary for performance on the present trial, those fixations did not reflect a general desire to learn about what the categories were like. Rather, they prepared the learners for making successful inferences on future trials.

Implications for the set-of-rules model

This anticipatory learning account of the feature inference task also has several implications for the second hypothesis we considered in this article, the set-of-rules model. While our results provide some support for that model, they also present several challenges. To begin with the positive, recall that the set-of-rules model predicts that learners will acquire those—and only those—category-to-feature associations needed to succeed at the inference task. According to Johansen and Kruschke (2005), the set of rules only includes “the most valid rule for each response dimension” (p. 1436, emphasis added). The results of Experiments 1, 3, and 4 were consistent with this view. In Experiment 1, inference participants inferred typical features on all four dimensions during training and in fact they inferred typical features on those dimensions during test. In Experiments 3 and 4, participants only predicted typical features on two of four dimensions and then at test consistently inferred typical features on the sometimes-queried dimensions but not the never-queried ones. Of course, Experiment 3 participants were above chance on the never-queried dimensions, but, as we have mentioned, this may be due to (weak) rules acquired at the beginning of training before participants knew which dimensions would be queried (or perhaps at the beginning of the test phase itself). When these potential sources of learning were eliminated in Experiment 4, no learning of the never-queried dimensions was exhibited.

But although the inference test results thus support the set-of-rules model, the eyetracking data suggesting unsupervised learning of the sometimes-queried dimensions in Experiments 3 and 4 present two potential challenges. We first present these issues and then propose alternatives to the set-of-rules model to address them.

Unsupervised “rule” learning

One reason that the unsupervised learning of feature dimensions is problematic for the set-of-rules model is that traditional views of “rule learning” assume that feedback is required to acquire rules. For example, according to Anderson’s (1987) theory of skill acquisition, skills consist of rules that are acquired as a result of experience in a domain. On this account, rule learning is analogous to instrumental conditioning in which condition-response associations are strengthened on the basis of feedback. Of course, because the rules involved in skill learning are assumed to be asymmetric (i.e., a rule’s consequent is available given the antecedent but not vice versa, Anderson, 1987; Anderson & Fincham, 1994; Bunsey & Eichenbaum, 1996), they differ from the bidirectional “rules” proposed by Johansen and Kruschke with which the category label can be predicted from the feature as easily as vice versa (an important property given that Johansen & Kruschke, 2005; Yamauchi & Markman, 1998, and the present exper-

iments have all demonstrated how information acquired through inference learning can be used to make subsequent classification decisions). Nevertheless, Johansen and Kruschke’s claim that rules are formed only on each “response dimension” indicates that their model shares with the traditional view that feedback is required for rules to be acquired. But as we have seen, this view of learning cannot explain the fixations to sometimes-queried dimension in Experiments 3 and 4, because no feedback is provided on those dimensions.

Conflicting representations

The second potential challenge concerns the presence of conflicting representations. Whereas the set of rules model specifies a single unambiguous rule relating the category label to a feature dimension, unsupervised learning on the non-queried dimensions raises the possibility of conflicting responses on each feature dimension. For example, for the inference problem Ax001 a learner might correctly predict $x = 0$ and then, on the basis of feedback, strengthen the rule relating the “A” category label and a 0 value on dimension 1 (supervised learning). The observation of 0 values on dimensions 2 and 3 might also strengthen the corresponding rules on those dimensions (unsupervised learning). But dimension 4 displays an exception feature, that is, a value in conflict with the rule relating “A” to a 0 value on dimension 4.

Moreover, it is clear that conflicts might also arise because which value should be predicted for a missing feature can vary from one category member to the next. Note that the inference tasks used in Yamauchi and Markman (1998), Johansen and Kruschke (2005), and the present experiments always had learners predict the same value on a dimension for a category. But in the real-world people are often confronted with exceptions. One will be right predicting that a dog has four legs until the first three-legged dog is encountered (Nilsson et al., 2005; Sweller et al., 2006). Dimensions of real-world categories can also have more than the two values (e.g., apples can be red, green, or yellow) that vary in their prevalence (apples are usually red). These sorts of conflicts point to the need to represent two or more potential responses per dimension, responses that differ in strength depending on the prevalence of the alternative values in the environment. These situations are inconsistent with Johansen and Kruschke’s assumption that the set of rules includes only one rule for each response dimension.

To summarize, the set of rules model is unable to account for the unsupervised learning of category-to-feature associations and multiple values per dimension. Note that these difficulties add to the challenges already faced by the set-of-rules model. As reviewed, that model is unable to account for the better learning of within-category correlations and of abstract commonalities shared by category members exhibited by feature inference learners as compared to classification learners (Chin-Parker & Ross, 2002; Erickson et al., 2005; Rehder & Ross, 2001). In contrast, these findings are readily accommodated by the anticipatory learning account we advocate, as now described.

From anticipatory to incidental learning

We have found that the feature inference task does not generally induce in learners the goal to learn the internal structure of categories. But if inferring features does not motivate people to learn what categories are “like,” we must provide an alternative account of the undisputed finding that inference learners often end up learning more about a category. We believe that the answer to this question lies in anticipatory learning and in the spread of attention over multiple feature dimensions that such learning induces. Specifically, attending to multiple dimensions results in the *incidental* learning of other aspects of the category's internal structure.

For example, recall that [Chin-Parker and Ross \(2002\)](#) found that feature inference learning resulted in participants learning more about within-category correlations than classification learning, at least when those correlations were not necessary for correct classification. On our view, the anticipatory learning induced by feature inference training led subjects to fixate most or all feature dimensions on each trial. This in turn led them to (incidentally) encode the co-occurrence information among those dimensions.

Anticipatory learning can also promote the learning of abstract information about categories. For example, [Rehder and Ross \(2001\)](#) and [Erickson et al. \(2005\)](#) found that categories defined by certain sorts of interfeature semantic relations are easier to learn by inference as compared to classification. We propose that this result obtained for two reasons. First, because the inference task led subjects to attend to more feature dimensions it resulted in them more readily noticing interfeature relations. Second, attending to multiple dimensions made it more likely that the item's features were encoded together in memory, allowing them to be compared with items presented on subsequent training trials. This in turn made it more likely that learners would notice the relational structures common to category members. Together, the results of [Chin-Parker and Ross](#), [Rehder and Ross](#), and [Erickson et al.](#) suggest that although inference participants may not be motivated to learn more than necessary, the spread of attention over multiple feature dimensions results in them learning more than necessary nevertheless.

Previous research has already documented how learners can acquire category information incidentally, notably in standard supervised classification learning. For example, the well-known study of [Allen and Brooks \(1991\)](#) found that subjects learned information about the context in which category exemplars typically appeared during a classification task even when they were told the correct classification rule beforehand. In these experiments, category members (a type of schematic animal) were presented against a background of environmental scenes. We suggest that, because during training learners were required to search for the item in the display, attention devoted to the background was sufficient for learners to incidentally encode that information as part of their category representation. Indeed, in a subsequent experiment, when items were presented at a fixed position on the computer screen, the effect of context disappeared. That is,

changing the information attended during training changed what information was (incidentally) learned (also see [Brooks, Squire-Graydon, & Wood, 2007](#); [Thibaut & Geisler, 2006](#)). In other words, the mediating role of attention in the incidental learning of category information is not unique to the feature inference task.

The results just reviewed suggest a complicated relationship between the experimental task, presented stimuli, and what is learned about categories. We suggest that a particular experimental task first leads learners to adopt a particular learning strategy. That strategy, in turn, leads them to attend to certain kinds of information. Subjects will then acquire information directly relevant to their learning goal; as we have seen that learning may be both supervised (reinforced with feedback) and unsupervised. But learners may also incidentally acquire information that was attended but not needed for the immediate task.

A new model of feature inference learning

To formalize our proposal regarding how anticipatory learning can lead to the incidental acquisition of additional category information, we now present a new model of feature inference learning referred to as the *partial exemplars model*. On the one hand, this model shares with the [Johansen and Kruschke set-of-rules model](#) the assumption that the category representations acquired through inference learning are exemplars with missing values. But it differs from the set-of-rules model by encoding *all* features attended on a feature inference trial, not just the category label and the predicted feature. Moreover, it assumes that those features are represented configurally, that is, in the same memory trace.

To illustrate these differences with a concrete example, [Table 3](#) presents the category representations stipulated by the set-of-rules model and our alternative model for the inference task of Experiment 4 assuming that dimensions 1 and 2 are the sometimes-queried dimensions and 3 and 4 are the never-queried ones. Note that, because dimensions 3 and 4 are never queried, the set-of-rules model postulates only two rules for each category: those between the category label and the typical values on dimensions 1 and 2. Following [Johansen and Kruschke](#), in [Table 3](#) these “rules” are represented as exemplars with missing values on the other three feature dimensions. In contrast, [Table 3](#) indicates how the partial exemplars model encodes the relative number of times the two sometimes-queried dimensions 1 and 2 display various combinations of 1s and 0s for each category. Specifically, Experiment 4 presented six feature inference trials for category A: category members A1 and A2 were queried on dimensions 1 and 2, A3 was queried on dimension 1, and A4 was queried on dimension 2 (recall that exception feature trials were not presented). As a consequence, category A's representation in [Table 3](#) encodes that inference learners saw two typical values ('00') on dimensions 1 and 2 on four presentations and mixed values ('01' and '10') on the remaining two.

This representation posited by the partial exemplars model has two advantages. First, it allows atypical as well typical features to be represented (e.g., the fact that apples can be green or yellow in addition to red). For example, in

Table 3 the partial exemplars encode the “associations” relating the category label A to not only the typical ‘0’ values but also the atypical ‘1’ values, reflecting the fact that those values were observed on sometimes-queried dimensions in Experiment 4. Importantly, however, because we assume that the model also encodes the relative number of presentations of each partial exemplar, it represents the fact the exception feature was observed on a dimension only once every six trials. Thus, by stipulating that the (partial) exemplars that encode typical dimension values are greater in number (or strength) than those that encode atypical values, the model can account for the finding that inference learners were far more likely to infer typical than atypical values at test.²

The second advantage of the partial exemplars model is that, because it encodes attended features configurally, it allows for the acquisition of category information not explicitly required by the inference task. For example, had a interfeature correlation existed between dimensions 1 and 2 in Experiment 4, that correlation would have been implicitly encoded in the partial exemplars. Because the partial exemplars model shares with the set of rules model the assumption of classification via a multiplicative similarity metric common to exemplar models (Medin, Altom, Edelson, & Freko, 1982), it allows sensitivity to any encoded correlation to be expressed on a subsequent test (Chin-Parker & Ross, 2002). It also enables the comparison of category members required for the learner to notice commonalities. For example, the partial exemplars in **Table 3** allow a learner to notice interfeature relations involving dimensions 1 and 2 that might be common across category members (Erickson et al., 2005; Rehder & Ross, 2001). Of course, the partial exemplars model predicts that an inference learner will only acquire those correlations and relations involving features that are attended, which in turn

² The possibility of multiple values per dimension suggested by the partial exemplars model suggests that some category-to-feature inferences will be made with greater certainty than others depending on the relative strength of those responses. Data presented by Johansen and Kruschke (2005) are suggestive of just this possibility. As mentioned, they tested a *nonexception condition* in which learners only predicted typical features (like Yamauchi & Markman, 1998, and the present Experiments 1, 3, and 4) and an *exception condition* in which they predicted nothing but exception features. In model comparisons involving an exemplar model, a prototype model, and a set of rules model, only the latter was able to provide an adequate account of both the exception and nonexception conditions. Nevertheless, the fits for the set of rules model was substantially better in the nonexception condition than the exception condition, a finding that was traced to the more inconsistent responding (i.e., more guessing) in the nonexception condition. We suggest that the more inconsistent responding in the exception condition was likely due to its large number of conflicts between supervised and unsupervised information. Whereas in the non-exception condition participants not only predicted typical features but also usually observed typical features on the other dimensions, in the exception condition predicted exception features but usually observed typical features on the other dimensions. Specifically, in the nonexception condition the probability that a non-queried dimension displayed a typical feature for the category structure tested in Johansen and Kruschke's Experiment 2 was .64. In other words, the predicted and observed features usually matched. In contrast, in the exception condition the probability that the predicted feature matched the value on a non-queried dimension was only .18. On our account, the frequent conflicts between predicted and observed information resulted from multiple conflicting responses per dimension. The result was less confident responding (i.e., more guessing) in the exception condition as compared to the nonexception condition.

depends on which dimensions they are asked to infer during training. That is, the incidental acquisition of category information is mediated by the anticipatory learning strategy required by the structure of the feature inference task.

To understand the importance of encoding configural information to incidental learning, it is illuminating to compare the partial exemplars model with one that fails to represent such information. In **Table 3**, the *multiple associations model* represents associations with both typical features (acquired through supervised learning) and atypical ones (unsupervised, anticipatory learning). However, it does so using the same sort of rules as the set-of-rules model, namely, as simple (non-configural) associations between the category labels and features. For example, for the feature inference task in Experiment 4, the multiple associations model encodes the associations between the “A” category label and typical (‘0’) and atypical (‘1’) feature values and also the observed 5:1 ratio of those values on dimensions 1 and 2 (**Table 3**). Critically, however, these simple rules would fail to encode any correlation that might have obtained between dimensions 1 and 2. And, they prevent the comparisons necessary to extract commonalities that obtain across category members.

To summarize, we have proposed a model that accounts for the unsupervised learning of category-to-feature associations and the incidental learning of additional category information. This proposal is tentative and will require systematic testing against these and other data sets. For example, one deficiency is that as described it assumes that category information is weighed equally regardless of whether it is acquired through supervised, unsupervised, or incidental learning, an assumption at odds with the well-known finding that feedback results in faster learning.³ But we suggest that a suitably elaborated partial exemplars model may form the basis of comprehensive account of the category representations formed via feature inference learning.

Learning what categories are “like”

We have seen that although feature inference learning can promote supervised, unsupervised, and incidental learning of category information, the present experiments found no evidence that learners were motivated to acquire categories' internal structure. However, this does not mean they lack such motivation in all circumstances. Interestingly, support for this claim comes from studies testing supervised classification, the very task that was supposed to result in less category learning than feature inference. For example, Hoffman and Murphy (2006) compared the supervised classification learning of categories with four

³ Indeed, Sweller et al. (2006) compared a condition that replicated Yamauchi and Markman (1998) with another that was identical except that during training subjects were also presented with exception feature trials and found that in the latter condition subjects were less likely to infer typical features at test (also see Nilsson & Ohlsson, 2005). And of course we have already seen in Johansen and Kruschke (2005) how changing which features are inferred during training has a dramatic effect on which are inferred during test. A partial exemplars model accommodate these results by representing information acquired through unsupervised learning less strongly than that acquired through supervised learning.

Table 3Category representations postulated by alternative models for the feature inference task of Experiment 4. *x* = unknown value.

Model	Category label	D1	D2	D3	D4	Relative # of observations
Set of rules	A	x	0	x	x	1
	A	0	x	x	x	1
	B	x	1	x	x	1
	B	1	x	x	x	1
Partial exemplars	A	0	0	x	x	4
	A	1	0	x	x	1
	A	0	1	x	x	1
	B	1	1	x	x	4
	B	1	0	x	x	1
	B	0	1	x	x	1
Multiple associations	A	x	0	x	x	5
	A	x	1	x	x	1
	A	0	x	x	x	5
	A	1	x	x	x	1
	B	x	1	x	x	5
	B	x	0	x	x	1
	B	1	x	x	x	5
	B	0	x	x	x	1

versus eight dimensions and found that subjects learned more features (i.e., associated more features with their correct category label) in the latter condition as compared to the former, even though those additional features were not necessary to succeed at the learning task. Bott, Hoffman, and Murphy (2007) used a blocking design in which subjects first learned to predict a response given a perfect predictor and then were presented with additional learning trials that each presented additional (probabilistic) cues. When they were predicting whether the computer would emit a low or high tone, subjects exhibited no learning of the associations between the additional cues and the outcome. That is, they exhibited blocking. But blocking was eliminated when subjects predicted letter strings instead (“Mobblies” or “Streaths”) that were plausibly the labels of actual categories (of automobiles, given that the “cues” were features of cars). Finally, Hoffman, Harris, and Murphy (2008) demonstrated that subjects will learn far more than necessary about a new category when subjects can use their prior knowledge to relate the new category to previously learned concepts. In each of these studies, subjects could succeed at the experimental task without learning anything else about the categories, and yet they learned more nonetheless.

These results suggest that classification versus feature inference learning may not be the key variable determining whether people become motivated to learn what a category is “like.” Rather it may depend on other subtle aspects of the experimental situation such as the materials and experimental instructions. All of these aspects may affect learners’ beliefs about what information about the category will likely be needed in the future. For example, if a category seems realistic, or useful, subjects may then be motivated to learn about it. Conceivably, our use of visual features (vs. the semantic features used by Bott et al., 2007) and opaque (and meaningless) category labels “A” and “B” (vs. “Mobblies” or “Streaths”) meant that our stimuli were not sufficiently “category like” for our participants to believe there was anything interesting to learn about

them above and beyond the experimental task. In addition, our experimental instructions may have emphasized the inference task so strongly that we inadvertently undermined any motivation to learn about the categories that might have otherwise existed. For example, telling Experiment 4 subjects beforehand which dimensions would be queried probably emphasized the experiment’s immediate goal (feature inference on two dimensions) even further at the expense of other learning. But although experimental variables can influence learners’ perception of the task which can in turn undoubtedly affect their performance, we have nevertheless shown that inference learning does not invariably induce the goal of learning what categories are like. Future research may identify those conditions that motivate subjects to learn more about categories beyond what is required by the experimental task, regardless of whether it involves feature inference, classification, or some other category-based judgment.

Features inferences and category labels

Finally, in addition to the implications that our eye-tracking data have for models of feature inference learning, there is one other aspect of our results that deserves emphasis. Although our goal of distinguishing models has led us to focus on whether inference learners fixate feature dimensions other than the queried one, it is important to note that each of our inference conditions found that the category label was the primary basis that learners used to predict the missing feature. Indeed, at the end of learning in Experiments 1, 3, and 4 subjects fixated the category label with probability >0.90 and they spent over half their time fixating the single category label as compared to the three non-queried dimensions. Moreover, across the 79 inference participants in Experiments 1, 3, and 4, unambiguous evidence that the category label *wasn’t* being used to infer features was found for only two subjects (in Experiment 1). These results corroborate a large number of studies showing that the category label plays a special role in

Table A

Fixations to dimensions as a function of whether they have a typical or exception feature for Experiments 1, 3, and 4.

Experimental condition	Fixation probability			Total proportion fixation time		Average proportion fixation time		
	Typical features	Exception features	Diff.	Typical features	Exception features	Typical features	Exception features	Diff.
Expt. 1/inference	0.49	0.41	0.08*	0.70	0.30	0.35	0.30	0.05
Expt. 1/classification	0.94	0.91	0.03	0.78	0.22	0.26	0.22	0.04*
Expt. 3/inference								
Sometimes-queried	0.59	0.60	0.01	0.75	0.77	0.75	0.77	0.02
Expt. 4/inference								
Sometimes-queried	0.58	0.57	0.01	0.87	0.88	0.87	0.88	–0.01

* Significant at alpha = 0.05.

category-based induction (Gelman & Markman, 1986; Johansen & Kruschke, 2005; Yamauchi & Markman, 2000a; Yamauchi et al., 2007).

Conclusions

We have shown that feature inference learning does not invariably lead to a desire to acquire categories' internal structure. However, it does lead them to engage in anticipatory learning in which on every trial they learn about the to-be-predicted feature (supervised learning) and about features that will need to be predicted on future trials (unsupervised learning). We have argued that the spread of attention over multiple feature dimensions thus induced enables the incidental learning of yet additional category information. So although inference learning may not be sufficient to energize people to learn categories, recognize that, if you want someone to learn a family resemblance category, including its prototypical features, within-category correlations, and other abstract commonalities, don't have them classify items. Have them predict features.

Acknowledgements

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

Appendix A

We present in Table A the probability of fixations and fixation times for typical and exception features during the last block of learning in Experiments 1, 3, and 4. In the table, *total proportion fixation time* is the time spent fixating typical and exception features excluding fixations to the category label. The *average proportion fixation time* corrects for differences in the number of the two types of features. For example, in Experiment 1's inference condition, each trial presented two typical features and one exception feature, so to make fixation times comparable, the total proportion fixation time to the typical features was divided by 2 to produce the typical features' average proportion fixation time. In Experiment 1's classification condition, each trial presented three typical features and one excep-

tion feature, so the total proportion fixation time to the typical features was divided by 3. Note that in Experiments 3 and 4 we only report fixations to the sometimes-queried dimension because of the small number of fixations to the never-queried dimensions.

Table A indicates that learners had a small preference for looking at the typical features in both the inference and classification conditions of Experiment 1. In the inference condition, fixation probabilities were 0.49 and 0.41 to the typical and exception features, respectively, and average proportion fixation times were 0.35 and 0.30. Only the difference in fixation probabilities was significant however. In the classification condition, fixation probabilities to the typical and exception features were 0.94 and 0.91 and average proportion fixation times were 0.26 and 0.22. Only the difference in average proportion fixation times was significant. In contrast, in Experiments 3 and 4 learners' fixations to the sometimes-queried dimension did not vary as a function of whether it displayed a typical or exception feature. Thus, learners' eye movements do not vary greatly as a function of whether a dimension contains information that is consistent or inconsistent with the rest of the stimulus.

References

- Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review*, 94, 192–210.
- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98, 409–429.
- Anderson, J. R., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology – Learning, Memory, and Cognition*, 20, 1322–1340.
- Anderson, J. R., & Fincham, J. M. (1996). Categorization and sensitivity to correlation. *Journal of Experimental Psychology – Learning, Memory, and Cognition*, 22, 259–277.
- Anderson, A. L., Ross, B. H., & Chin-Parker, S. (2002). A further investigation of category learning by inference. *Memory & Cognition*, 1, 119–128.
- Bott, L., Hoffman, A., & Murphy, G. L. (2007). Blocking category learning. *Journal of Experimental Psychology – General*, 136, 685–699.
- Brooks, L. (1978). Non-analytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Erlbaum.
- Brooks, L. R., Squire-Graydon, R., & Wood, T. J. (2007). Diversion of attention in everyday concept learning: Identification in the service of use. *Memory & Cognition*, 35, 1–14.
- Bunsey, M., & Eichenbaum, H. (1996). Conservation of hippocampal memory function in rats and humans. *Nature*, 379, 255–257.
- Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within-category correlations. *Memory & Cognition*, 3, 353–362.

- Chin-Parker, S., & Ross, B. H. (2004). Diagnosticity and prototypicality in category learning: A comparison of inference learning and classification learning. *Journal of Experimental Psychology – Learning, Memory, and Cognition*, 30, 216–226.
- Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, 36, 1827–1837.
- Erickson, J. E., Chin-Parker, S., & Ross, B. H. (2005). Inference and classification learning of abstract coherent categories. *Journal of Experimental Psychology – Learning, Memory, and Cognition*, 31, 86–99.
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, 23, 183–208.
- Hoffman, A. B., Harris, H. D., & Murphy, G. L. (2008). Prior knowledge enhances the category dimensionality effect. *Memory & Cognition*, 36, 301–315.
- Hoffman, A., & Murphy, G. L. (2006). Category dimensionality and feature knowledge: When more features are learned as easily as fewer. *Journal of Experimental Psychology – Learning, Memory, and Cognition*, 32, 301–315.
- Johansen, M. K., & Kruschke, J. K. (2005). Category representation for classification and feature inference. *Journal of Experimental Psychology – Learning, Memory, and Cognition*, 31, 1433–1458.
- Kowler, E., Anderson, E., Doshier, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, 35, 1897–1916.
- 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.
- Lassaline, M. E., & Murphy, G. L. (1996). Induction and category coherence. *Psychonomic Bulletin & Review*, 3, 95–99.
- Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Science*, 4, 6–14.
- Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, 129, 592–613.
- Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology – Learning, Memory, and Cognition*, 8, 37–50.
- Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, 19, 242–279.
- Nilsson, H., & Olsson, H. (2005). Categorization vs. inference: Shift in attention or in representation? In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 1642–1647). Stresa, Italy: Cognitive Science Society.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1990). *The adaptive decision maker*. New York: Cambridge University Press.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3–25.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 Years of research. *Psychological Bulletin*, 124.
- Rehder, B., & Hoffman, A. B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology*, 1, 1–41.
- Rehder, B., & Hoffman, A. B. (2005b). Thirty-something categorization results explained: Selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology – Learning, Memory, and Cognition*, 31, 811–829.
- Rehder, B., & Ross, B. H. (2001). Abstract coherent categories. *Journal of Experimental Psychology – Learning, Memory, and Cognition*, 27, 1261–1275.
- Ross, B. H. (2000). The effects of category use on learned categories. *Memory & Cognition*, 28, 51–63.
- Sakamoto, Y., & Love, B. C. (2006). Sizable sharks swim swiftly: Learning correlations through inference in a classroom setting. In R. Sun, & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 2087–2092). Mahwah, NJ: Erlbaum.
- Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, 75(3) (13, Whole No. 517).
- Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *The Quarterly Journal of Experimental Psychology*, 38, 475–491.
- Solomon, K. O., Median, D. L., & Lynch, E. (1999). Concepts do more than categorize. *Trends in Cognitive Science*, 3, 99–104.
- Sweller, N., Hayes, B. K., & Newell, B.R. (2006). Category learning through inference and classification: Attentional allocation causes differences in mental representation. In *Poster Presented at the 47th Annual Meeting of the Psychonomic Society*, November 16–19, Houston, TX.
- Thibaut, J., & Geisler, W. S. (2006). Exemplar effects in the context of a categorization rule: Featural and holistic influences. *Journal of Experimental Psychology – Learning, Memory, and Cognition*, 32, 1403–1415.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352.
- Yamauchi, T., Kohn, N., & Yu, N. (2007). Tracking mouse movement in feature inference: Category labels are different from feature labels. *Memory & Cognition*, 35, 544–553.
- Yamauchi, T., Love, B. C., & Markman, A. B. (2002). Learning nonlinearly separable categories by inference and classification. *Journal of Experimental Psychology – Learning, Memory, and Cognition*, 3, 585–593.
- Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language*, 39, 124–148.
- Yamauchi, T., & Markman, A. B. (2000a). Inference using categories. *Journal of Experimental Psychology – Learning, Memory, and Cognition*, 3, 776–795.
- Yamauchi, T., & Markman, A. B. (2000b). Learning categories composed of varying instances: The effect of classification, inference, and structural alignment. *Memory & Cognition*, 28, 64–78.