Title
A New Theory of Classification and Feature Inference Learning: An Exemplar Fragment Model

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 31(31)

ISSN
1069-7977

Authors
Colner, Robert
Rehder, Bob

Publication Date
2009

Peer reviewed

A New Theory of Classification and Feature Inference Learning: An Exemplar Fragment Model

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

Abstract

In addition to supervised classification learning, people can also learn categories by predicting the features of category members. One account of feature inference learning is that it induces a prototype representation of categories. Another is that it results in a set of category-to-feature rules. Because neither model provides an adequate account of existing data, we propose instead that inference learning induces an *anticipatory learning strategy* in which learners attend to aspects of training items they think will be needed in the future, and by so doing *incidentally* encode information about the category's internal structure. The proposal is formalized by an *exemplar fragment model* (EFM) that represents *partial* exemplars, namely, those parts that are *attended* during training. EFM's attention weights are approximated by eyetracking data, resulting in fewer free parameters as compared to competing theories.

When people classify objects, problem solve, describe concepts, or infer missing information, they must access conceptual knowledge. Thus, the question of how people learn and represent concepts has been central to the overall mission of cognitive psychology.

Researchers have developed sophisticated formal theories that explain many aspects of concept acquisition. These theories are largely based on *supervised classification learning* in which subjects classify items whose category membership is unknown and receive immediate feedback. Recently, to understand the interplay between how categorical knowledge is used and the concept acquired, researchers have begun to investigate a wider range of learning tasks (Brooks, 1978; Yamauchi & Markman, 1998, 2000a, 2002; Chin-Parker & Ross, 2002; Ross, 2000). For example, classification learning has been compared 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).

A Prototype Model of Feature Inference

Differences in how category information is acquired across classification and inference tasks were initially explained in terms of exemplars and prototypes. Yamauchi & Markman (1998) argued that inference learners represent categories as prototypes 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.

In their seminal study, Yamauchi & Markman (1998) contrasted classification and inference learning with a family resemblance category structure consisting of four items in two categories (labelled 'A' and 'B' in Table 1). Each category member has one *exception feature* from the other category. 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). Following learning, all participants completed a transfer test in which each feature was predicted in every training item, including both typical and exception features.

Test performance on the typical and exception feature trials in the classification and inference conditions in Yamauchi & Markman are presented in Figure 1 (see "YM" conditions) which present the proportion of responses that were consistent with the categories' prototype. (Figure 1 also includes the results from a number of other studies and fits of a new model described below.) The critical result was that learners responded with the category's typical feature more often than did the classification learners. The authors concluded that classification learners more often based inferences on the training exemplars whereas inference learners based theirs on the category prototypes. (Another result was that both groups were more likely to infer an exception feature on exception feature trials than typical feature trials, a point we return to below.)

Table 1. Category structure in Yamauchi & Markman's (1998).

Category	D1	D2	D3	D4
A1	0	0	0	1
A2	0	0	1	0
A3	0	1	0	0
A4	1	0	0	0
B1	1	1	1	0
B2	1	1	0	1
B3	1	0	1	1
B4	0	1	1	1

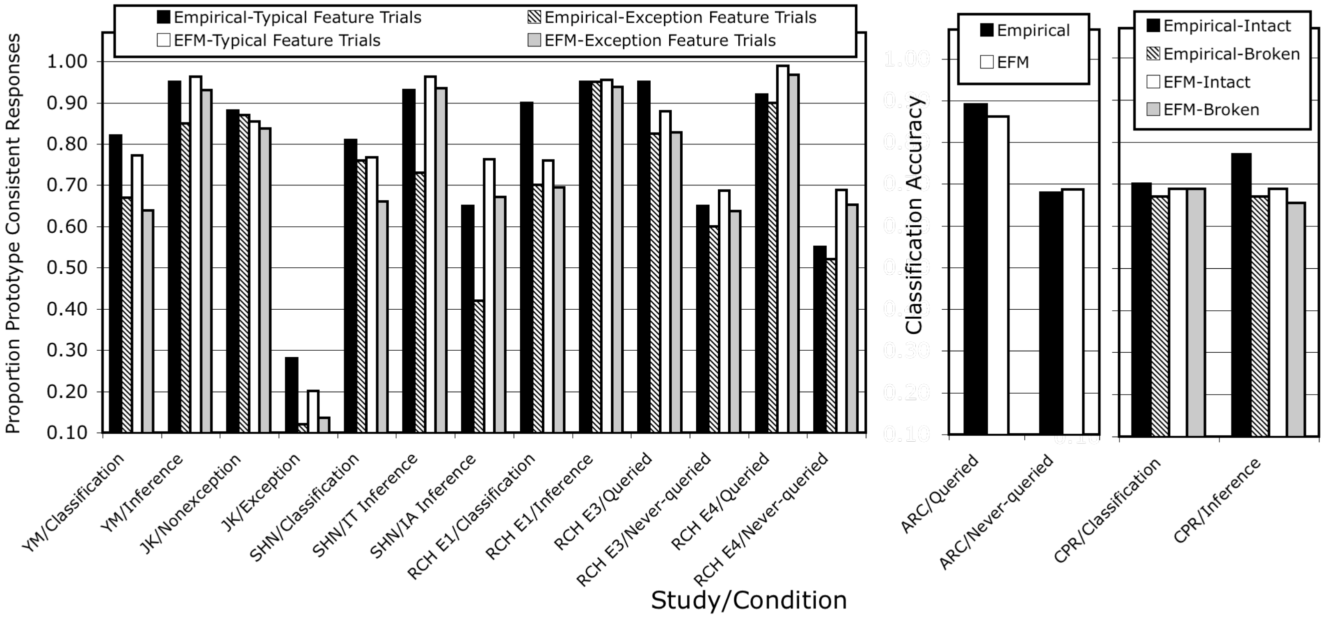

Figure 1. Empirical results and EFM model fits. *Note.* YM = Yamauchi & Markman. JK = Johansen & Kruschke. SHN = Sweller, Hayes, & Newell. RCH = Rehder, Colner, & Hoffman. ARC = Anderson, Ross, & Chin-Parker. CR = Chin-Parker & Ross.

To further test the claim that inference and classification learners represent categories differently, Yamauchi & Markman modified the General Context Model (GCM) to account for inference data by treating the category label as an additional feature. The model provided a good fit to the test data in the classification condition but not the inference condition (although, see Kruschke et al. 1999). These results led Yamauchi & Markman to conclude that inference learners indeed represent prototypes rather than exemplars.

Evidence Against Prototypes and for Rules

There is an alternative interpretation of the feature inference task, however. Johansen & Kruschke (2005) proposed that, rather than prototypes, inference learners in Yamauchi & Markman (1998) 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 with their typical features. Of course, the prototype and set-of-rules model may seem to be equivalent, because they both predict that in many circumstances (e.g., those in Yamauchi & Markman) 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 exact inferences made during training. For example, Johansen & Kruschke (2005, Expt. 2) compared a *nonexception condition* in which learners were only presented with typical feature trials with an *exception condition* in which they were only presented with exception feature trials. Whereas at test the nonexception group inferred typical features, the exception group inferred exception features (see Fig. 1, "JK" conditions). That is, they responded on the basis of the inferences required during training rather than on the categories' prototypes. Moreover, only the set-of-rules model provided a reasonable account of both the typical and exception conditions.

Related evidence was provided by Sweller, Hayes, & Newell (2006) in which subjects were tested in Yamauchi & Markman's classification and inference condition and a second inference condition in which both typical and exception feature trials were presented. In this condition (SHN/IA Inference condition, Fig. 1), subjects were less likely to predict typical features at test as compared to standard inference learning (SHN/IT Inference) and even classification learning (SHN/Classification) again suggesting that inference learners were not simply learning the category prototypes (also see Nilsson & Ohlsson, 2005).

Evidence Against Both Prototypes and Rules

The prototype and set-of-rules models have furthered our understanding of feature inference learning. However, examination of the literature reveals evidence against the set-of-rules model as well as additional evidence against the prototype model, as we now review.

Yamauchi & Markman (1998). As mentioned, inference learners were more likely to respond with an exception feature on exception feature trials than typical feature trials. This result indicates that they encoded some *configural* information about the categories' exemplars, that is, the combinations of 1s and 0s that appeared during training. For example, inference learners were more likely to predict x = 1 for test item A000x than A100x, apparently because A000x is so similar to item A1 in Table 1 which has a '1' on dimension 4. But such configural information is represented by neither a category prototype nor a set of rules that merely associates category labels and features.

Chin-Parker & Ross (2002). Subjects learned categories that possessed within-category correlations either by classification or inference. The correlations were not necessary for correct classification. They then performed a double feature classification test consisting of either *intact* feature pairs (features that appeared together during training) or *broken* pairs (features that never appeared together). Inference but not classification learners were more accurate on intact vs. broken pairs, indicating the sensitivity of the former to the within-category correlation (see Fig. 1, "CR" conditions). This finding again suggests that inference learning promotes the learning of configural information such as feature correlations was not explicitly tested during training. As mentioned, neither the prototype nor the set-of-rules model represents configural information such as feature correlations.

Anderson, Ross, & Chin-Parker (2002). Subjects completed a feature inference task in which only two of the four feature dimensions were queried during training. On a subsequent single-feature classification test, participants were more accurate on the two queried dimensions than the two never-queried ones (see Fig. 1, "ARC" conditions). This result is inconsistent with a prototype model that predicts that typical features on all dimensions should be represented equally well and again emphasizes the importance of which specific features are predicted during training. However, the fact that learners were above chance on the never-queried dimensions is inconsistent with the set-of-rules that assumes that only rules on queried dimensions are represented and again emphasizes that inference learners can acquire category information not explicitly tested during training.

Rehder, Colner, & Hoffman (in press). To gather more information about feature inference learning, Rehder et al. (in press, Expt. 1) replicated the Yamauchi & Markman (1998) study with an eyetracker. The gathering of eye movement data as another dependent variable is useful because models make different claims regarding the allocation of attention during feature inference learning. Besides replicating the essential results from Yamauchi & Markman (see Fig. 1, "RCH E1" conditions), they found that during training the large majority of inference learners fixated not only the category label but also most of the other features displayed by the training item, and did so despite the fact that the category label was perfectly predictive of the missing feature. This result provides prima facie evidence against the set-of-rules model, because the view that feature inference involves applying category-to-feature rules suggests that the reasoner will only fixate the antecedent of the rule (the category label). Indeed, Rehder et al. (Expt. 2) found that learners quickly limited fixations to the antecedent when one-dimensional feature-to-category classification rules were being acquired (also see Rehder & Hoffman, 2005a). Fixating most feature dimensions on most trials appears to support the notion that inference learners were trying to acquire the category prototypes.

However, this conclusion was tested by Rehder et al.'s Expt. 3 that replicated the Anderson et al. (2002) study described earlier in which only two of four of the feature dimensions were queried. The prototype model predicts that people should continue to fixate features that are never queried (so they can learn the typical features on those dimensions) but, contra this prediction, inference learners were more likely to fixate sometimes-queried dimensions than the never-queried. Indeed, the never-queried dimensions were virtually never fixated by the end of training. (As in Anderson et al., they were also far more likely to predict typical features for the sometimes-queried dimensions than for the never-queried ones which in turn were above chance, see Fig. 1, "RCH E3" conditions.) Rehder et al. concluded that inference learners fixated other feature dimensions during Expt. 1 training trials not because they were learning the category prototypes but rather because they anticipated being queried on those dimensions on future trials, a view they referred to as the *anticipatory learning hypothesis*. Moreover, they argued that the above-chance performance on (and early fixations to) the never-queried dimensions in Expt. 3 arose because inference learners initially thought they would queried on those dimensions. Indeed, when inference learners were informed at the start of the experiment which dimensions would be queried, fixations to the never-queried dimensions were almost entirely absent (and subjects were at chance on those dimensions) (Rehder et al. Expt. 4, see "RCH E4" conditions in Fig. 1).

A logical objection. Finally, we observe that both the prototype and set-of-rules model embody the unrealistic assumption that people represent only one possible value per feature dimension. As a result, people but not these models can represent the fact that most apples are red but some are yellow and green (but none are blue). Indeed, without elaboration, the set-of-rules model is unable to model the study of Sweller et al. (2006) that required different responses on the same dimension on different trials.

Summary. On the basis of these studies, we conclude that the prototype and set-of-rules models fail to account for (a) the learning of configural information, (b) the (unsupervised) learning of category information not explicitly tested during training, (c) eye fixation data, or (d) the fact that people know multiple values per dimension.

Instead of supporting the prototype or set-of-rules model, we argue that these studies together motivate a new account of the feature inference task—the anticipatory learning account—that postulates that inference learners represent neither prototypes nor a set of rules but rather allocate attention to those aspects of categories they think will be needed in the future, and by so doing *incidentally* learn many additional aspects of a category's structure (e.g., feature correlations). In other words, the inference task leads participants 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). This anticipatory strategy spreads attention over multiple feature dimensions and enables the incidental learning of additional category information not explicitly required by the task.

The Exemplar Fragment Model

To formalize our proposal regarding how anticipatory learning leads to the incidental acquisition of category information, we now present a new model of category learning, the exemplar fragment model (EFM). Unlike the GCM, which assumes that each presented exemplar is represented completely, or the set-of-rules model, which only encodes an association between the category label and the predicted feature, our model assumes that the dimensions encoded on a given trial are those that are *attended*. Moreover, the strength of each dimension's encoding varies as a function of (a) how much attention it receives and (b) whether it was the queried dimension (that received feedback). Importantly, because the learner's attention to (and thus encoding of) the same exemplar can vary from trial to trial, the EFM represents exemplar *tokens* (one representation for each subject's exposure to that exemplar) rather than *types*. Finally, EFM also assumes that inferences (of a category label or a missing feature) are affected by which aspects of the current stimulus are being attended.

The representations proposed by EFM have three advantages. First, they naturally represent the fact that people know of more than one value per feature dimension (apples are red but occasionally green or yellow). But because the model encodes the relative number of presentations of each exemplar, EFM, like humans, is more likely to infer typical than atypical values at test.

The second advantage is that, because EFM encodes attended features configurally, it allows for the acquisition of category information not explicitly required by the inference task. For example, if an interfeature correlation exists between two dimensions, that correlation is implicitly encoded in the exemplar fragments. Because EFM assumes classification via a multiplicative similarity metric common to exemplar models (Medin et al., 1982), it allows sensitivity to any encoded correlation to be expressed on a subsequent test. Of course, EFM only encodes those feature configurations that are attended during training.

The third advantage is that EFM assumes that predicted dimensions are encoded more strongly than those that are only observed. This allows EFM to exhibit sensitivity to the specific inferences that are carried out during training.

We now define EFM in the same terms as the standard exemplar model. Below are the equations associated with the GCM, generalized so that they predict values on any queried dimension, not just the category label. Specifically, the probability that a test item t has value 1 rather than 0 on the queried dimension q is,

$$P(t_q = 1|t) = TotalSim^{\gamma}(t, 1) \, / \, (TotalSim^{\gamma}(t, 1) + TotalSim^{\gamma}(t, 0)) \quad (1)$$

$$TotalSim\,(t, v) = \Sigma_{m \in M, m_q = v}\, Sim\,(t, m) \quad (2)$$

$$Sim\,(t, m) = e^{-cDist(t,m)} \quad (3)$$

$$Dist\,(t, m) = \Pi_{i=1..n}\, d_i(t, m) \quad (4)$$

$$d_i(t, m) = w_i|t_i - m_i| \quad (5)$$

where M is the set of stored category members, n is the number of dimensions (including the category label), $d_i(t, m)$ is the distance between t and m on dimension i, c is a sensitivity parameter, and γ is a response scaling parameter. The w_is are *attention weights* that multiply the mismatches between t and m on each dimension and so allow dimensions to have unequal influence on inferences.

EFM introduces two changes to this definition of an exemplar model. From their inception, so-called "attention weights" were understood to subsume more than one factor. A dimension might have a low attention weight either because a classifier has learned to ignore that dimension in to-be-classified stimuli or because it was ignored during training and thus poorly encoded. EFM extends the class of exemplar models by decomposing an attention weight w_i into one component ($EncodingWt_{m,i}$) that represents how strongly dimension i of stored exemplar m was encoded and a second component ($TestWt_{t,i}$) that represents the degree to which a dimension is attended in test stimulus t. The value of w_i is formed by multiplying $EncodingWt_{m,i}$ and $TestWt_{t,i}$ and then normalizing the result (so that all ws sum to 1).

$$w_i = (EncodingWt_{m,i} * TestWt_{t,i}) \, / \, \Sigma_{j=1..n}\, EncodingWt_{m,j} * TestWt_{t,j} \quad (6)$$

This definition of w_i exhibits the following important properties. First, if dimension i wasn't encoded when m was presented, then $EncodingWt_{m,i}$ is 0 and so is w_i; thus a mismatch between t and m on dimension i has no effect on their similarity. Second, if dimension i of the test stimulus is not attended, then $TestWt_{t,i} = 0$, and again a mismatch on dimension i becomes irrelevant.

The second change allows the encoding weight of the *being predicted* dimension q to affect the response by replacing Eq. 2 with Eq. 2',

$$TotalSim\,(t, v) = \Sigma_{m \in M, m_q = v}\, EncodingWt_{m,q} * Sim\,(t, m) \quad (2')$$

In other words, even if t and m are highly similar on all other dimensions, if the queried dimension q of m was never encoded ($EncodingWt_{m,q} = 0$) then m will have no effect on whether a 1 or 0 is predicted for that dimension.

Approximating Encoding Weights

EFM's separation of encoding weights from test weights potentially grants it the flexibility needed to account for the feature inference task. However, by itself the usefulness of EFM is limited because of the excessive number of parameters it introduces (encoding weights for each observed example and test weights for each test stimulus). In this article we address this concern by approximating these weights from eyetracking data.

Not all studies of feature inference learning used eyetracking of course. However, the eyetracking results in Rehder et al. (in press) can be used to model not only the behavioral results from that study, they can be extrapolated to a number of other studies. The eyetracking result from the feature inference conditions in Rehder et al. are presented in Table 2. Note that on each inference training trial, there was one queried dimension and four nonqueried dimensions: the *category label*, the *sometimes queried dimensions* that were queried during training but not on the current trial, and the *never queried dimensions* that were never queried. (The number of sometimes and never queried dimensions was 3 and 0 in Expt. 1, because all dimensions were predicted, and 1 and 2 in Expts. 3 and 4, because only two dimensions were predicted.) Table 2 presents the proportion of time the average subject spent fixating each type of dimension in each experiment. Recall that, using eye movement data, EFM can represent each exemplar token observed by each subject. However, because the goal of this article is to account for group level data only, we use eye movement data averaged over subjects.

Table 2. Proportion fixation times to dimensions in Rehder, Colner, & Hoffman's (in press) inference conditions.

	Training Eye Fixations			Encoding Weights				Test Eye Fixations		
Experiment	Category Label	Sometimes Queried	Never Queried	Category Label	Queried	Sometimes Queried	Never Queried	Category Label	Sometimes Queried	Never Queried
1	.538	.154	–	.179	.667	.051	–	.588	.137	–
3	.615	.281	.102	.205	.667	.094	.010	.532	.266	.201
4	.690	.260	.049	.230	.667	.087	.005	.608	.236	.156

As mentioned, a second factor that influences encoding weights, but not test weights, is the presence of feedback. EFM assumes that on any given trial the dimension that received feedback will be more strongly encoded than those that were only observed. Accordingly, a *feedback multiplier parameter (FM)* determines the relative strength of encoding for the queried dimension vs. the observed dimensions.

The encoding weights for each experiment are derived from the training eye fixations taking into account the feedback multiplier *FM*, as specified by Eqs. 7 and 8,

$$EncodingWt_{m,q} = (FM/n) \,/\, (FM/n\ + 1) \qquad (7)$$

$$EncodingWt_{m,i} = TrainProportionFixationTime_{m,i} \,/\, (FM/n\ + 1) \qquad (8)$$

where q is the queried dimension and n is the number of dimensions. For example, Table 2 presents the values of $EncodingWt_{m,i}$ when $FM = 10$ for the three inference conditions in Rehder et al. (in which $n = 5$). As a result, the queried dimension q in each stored exemplar m has ten times the encoding strength of the average nonqueried dimension.

Table 2 also presents the average proportion eye fixations observed during test in Rehder et al.; these fixations are used as the values for $TestWt_{t,i}$.

$$TestWt_{t,i} = TestProportionFixationTime_{t,i} \qquad (9)$$

We also use eye fixation data from Rehder et al.'s Expt. 1 classification condition (not shown in Table 2) to approximate the encoding and test weights in that condition. Because each feature dimension was fixated about equally during training, the encoding weights were derived from Eqs. 7 and 8 assuming that $TrainProportionFixationTime_{m,i}$ = .25. The test weights were derived from the proportion fixation times observed during test: .325 on the category label and .225 on each feature dimension.

Table 3 indicates how, with a few exceptions, the encoding and test weights used to model the conditions in Rehder et al. are used to model the other studies. The exceptions are as follows. Because Anderson et al. presented single feature classification tests, the test weight on this single dimension is 1. Because Chin-Parker & Ross presented double feature classification trials at test, in the inference condition the test weights on those two dimension are .5. Note that, because of the Chin-Parker & Ross's category structure, each dimension was a perfect predictor of the category label. Thus, in the classification condition we apply previous results showing that learners will attend exclusively to a single perfectly diagnostic dimension (Rehder & Hoffman, 2005a) and assume an encoding and test weight of 1 on one dimension and 0 on all others.

Table 3. Application of Rehder et al.'s weights to other studies.

Expt./Condition	Encoding Weights	Test Weights
E1/Inference	YM/Inference, JK, SHN/Inference [IT&IA] CPR/Inference	YM/Inference, JK, SHN/Inference [IT&IA]
E1/Classification	YM/Classification SHNClassification	
E2/Inference	ACR	

Note that because $EncodingWt_{m,i}$ and $TestWt_{t,i}$ are measured directly from eyetracking data, EFM has no free attention weight parameters, a difference that results in a sharp reduction in its number of parameters relative to standard exemplar models. In the following simulations, we assume one c parameter for each study, and γ and *FM* parameters that are common across studies.

Results

In this article our primary goal is to establish that EFM is sufficient to provide a qualitative account of the key feature inference learning results we have reviewed. Accordingly, we used an informal model fitting approach in which parameters were tuned by hand. This results in values of c of 9.1, 2.6, 7.7, 6.3, 5.5, and 1.1 for Yamauchi & Markman, Johansen & Kruschke, Sweller et al, Anderson et al., Rehder et al, and Chin-Parker & Ross, respectively; the best fitting values for γ and FM were 1.1 and 10.5. The empirical results from each study are presented in Figure 1 alongside the EFM fits. As the figure indicates, EFM reproduces most of the important results from these studies.

Yamauchi & Markman. EFM correctly predicts more prototype consistent responding in the inference condition versus the classification condition. It also correctly predicts more prototype consistent responding on the typical feature trials than the exception feature trials.

Johansen & Kruschke. EFM correctly predicts that subjects will respond with the typical features in the nonexception condition and with atypical features in the exception condition. As for Yamauchi & Markman, it correctly predicts more prototype consistent responding on the typical feature trials than the exception feature trials.

Sweller, Hayes, & Newell. EFM correctly predicts that the IT inference condition but not the IA inference condition produces more prototype consistent responding as compared to the classification condition. One deficiency is that EFM did not produce the lower rate of prototype consistent responding in the IA condition as compared to the classification condition.

Chin-Parker & Ross. EFM correctly predicts a sensitivity to feature correlations in the inference condition and the absence of this sensitivity in the classification condition.

Anderson, Ross, & Chin-Parker. EFM correctly predicts more prototype consistent responding on the queried dimensions than the never-queried ones which in turn are above chance.

Rehder, Colner, & Hoffman. For Expt. 1 EFM correctly produces more prototype consistent responding in the inference condition than the classification condition. For Expts. 3 and 4, it correctly predicts more prototype consistent responding on the queried dimensions than the never-queried ones, which in turn are above chance. One deficiency is that it fails to produce the lower performance on the never-queried dimensions in Expt. 4 as compared to Expt. 3.

General Discussion

The EFM was presented as a formal model of the anticipatory learning hypothesis. It is distinguished by its ability to account for both supervised learning of the predicted dimension as well as unsupervised learning of dimensions merely observed. On this account, the demands of the task are the key determiner of attention. In the feature inference task, because multiple features are queried during training, attention is spread among the queried features to learn them in *anticipation* of future trials. This in turn enables the *incidental* encoding of category information not explicitly tested during training. This approach allowed EFM to exhibit the key properties we have noted, namely, the encoding of configural information, information not explicitly tested during training, and multiple values per dimension. To our knowledge, EFM is unique in providing a qualitative account of the key results in six studies, and we expect that further development (e.g., formal model fitting) will result in excellent quantitative fits as well.

EFM can be conceived of as an active learning theory in that the interesting aspects of the model result from learners' active contributions. Only the feature dimensions that were actively sampled by the learner are encoded, which in turn determines the model's response via similarity with the currently fixated dimension of the test item. It is an open question if the sampling of features observed in these studies represents optimal information gain, though some related work suggests it might (Nelson & Cottrell, 2005).

Another unique property of EFM of course is its use of eyetracking data. Previous applications of exemplar models have made the (obviously incorrect) assumption that category exemplars are encoded perfectly in memory. Eyetracking allowed us to determine what information was attended and thus approximate what was encoded. Veridical representations of learners' category knowledge was key to EFM's success at modeling test performance from multiple studies.

Finally, it is important to consider the EFM within the broader theoretical scope of concept learning models. Specifically, a multiple-systems view of category learning may provide a useful framework to evaluate the parameters of the EFM. For example, the COVIS multiple-systems model assumes that two separate memory systems—an explicit verbal (rule-based) system and a procedural system—are involved in the acquisition of (perceptual) concepts (Ashby et al. 1998). Fitting the EFM to inference data required setting the feedback multiplier (FM) to a relatively high value (~10). In effect, this validates the central assumption of the set-of-rules model, that associations between the category label and the features acquired on the basis of feedback are strongly represented. Therefore, an alternative modeling approach involving combining explicit category-to-feature rules with stored exemplars might prove fruitful.

Acknowledgments

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

References

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

Ashby G., Alfonso-Reese, L., Turken, A., Waldron, E. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105,* 442-81.

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.

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

Chin-Parker, S., & Ross, B. H. (2004). Diagnosticity and prototypicality in category learning: a comparison of inference learning and classification learning. *JEP:LMC, 1*, 216-26.

Johansen, M. K., & Kruschke, J. K. (2005). Category representation for classification and feature inference. *JEP:LMC, 6*, 1433-58.

Kruschke, J. K., Johansen, M. K., & Blair, N. (1999). Exemplar Model Account of Inference Learning. Unpublished Manuscript.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *JEP:LMC. 8*, 37-50.

Nelson, J., & Cottrell, G. (2007). A probabilistic model of eye movements in concept formation. *Neurocomputing*, *70, 2256-72.*

Nilsson, H., & Olsson, H. (2005). Categorization vs. inference: Shift in attention or in representation? Proceedings of the 27th Annual Conference of the Cognitive Science Society (pp. 1642-1647). Stresa, Italy: Cognitive Science Society.

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

Rehder, B., Colner, R., Hoffman, A. (in press) Feature inference learning and eyetracking. *Journal of Memory and Language.*

Sweller, N., Hayes, B. K., & Newell, B. R. (2006). Category learning through inference and classification: Attentional allocation causes differences in mental representation. Poster presented at The 47th Annual Meeting of the Psychonomic Society. November 16-19, Houston, TX.

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

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *JEP:LMC, 3*, 776-95.

Yamauchi, T., Love, B. C., & Markman, A. B. (2002). Learning nonlinearly separable categories by inference and classification. *JEP:LMC, 3*, 585-93.