Title

A Configural-Cur Network Model of Animal and Human Associative Learning

Authors
Gluck, Mark A.
Bower, Gordon H.
Hee, Michael R.

Publication Date
1989

Peer reviewed

Eleventh Annual Conference of the Cognitive Science Society, Ann Arbor, MI. August 16-19, 1989.

A CONFIGURAL-CUE NETWORK MODEL OF ANIMAL AND HUMAN ASSOCIATIVE LEARNING

Mark A. Gluck Gordon H. Bower Michael R. Hee

Department of Psychology
Stanford University

ABSTRACT

We test a configural-cue network model of human classification and recognition learning based on Rescorla & Wagner's (1972) model of classical conditioning. The model extends the stimulus representation assumptions from our earlier one-layer network model (Gluck & Bower, 1988b) to include pair-wise conjunctions of features as unique cues. Like the exemplar context model of Medin & Schaffer (1978), the representational assumptions of the configural-cue network model embody an implicit exponential decay relationship between stimulus similarity and and psychological (Hamming) distance, a relationship which has received substantial independent empirical and theoretical support (Shepard, 1957, 1987). In addition to results from animal learning, the model accounts for several aspects of complex human category learning, including the relationship between category similarity and linear separability in determining classification difficulty (Medin & Schwanenflugel, 1981), the relationship between classification and recognition memory for instances (Hayes-Roth & Hayes-Roth, 1977), and the impact of correlated attributes on classification (Medin, Altom, Edelson, & Freko, 1982).

In earlier papers, we have explored a simple adaptive network as a model of human learning (Gluck & Bower, 1986, 1988a, 1988b; Gluck, Corter, Bower, & Kyleberg, 1988). We have used Rescorla and Wagner's (1972) description of classical conditioning and extended it to human classification learning. The learning rule is the same as the least mean squares (LMS) learning rule for training one-layer networks (proposed by Widrow & Hoff, 1960). The model has been fit to data from experiments on probabilistic classification learning with multiple cues. While this simple model can be applied to only a restricted range of experimental circumstances, it has shown a surprising accuracy in predicting human behavior within that range--people's choice percentages during learning, the relative difficulty of learning various classifications, as well as their responses to generalization tests involving novel combinations of cues.

This paper extends the stimulus representation assumptions used previously. We assume in this "configural-cue" model that pair-wise conjunctions of stimulus features are encoded as unique elements. This configural cue assumption is common in the animal learning literature (e.g., Wagner

Correspondence regarding this article should be addressed to: Mark A. Gluck, Department of Psychology, Jordan Hall; Bldg. 420, Stanford University, Stanford, CA 94305; or email to *gluck@psych.stanford.edu*. For their thoughtful comments and advice on this work we are indebted to: James Corter, W. K. Estes, Robert Nosofsky, Jay McClelland, Doug Medin, Misha Pavel, Robert Rescorla, and Allan Wagner. This research was supported by NSF Grant BNS-8618049.

and Rescorla, 1972), and has been used to explain a range of results (e.g., Rescorla, 1972, 1973). This paper shows how this extended model accounts for several aspects of complex category learning by humans.

BACKGROUND

The ingredients of the basic network model are shown in the left side of Figure (1A). Presentation of a stimulus or pattern of cues corresponds to activating one or more of the sensory elements on the left. They send their activations to a single output unit along associative lines which have amplifier weights, the w_i. The weighted inputs are summed at the output node, and this output $\sum_{j=1}^{n} w_j a_j$, is converted into some response measure. In a classical conditioning situation, the inputs are single to-be-conditioned stimuli such as lights and bells that are paired with the unconditional stimulus, such as food for a hungry dog; the output node reflects the animal's expectation of the unconditional stimulus given the cues presented. In a classification experiment involving human adults as subjects, the stimuli might be patterns of, say, medical symptoms displayed by a patient, and the output reflects the degree to which the model expects such a patient to have some target disease (classification) versus alternative diseases.

The network operates in a training environment in which reinforcing feedback (the UCS or the correct classification) is given just after each stimulus pattern. The central axiom of the model is its learning rule, which is that the weights, the w_i's, change on each trial according to Equation 1:

$$\Delta w_i = \beta a_i (\lambda - \sum_{j=1}^{n} w_j a_j) \qquad (1)$$

Here, λ is the training signal which might be +1 for the correct category and 0 for an incorrect category. The cue-intensity parameter, a_i is assumed to be 1 if cue i is present on the trial, and 0 if it's not. The learning rate, β, is a parameter (on the order of .01 in most simulations) that determines how much the weights change when the output differs from the training signal, λ.

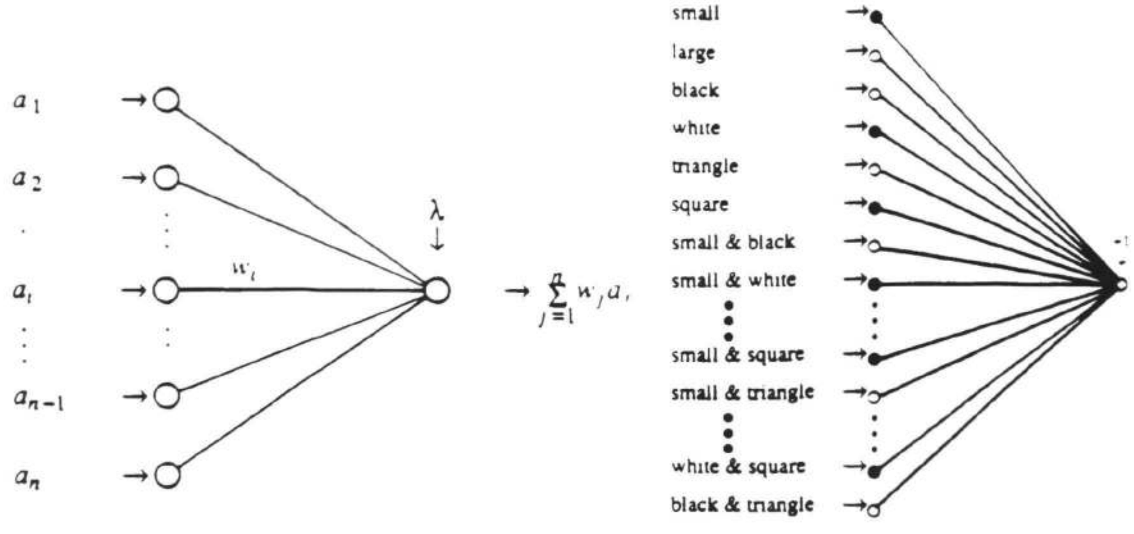

Figure 1. (A) A simple one-layer network which can learn the associations between three cues (CSs) and one outcome (US). (B) A configural-cue network with the cues for a small white square activated.

Equation 1 is variously called the delta rule, the least-mean-square (LMS) rule, or the Rescorla-Wagner conditioning rule (for a discussion, see Gluck & Bower, 1988a). Importantly, it corrects all weights according to the degree of error between the network's current output and what was desired for this pattern. This defines a learning process whereby the weights on the input lines converge to values that reflect the relative correlations of the stimulus features with the feedback signal. In a medical setting, these weights reflect the differential validity of each symptom (cue) for each disease (category). We have applied this baseline model to a variety of classification experiments (see Gluck & Bower, 1986, 1988a, 1988b). In each case, the simplest identifications have been used, viz., presentation of a specific medical symptom (e.g., stomach cramps) corresponded in the model to turning on a specific input node. Thus, a pattern of medical symptoms exhibited by a patient was represented by activation of the corresponding input nodes in the model. These identifications were successful in fitting the data of the early experiments by us and others (Estes et al., in press; MacMillan, 1987; Nosofsky, personal communication).

However, this approach, of theoretically identifying each experimental stimulus cue with a single input node in the model, encounters several difficulties. Most familiarly, one-layer networks with such manifest stimulus identifications are incapable of learning classifications that are not "linearly separable". An example is the exclusive-or (XOR) problem, wherein stimulus patterns (0,0) and (1,1) belong to one category, while patterns (1,0) and (0,1) belong to another. A common approach for solving such non-linear classification problems is to postulate additional, "hidden units" which connect between the input and output units (Parker, 1986; Rumelhart, Hinton, & Williams, 1986). While these multi-layer networks have great power for learning complex discriminations, they are insufficiently constrained to serve yet as testable, psychological models of simple learning. They require large numbers of assumptions regarding their structure (e.g., the basic representation of stimuli and responses, the number and connectivity of hidden units, etc.), their learning rule, and their method for calculating response probabilities.

For such reasons, we preferred initially to explore the viability of a simple extension of the elementary model, one which postulates that conjunctions of elementary stimulus features can serve as "higher-order" features of a stimulus pattern. Thus, given the presentation of an experimental pattern consisting of elementary features BCD, we will assume that this is reflected in activation of input nodes corresponding to the single elements B, C, and D, and the pair-wise conjuncts BC, BD, and CD. As another illustration, Figure 1B shows a network that is learning to classify geometric patterns varying in size, color, and shape: presentation of a "small white square" causes activation of the input nodes blackened in the figure for single and pair-wise cues.

We will assume that such "configural" features obey the same activation and learning rules as do the single features, viz., Eq.1. The inclusion of such configural features as "inputs" now enables the one-layer model to learn the XOR problem as well as other non-linearly separable discriminations.

To include configural cues is hardly a novel move for theories of discrimination learning. Learning theories have traditionally recognized configural learning (Pavlov, 1927; Woodbury, 1943). Wagner and Rescorla (1972, p. 306) explicitly expanded their theory of conditioning to include configural cues; and in a series of studies, Rescorla (1972, 1973) found that configural cues have many of the same associative properties as single cues. In particular, Rescorla found that configural cues can acquire both excitatory and inhibitory associations, that their associative strengths summate with those of single cues to determine behavior, that configural cues can modify the effectiveness of a given reinforcing event, and that their strength can be attenuated by making them irrelevant to the discrimination being trained. Thus, our introduction of configural cues into the one-layer network is supported by a considerable history.

We will impose one arbitrary limitation upon the configural cue model tested in this paper, namely, that only *pair-wise* conjunctions of elementary features will be allowed. An alternative, proposed by Reitman & Bower (1973), Hayes-Roth & Hayes-Roth (1977), and Gluck & Bower (1988a), is to introduce as higher-order features the entire power-set of all subsets of each n-dimensional stimulus presented in the experiment. This power-set model rapidly becomes unwieldy, so we have restricted our explorations to the pair-wise conjunction version of it.

In the following, the predictions of the configural cue model are compared to the data from three representative, critical experiments from the literature on human classification learning. The fit of the configural cue model to the observed data will be compared to the fit of two other models: (1) the single-cue-only model, and (2) an alternate extension of the network model recently proposed by Estes (in press). Estes suggested using as inputs only the single cues and the full patterns, so that presentation of BCD would activate input nodes B, C, D, and (BCD). We will call this the "feature-pattern" model.

LINEAR SEPARABILITY IN CLASSIFICATION LEARNING

We provide three illustrations extending the configural-cue model to account for several aspects of complex human category learning. First, the inability of the simple network model to solve non-linearly-separable classifications has historically been a major reason for introducing configural cues into one's theory. Therefore, we wished to apply the configural-cue model to such a non-linear learning task. An experiment by Medin & Schwanenflugel (1981) provides relevant data. Figure 2 schematizes the 6 stimulus patterns that two groups of college students learned to classify as A's or B's. The two values of the four stimulus dimensions are denoted 1 and 0. To recognize the linear separability of the left-hand classification, note that the number of 1's in Dimensions 1, 3, and 4 equal 2 for the A-stimuli, but is less than 2 for any B stimulus; however, no such linear combination of feature valves will separate the two classes of patterns in the right-hand classification. Note too that the two classifications are perfectly balanced in terms of the average number of shared features among patterns *within* each class (average of 1.33 shared features) and shared features of patterns *across* different classes (average of 1.78). Medin and Schwanenflugel found that their subjects learned this nonlinearly-separable problem more easily than their linearly-separable problem (see Figure 3A). Their model predicted this because it calculates the similarity of two patterns in a nonlinear fashion, so that confusions of a test pattern with memories of two A-patterns with which it shares 1 and 3 features will be much greater than its confusions with two B-patterns with which it shares 2 features each.

We attempted to simulate the Medin & Schwanenflugel results with the network model using three different representations of the stimuli: the single-cue (baseline) model, the pair-wise configural-cue model, and the feature-pattern model. In all the simulations, we used a learning rate of $\beta = 0.01$, one output node, and reinforced the network with $\lambda = +1$ for category A exemplars, and $\lambda = -1$ for category B exemplars. Each network had two cue nodes for each dimension -- one node represented the presence of a cue, the other its complement. The configural-cue network had additional nodes for all pair-wise combinations of feature values.

	Linearly Separable Task		*Non-linearly Separable Task*	
	Exemplar	Dimension	Exemplar	Dimension
		1 2 3 4		1 2 3 4
Category A	A1	0 1 1 1	A1	1 1 0 0
	A2	1 1 1 0	A2	0 0 1 1
	A3	1 0 0 1	A3	1 1 1 1
Category B	B1	1 0 0 0	B1	0 0 0 0
	B2	0 0 0 1	B2	0 1 0 1
	B3	0 1 1 0	B3	1 0 1 0

Figure 2. Classification tasks in Medin & Schwanenflugel (1981), Experiment #3.

Figure 3B shows the average mean squared error for each training epoch for the single-cue model. The average MSE for the single-cue model trained on the non-linearly-separable task never reaches zero, meaning that this discrimination is not perfectly learnable by the single-cue model. The pair-wise configural-cue model does, however, predict the correct ordering of the results: it learns the non-linear task faster than the linear one (Figure 3C). Thus, the addition of feature pairs to the input nodes improved the network performance. These theoretical results suggest that the configural-cue model, like Medin & Schaffer's context model, is more sensitive to exemplar similarity (as computed by a non-linear multiplicative similarity rule) than to the linear separability of the patterns in the different categories. As noted by Nosofsky (1984), the multiplicative similarity rule is equivalent to assuming stimulus generalization is an exponential decay function of psychological distance, the latter indexed by the number of featural mismatches. This exponential relationship between similarity and psychological distance has received substantial independent empirical and theoretical support (Shepard, 1957, 1987). That the configural-cue model embodies the same similarity-distance relationship can be seen by computing how the number of overlapping active nodes (similarity) changes as a function of the number of overlapping component cues (distance). If two triplet patterns share one feature (ABC, XYC), they will have only one active node in common and five nodes nonoverlapping; if they share two features (ABC, XBC), they will have three active nodes in common (two component cues and one configural-cue node) and three nonoverlapping nodes; if they share three features in common, they will have six active nodes in common (three component cues and three configural-cue nodes). This implies that the configural-cue network, like the context model, will judge a test pattern to be more similar to a category of two exemplars with which it shares 1 and 3 features (for an average of 3.5 nodes in common), than an alternate category of two exemplars with which it shares 2 features each (for an average of 3 nodes in common).

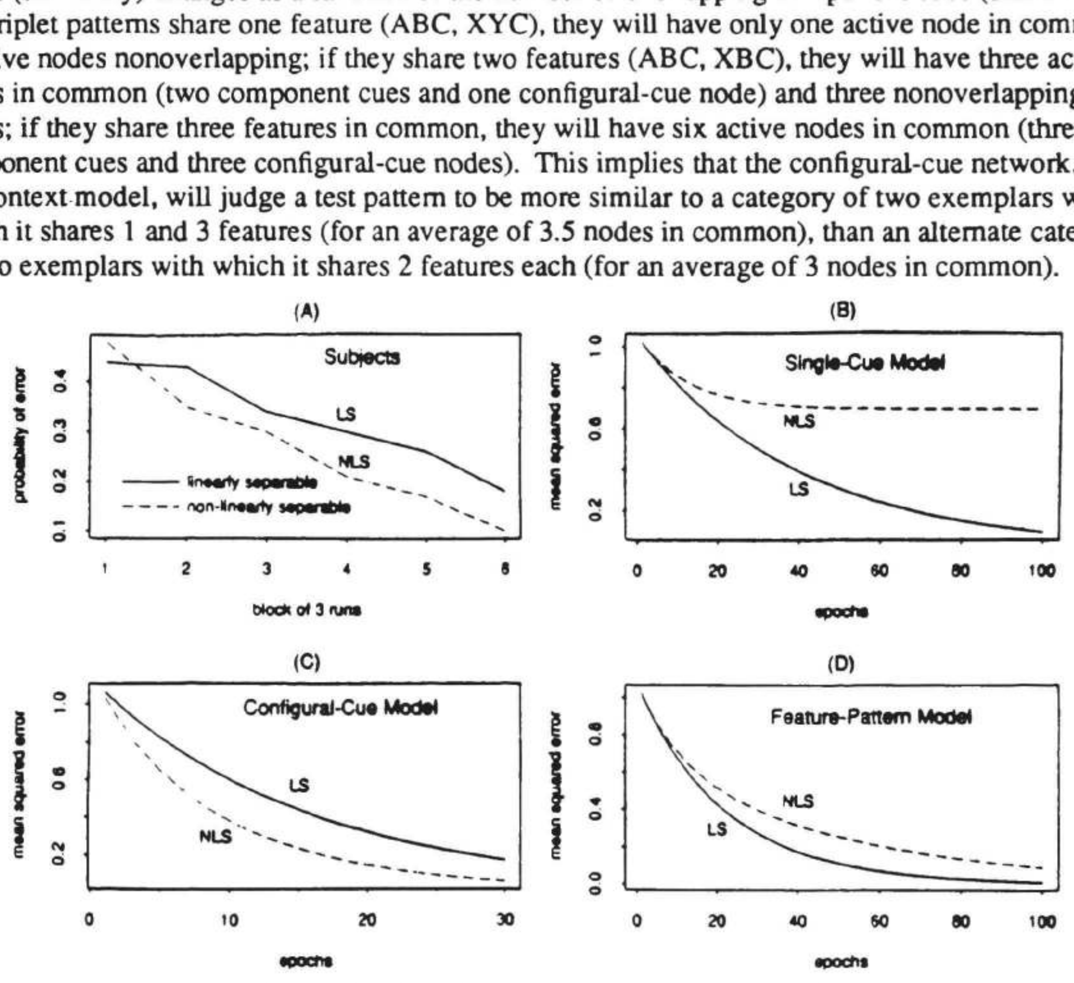

Figure 3. Predicted difficulty of non-linearly versus linearly separable classification tasks in Medin & Schwanenflugel (1981), Experiment #3. LS: linearly separable classification task. NLS: non-linearly separable task. The mean squared error (MSE) represents the absolute difference (squared) between the actual and predicted category classifications averaged over all presentation exemplars. Task difficulty is predicted by the rate at which the model reduces the MSE to zero. (A). The data on percentage errors, showing that the LS problem is more difficult (slower to learn); adapted from Medin & Schwanenflugel (1981). (B). The incorrect predictions of the one-cue network model showing that only the LS task is learnable. (C) The closer predictions of the "pair-wise" configural-cue model showing that the LS category is more difficult (slower to learn). (D). The less accurate predictions of the "feature-pattern" model.

Interestingly, the feature-pattern model which uses only single cue plus full patterns (see Fig. 3D) mispredicts the ordering of the data. Although the addition of nodes representing entire patterns allows the model to learn complex, non-linear discriminations, it expects the non-linearly separable task to be learned more slowly than the linearly-separable one--contrary to fact.

RECOGNITION MEMORY VERSUS CLASSIFICATION

In testing models of category learning, we may examine how the classification of a given test pattern depends on the subjects' remembering specific exemplars that were shown during training. Such "recognition memory" can be tested by asking subjects to judge whether each test pattern is an "old" training instance, or a "new" instance not experienced before. Prototype theories, which assume that people extract only a mean centroid from the training exemplars, expect a strong correlation between the classification and "old" judgments for test exemplars, since both decisions could presumably only be based on the distance of the exemplar from the prototype.

An experiment by Hayes-Roth and Hayes-Roth (1977) examined this issue; they found a surprisingly low correlation between subjects' classifications and their Old (vs. New) judgments over a variety of test patterns. It was of interest to see whether the configural-cue model could duplicate this surprising lack of correlation between classification and recognition memory.

In the Hayes-Roth & Hayes-Roth task, subjects learned to classify into three categories (Club 1, Club 2, or Neither) descriptions of people who varied along three dimensions with four values per dimension (labeled 1-4). The presence of a majority of 1's with no 4's *(e.g., 112,131)* signified membership in club 1, whereas a majority of 2's with no 4's *(e.g., 212, 221)* indicated club 2. An equal number of 1's and 2's with no 4's indicated membership in either category. If any 4's were present, the person belonged to neither category. The 3's were irrelevant. Specific patterns ("persons") were presented with widely varying frequencies. For example, the most prototypical category members *(e.g., 111, 222, 333, 444)* were never presented during the training phase; however, they were shown on subsequent recognition and classification tests.

As noted, Hayes-Roth & Hayes-Roth found that classification of an exemplar correlated poorly with its recognition. For instance, subjects gave the non-presented category prototypes *111* and *222*, the highest classification ratings; in contrast, these prototypes were rarely recognized as "old" training instances. Also, certain exemplars which were presented often during training *(e.g., 112, 121)* received the highest recognition ratings, but weaker classification responses.

Nosofsky (1988) suggested that an exemplar model could predict these results if recognition was based on the summed similarity of that pattern to *all* stored exemplars. Using this rule within his model, Nosofsky fit the Hayes-Roth & Hayes-Roth data. His amended model correctly predicted both the high classification of category prototypes *111* and *222*, along with the high recognition of frequently-presented exemplars (see Figures 4A and 4B).

We fit the single-cue and configural-cue models to the data of the Hayes-Roth and Hayes-Roth experiment. As before, the exemplars were presented in a random order to the network (β = 0.01) for one complete pass through all the exemplars. The network had four input nodes (cue values) for each of the three dimensions (12 total) connected to three output nodes (Club 1, 2 or Neither). The probability of assigning a test pattern to Club 1 vs Club 2 was set equal to the strength of the one output activation divided by the summed strength of the output activations for both clubs (with negative activation values being converted to 0). In contrast, the recognition-memory rating of a test pattern was predicted from the summed activation of all three output nodes (including the Neither node) to that pattern, which is a rule similar to Nosofsky's (1988). Figure 4A illustrates the network model's recognition and classification responses to test patterns (averaged over 10000 simulations),

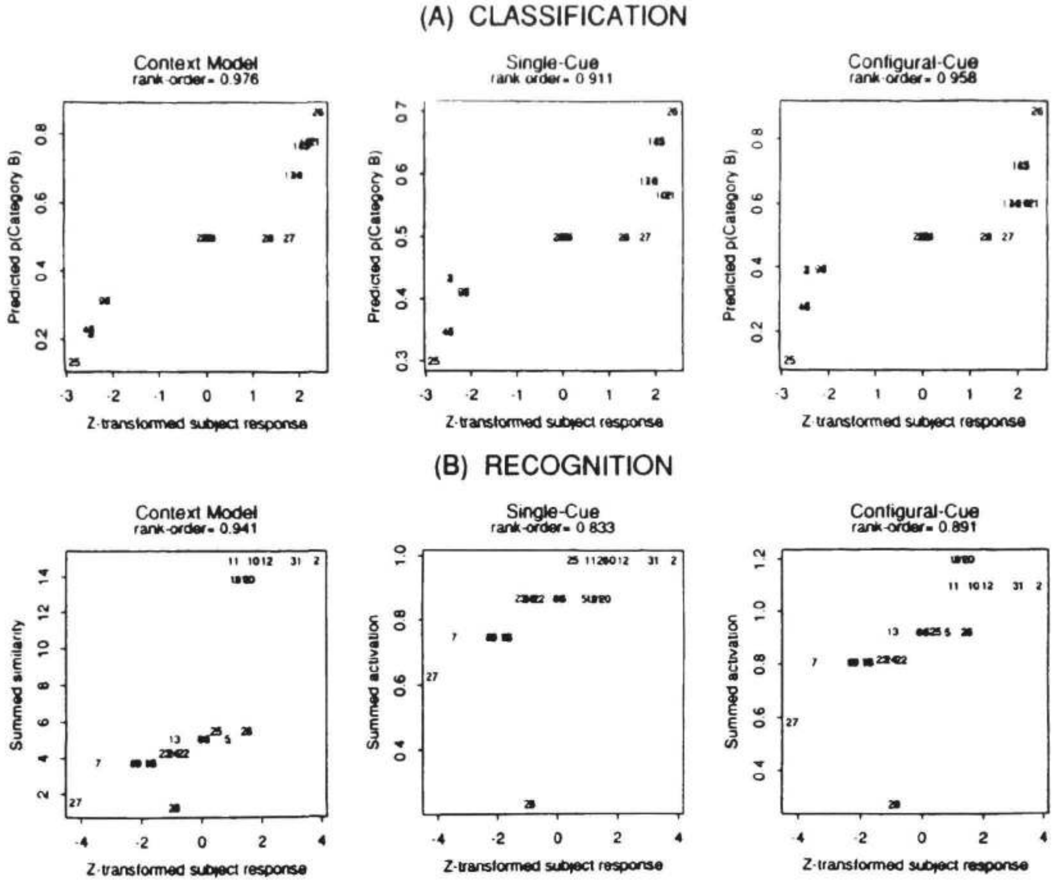

Figure 4. Predicted responses of subjects in the Hayes-Roth & Hayes-Roth (1977) learning task. Predictions of various models are plotted against subject's performance (z-scores) for both the classification (4A) and the recognition (4B) phases of the experiment. The Spearman rank-order correlation between a model's predictions and the subjects' performance is reported. Each number on each plot represents the z-score for classification (4A) or for recognition (4B) for one of the 28 test patterns. (A) Classification ratings for Nosofsky's context model, the single-cue coding model, and the configural-cue network model. (B) Recognition ratings for the three models.

plotted against the observed ratings (transformed z-scores) by subjects. While the single-cue, baseline model correctly predicted subject's classificatory responses (a rank order coefficient of 0.91), it predicted recognition memory less successfully (rank order coefficient of 0.83).

In contrast, the configural cue model was more successful overall. Figure 4 shows its predictions for these data. Predictions of classificatory responses are accurate (rank order correlation = 0.96); importantly, the accuracy of recognition predictions improves over that of the single-cue model (rank order correlation = 0.89). Thus, the configural-cue model accounts for both classification and recognition memory with only a single parameter, viz., the learning rate, β.

Despite this overall success, the configural cue model evidences shortcomings similar to Nosofsky's. Both models, for example, predict a much lower recognition of the "Neither" prototype *444* than was actually obtained. Similarly, both models predict chance classification of the "Neither" (*444*) and the "Unknown" (*333*) prototypes, whereas subjects were biased towards one particular category. Examination of the data reveals no reason for these discrepancies.

Finally, we compared the predictions of Estes' feature-pattern encoding model to the results of the configural-cue network for the Hayes-Roth & Hayes-Roth data. The feature-pattern model's predictions for both sets of data were very similar to those for the pair-wise configural cue model, and yielded no discriminating comparisons.

CORRELATED ATTRIBUTES AND CATEGORY LEARNING

An obvious limitation of the single-cue model is that it is insensitive to the predictive validity of pairs of features. The weights attached to each single cue reflect the associations between it and the several categories, but these cannot capture correlations between cue-combinations and the categories. People, on the other hand, are sensitive to predictive combinations of features. Medin, Altom, Edelson, & Freko (1982) tested subject's use of combinations of symptoms in a simulated medical classification task. Their Experiment #3 put people's classification of patterns according to co-occurring features into opposition to their tendency to classify patterns according to the number of singly representative cues. Subjects first learned to classify patterns of symptoms into a single disease category. Each pattern consisted of five binary dimensions; these are illustrated in Figure 5A where a '1' or a '0' on each dimension indicated a symptom value or its complement.

The fourth and fifth symptom dimensions were perfectly correlated with each other. Also, for any dimension, the total number of '1's across presented patterns exceeded the total number of '0's. Thus, the presence of a '1' in a particular dimension indicated its more typical or characteristic value. The goals of the study were (1) to assess whether people would use the correlation between symptom-dimensions four and five to classify instances, and (2) to see how this information would be combined with information about the typicality of the individual features to determine choice.

Subjects studied the individual cases shown in Figure 5A and subsequently received transfer test pairs containing both new and old patterns (Figure 5B). For each transfer test pair, subjects had to decide which exemplar was more likely to be a member of the category defined by the collection of training instances in Figure 5A . On the critical transfer tests, subjects chose between some exemplar preserving the relationship between the fourth and fifth dimensions versus another exemplar that violated this correlation but had more characteristic features (more '1's).

Because the single-cue model, like all independent cue models, considers each feature separately, it predicts that subjects will select the transfer pattern containing more characteristic attributes as the more likely member of the category. However, the data showed that people preferred the pattern containing the correlated features as more likely to be a member of the category. Thus, even though a test pattern had fewer diagnostic features present, subjects were more likely to say it was a member of the category when the fourth and fifth symptoms preserved the correlation presented during training

(A)

Exemplar	Dimension 1 2 3	 4 5
a	0 1 0	1 1
b	1 1 0	1 1
c	0 0 1	1 1
d	1 0 1	1 1
e	1 1 1	1 1
f	1 1 1	1 1
g	1 0 0	0 0
h	0 1 1	0 0
i	1 1 1	0 0

(B)

Exemplar A preserved correlation	network activation		Exemplar B more '1's	network activation
1 1 1 **0 0**	*1.017*	vs.	**1 1 1** 0 1	0.923
0 0 1 **1 1**	*0.992*	vs.	**1 1 1** 0 1	0.923
0 1 0 **1 1**	*0.991*	vs.	**1 1 1** 1 0	0.923
0 0 1 **0 0**	*0.914*	vs.	0 0 **1** 0 **1**	0.860
1 0 0 **0 0**	*0.981*	vs.	**1** 0 0 **1** 0	0.887
average	*0.919*		average	0.903

Figure 5. Schematic design of Medin, Altom, Edelson, & Freko (1982), Experiment #3. (A) Training exemplars. A '1' on a particular dimension indicates its more common, or characteristic, value. Dimensions 4 and 5 are perfectly correlated with each other and with the correct category. (B) Transfer choice test pairs. After training, subjects were presented with each choice test pair and asked to choose the exemplar most likely to be a member of the collection described by (A). The choice tests compared exemplars preserving the correlation between dimensions 4 and 5, to those that violated the correlation, but contained more characteristic values (more '1's). In all choice tests the configural model correctly predicts that people will prefer the exemplar preserving the correlation between dimensions 4 and 5 as a more likely member of the category.

Our simulation of this experiment with the configural cue model used all 10 single-cue and all 32 cue pairs as input nodes linked to one output node representing category membership. Since all presented exemplars were members of the category, all presentations were consistently reinforced ($\lambda = +1$). Figure 5B shows that the output activation of the configural-cue model is greater for the exemplar that preserves the correlation between dimensions four and five compared to the activation produced by the exemplar with more characteristic attributes. Because the network's output activation translates into choice probability, the simulation will correctly predict that subjects will prefer those patterns that preserve the correlation in the transfer choice tests. The model expects this result because feature-conjuncts (4 & 5) are perfect predictors of category membership whereas single cues are imperfect predictors; in such cases, the competitive nature of the LMS learning rule implies that a more valid predictive feature (or conjunct) will dominate and beat down the learning of less valid features. This phenomenon, called "overshadowing", is familiar in conditioning studies.

The ability of the configural cue model to predict this configural-cue preference found by Medin et al. is not completely trivial, because the predictions depend on the balance of associative strength to the conjunct cues versus the more characteristic, single cues. Several plausible models do not calculate the balance of these factors appropriately. For instance, we applied to these data Estes' feature-pattern model which has nodes representing the presence of entire patterns as well as single features. Although this is one way to add configural pattern information into the learning process, the outcome was unsuccessful in this case: in four of the five transfer tests, the feature-pattern model expected subjects to prefer that stimulus with the greater number of characteristic features (1's) to the one preserving the correlation of features 4 and 5.

DISCUSSION

We have also applied the configural-cue model to explain and predict the priority of basic levels in category hierarchies, and this is reported elsewhere (Corter, Gluck, & Bower, 1989). The configural-cue model predicts that the basic-level categories of a hierarchy of categories are learned more quickly than other levels, and examples are recognized faster at this level. These results are consistent with much empirical data regarding both natural and artificially-learned categories (Jolicouer, Gluck, & Kosslyn, 1984; Corter, Gluck & Bower, 1988). In Gluck & Bower (1988a), we also applied the configural-cue model to a classic experiment by Shepard, Hovland, & Jenkins (1961) who studied the difficulty subjects had in learning six classifications varying in complexity. The model predicted the same order of difficulty of learning the classification rules as was revealed in the data, except for one slight misordering.

By expanding the representation of stimuli to include pair-wise configurations of features, the network model appears to account for a wider range of learning results from both the animal and human learning literatures. Some of this success can be traced to its using a similarity metric like that of Medin & Shaffer, viz., an implicit exponential decay relationship between stimulus similarity and psychological distance (number of feature mismatches). The configural-cue model has several obvious limitations, including the exponential growth of input nodes with increasing pattern size. Nevertheless, we believe that this model is interesting for four reasons. First, it is simple, understandable, and accounts for a surprisingly wide range of empirical phenomena. Second, it is theoretically parsimonious and uses assumptions for which independent evidence already exists. Third, its successes are instructive in identifying empirical phenomena which can be explained as emergent from the same elementary, associative processes found in lower species. Fourth, explanations of the failures of this model can suggest more sophisticated versions of the network model. Such failures may also indicate performances arising from an entirely different class of learning mechanisms, i.e., the rule-based or symbolic processes which have been well studied by cognitive psychologists.

REFERENCES

Corter, J. E., Gluck, M. A., & Bower, G. H. (1988). Basic levels in hierarchically structured categories Montreal, Canada. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society, Montreal, Canada.*. Hillsdale, NJ: Lawrence Earlbam Associates.

Corter, J. H., Gluck, M. A., & Bower, G. H. (1989). *Basic levels in hierarchical category structures: An adaptive network interpretation*. Unpublished Manuscript, Stanford University, Stanford, CA 94305..

Estes, W. K., Campbell, J. A., Hatsopoulos, N., & Hurwitz, J. B. (in press). Base-rate effects in category learning: A comparison of parallel network and memory storage-retrieval models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, .

Gluck, M. A., & Bower, G. H. (1986). Conditioning and categorization: Some common effects of informational variables in animal and human learning. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society.*. Amherst, Mass..

Gluck, M. A., & Bower, G. H. (1988a). Evaluating an adaptive network model of human learning. *Journal of Memory and Language, 27*, 166-195.

Gluck, M. A., & Bower, G. H. (1988b). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117*(3), 225-244.

Gluck, M. A., Corter, J. H., Bower, G. H., & Kylberg, R. L. (1988). *Learning of basic levels in hierarchically structured categories.*. Presented at the Annual Conference of the Psychonomic Society, Chicago, IL.

Hayes-Roth, B., & Hayes-Roth, F. (1977). Concept learning and the recognition and classification of exemplars. *Journal of Verbal Learning and Verbal Behavior, 16*, 321-338.

Jolicoeur, P., Gluck, M., & Kosslyn, S. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16*, 243-275.

MacMillan, J. (1987). *The role of frequency memory in category judgments*. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 8*, 37-50.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear seperability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 355-368.

Nosofsky, R. (1988). Similarity, frequency, and category representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 54-65.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 104-114.

Parker, D. (1986). A comparison of algorithms for neuron-like cells. In *Proceedings of the Neural Networks for Computing Conference*. Snowbird, Utah..

Pavlov, I. (1927). *Conditioned Reflexes*. London: Oxford University Press.

Reitman, J. S., & Bower, G. H. (1973). Storage and later recognition of exemplars of concepts. *Cognitive Psychology, 4*, 194-206.

Rescorla, R. A. (1972). "Configural" conditioning in discrete-trial bar pressing. *Journal of Comparative and Physiological Psychology, 79*(2), 307-317.

Rescorla, R. A. (1973). Evidence for "unique stimulus" account of configural conditioning. *Journal of Comparative and Physiological Psychology, 85*(2), 331-338.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*. New York: Appleton-Century-Crofts.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propogation. In D. Rumelhart, & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition (Vol. 1: Foundations)*. Cambridge, M.A.: MIT Press.

Shepard, R. (1987). Towards a universal law of generalization for psychological science. *Science, 237*, 1317-1323.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika, 22*, 325-345.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75*, 1-42.

Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Applications of a theory. In R. A. Boakes, & S. Halliday (Eds.), *Inhibition and learning* (pp. 301-36). New York: Academic Press.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. *Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record, 4*, 96-194.

Woodbury, C. B. (1943). The learning of stimulus patterns by dogs. *Journal of Comparative Psychology, 35*, 29-49.