Title
Exemplar Competition: A Variation on Category Learning in the Competition Model

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 14(0)

Authors
Taraban, Roman
Palacios, J. Marcos

Publication Date
1992

Peer reviewed

Exemplar competition: A variation on category learning in the Competition Model*

Roman Taraban
Department of Psychology
tirmt@ttacs.ttu.edu
J. Marcos Palacios
Computer Science
cjejp@ttacs.ttu.edu
Texas Tech University
Lubbock, TX 79409

Abstract

Two cue validity models for category learning were compared to the exemplar model of Medin & Schaffer (1978). The cue validity models tested for the use of two cue validity measures from the Competition Model of Bates & MacWhinney (1982, 1987, 1989) ("reliability" and "overall validity"); one of these models additionally tested for "rote" associations between items and categories. Twenty-four undergraduate subjects learned to classify pseudowords into two categories over 40 blocks of trials. The overall fit of the cue validity model without rote associations was poor, but the fit of the model that included these was nearly identical to the exemplar model (R^2 = .89 *vs* .90). However, both cue validity models failed to capture differences predicted by exemplar similarity, but not cue validity, that were apparent as early as the first block of learning trials. The critical parameters in the Medin-Schaffer model were fit as a logarithmic function of the learning block to provide a uniform account of learning across the 40 blocks of trials. The evidence that we provide suggests that competition at the level of exemplars should be considered as a possible extension of the Competition Model.

*This paper is based in part upon work supported by the Texas Advanced Research Program under Grant No. 0216-44-5829 to the first author.

Models of category learning have appeared in at least two distinct guises. *Independent-cue* models (Anderson, 1991; Beach, 1964; Reed, 1972; Rosch & Mervis, 1975) posit the summing of weighted "evidence" for a category derived from information provided by individual cues or features. *Exemplar* models (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1984) usually require the analysis of exemplars into simpler components, but compute the evidence for a category on the basis of between-*item* similarity.

The *Competition Model* of Bates and MacWhinney (1982, 1987, 1989) is an independent cue model that has been quite successful in accounting for the learning of natural language categories. An important thesis in this model is that children and adults weight cues differently depending on their level of learning. These differences are described through various cue validity measures that assess the relative contribution of a cue to category selection. Taraban, McDonald, & MacWhinney, 1989), for instance, used human and computer simulation data to argue that *overall validity* provides the best characterization of cue weights early in learning; later in learning the weights are best described by *reliability* and then by a least-mean squares solution. McDonald & MacWhinney (1991) have provided evidence for early use of *overall validity* and later reliance on *conflict validity*.

Although the Competition Model provides the best current account for learning linguistic categories, the exemplar view has not been explored and it is still not known whether the Competition Model could benefit from an exemplar approach. In this paper we are not concerned with the standard Competition Model questions that focus on shifts in weights of independent cues. Instead we set up contrasting predictions for independent-cue models and an exemplar model in a learning experiment to test whether the exemplar model provides a better fit to performance at *any* stage in learning. In an experimental setting, it is difficult to systematically explore language learning with natural language materials, so in some of these studies the experimenters have resorted to using artificial materials (e.g. McDonald & MacWhinney, 1991). We have adopted the same approach in the present study using a very simple set of pseudowords for which subjects learned category labels over the course of a single, long, experimental session.

Three models: Cue Validity, Cue + Rote, Exemplar

Reliability is closely related to formulations in Beach (1964) and Reed (1972). For any given cue and a category X, reliability corresponds to the conditional probability $P(X|cue)$. In the *Cue Validity* model, we fit one parameter for each letter position in the pseudoword stimuli to allow for differences in attention to cue reliabilities in those positions. As indicated in (1), the "evidence" for some category X given a test item t is a weighted sum of cue reliabilities. *Overall validity* corresponds to the product of the overall frequency of a cue and its reliability. In the context of the present study, it is important to point out that the overall frequency of each cue was 0.5. Thus, a fitted *overall validity* model differs from a *reliability* model by a constant factor – i.e. we could fit the *overall validity* model directly from (1) by simply multiplying each fitted parameter by 2. This means that (1) should give a good account of a substantial part of learning performance, based on current Competition Model thinking.

$$E_{X|t} \equiv \sum a_i * \text{reliability}_i \quad (1)$$

Is a weighted model like (1) sufficient for describing category learning? Clearly it is not, particularly if the categories are "non-linearly" separable, a condition which by definition precludes complete learning. MacWhinney, Leinbach, Taraban, & McDonald (1989) discuss the possibility that cue-to-category associations like those represented in (1) are supplemented by "rote" associations of items to their respective category. The *Cue + Rote* model discussed in this paper is identical to (1), except that the sum includes an additional product (a_i * item) that estimates the strength of association of pseudowords to their respective categories, with the value of item equal to 1 for its association to its own category, and 0 for its association to the competing category. Does adding a parameter for rote associations render the reliabilities superfluous? The answer is "no." If subjects simply learned "paired associations" there would be no between-item differences in fit to a category (viz. typicality), which is, in general, unlikely for categories and not the case for our stimuli, as described later.

The *Exemplar* model presented in (2) is the one used in Medin & Schaffer (1978). In this paper, (2) computes the *overall similarity* of an item t to a category X. $Similarity(t, x) = \prod s_i$, with an s_i fitted for each letter position, computes the similarity of an item t to a particular category member x. As in Medin & Schaffer, $s_i = 1$ if $letter_i$ in x and in t match, and $0 \leq s_i \leq 1$ if they mismatch. In the tests done by Medin & Schaffer (1978), independent-cue models that did include item-level (rote) information generally did not appear to do more poorly than the exemplar model, motivating a further examination here of both types of models.

$$E_{X|t} \equiv \frac{\sum_{x \in X} Sim(t,x)}{\sum_{x \in X} Sim(t,x) + \sum_{y \in Y} Sim(t,y)} \quad (2)$$

In order to compare the models, we chose to use an instantiation of Type V stimuli in Shepard, Hovland, & Jenkins (1961). This set was important since cue validity and exemplar similarity predict different patterns of performance

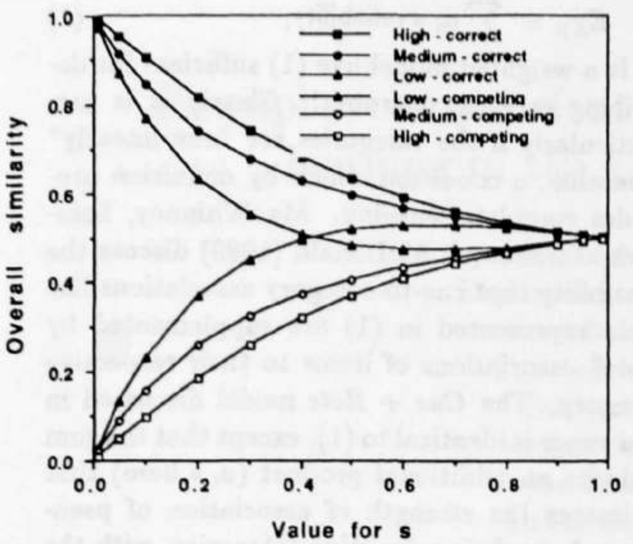

Figure 1: Overall similarity values for stimuli in Table 1, using (2).

across the learning trials. First, as shown in Figure 1, similarity calculations for the stimuli in Table 1 result in three groups, which we will term the *high-*, *medium-*, and *low-similarity* groups. The stimuli fall into these three groups for any value of s between 0 and 1, where s is the parameter estimated for (2) above. A sample set of similarities is shown in Table 1 for $s = \frac{1}{e}$. On the other hand, the sum of cue validities for each item in Table 1, for $a_i = 0.33$, shows that cue validities result in only two distinct groups. This is true whether the cue validity measure is "reliability" or "overall validity," as explained above.

Pseudo–word	*Category Label*	*Overall Sim*	$\sum$ *Cue Validity*
zub	*Jets*	.70 (.30)	.58 (.42)
zud	*Jets*	.64 (.36)	.58 (.42)
zob	*Jets*	.64 (.36)	.58 (.42)
vod	*Jets*	.51 (.49)	.42 (.58)
vub	*Sharks*	.70 (.30)	.58 (.42)
vud	*Sharks*	.64 (.36)	.58 (.42)
vob	*Sharks*	.64 (.36)	.58 (.42)
zod	*Sharks*	.51 (.49)	.42 (.58)

Table 1. The *overall similarities*, using (2) and $s = \frac{1}{e}$), and *cue validities*, (using $a_i = 0.33$), are for the item's category; the value for the competing category is shown in parentheses.

The crucial comparison in this experiment was between the ***high similarity*** (zub, vub) and ***medium similarity*** (zud, zob, vud, vob) groups. Using the estimates shown in Table 1, the *Exemplar* model predicts a difference between these groups, based on their relative *similarities*. Neither the *Cue Validity* model nor the *Cue + Rote* model predicts a difference, and, in fact, there is no set of parameters for these two models that could separate the items into the *high* and *medium* subsets. In this experiment we tested to see whether the exemplar model provided a better fit to the data than either of the cue validity models at any point in learning.

Method

Subjects. Twenty-four undergraduates participated in this experiment for course credit.

Stimuli. The stimuli are shown in Table 1. Each category consisted of 4 three-letter pseudowords, which were presented to subjects as codenames for gang members in the Jets and the Sharks.

Procedure. Each subject was presented with 40 blocks of trials on an IBM AT clone, with the pseudowords appearing in random order within each block. Subjects used a rating scale of 0-9 to indicate membership for both gangs – i.e. subjects rated the pseudoword twice on each trial. The order of ratings was random. Feedback was provided after each trial to indicate the correct gang. Subjects were warned that early on in the experiment they would know little about the gang membership, so they should avoid extreme ratings.

Results

Since subjects were instructed to use whole number ratings, a middle rating (4.5), important in the early trials, was not available to them, and subjects tended to begin with ratings of 5. In order to convert the ratings to the range 0-1, to correct for the artifact of the rating scale, and to assure that the sum of residuals in the analyses was 0, each rating was divided by 9 and then 0.069 was subtracted.

In the current experiment, items should elicit high ratings for the item's own correct category and low ratings for the competing category. An examination of Table 2 shows higher ratings for high- vs medium- vs low-similarity items for the items' correct category; similarly, lower ratings for high- vs medium- vs low-similarity items for the items' competing category. An ANOVA using Similarity (high, medium, low), Rating Type (either for its own category or for the competing category), and Block showed a significant effect for the crucial 2-way interaction in these data: Similarity X Rating Type [$F(2,46) = 6.58$, $p < .004$, by subjects; $F(2,5) = 6.69$, $p < .04$, by items]. Importantly, the effect of the 3-way interaction was non-significant [F-values < 1, by subjects and items]. This suggests that there was a significant difference between the *high*, *medium*, and *low* items and that the effect did not vary significantly across the blocks of trials. One-*df* F-tests were used to verify that there was a significant difference between the mean *high* and mean *medium* ratings for items' own category (.70 vs .66: $F(1,23) = 13.94$, $p < .002$, by subjects; $F(1,4) = 9.14$, $p < .04$, by items), and between the mean *high* and mean *medium* ratings for items' competing category (.29 vs .33: $F(1,23) = 7.61$, $p < .02$, by subjects; $F(1,4) = 6.03$, $p = .07$, by items). As is evident in Table 2, the differences between ratings for *high* and *medium* similarity items clearly emerges in block 1, at least for items' own category. (Subjects' mean ratings for all the blocks are shown in Figures 3A and 3B.)

Correct category	*High*	*Med*	*Low*
Overall	.70	.66	.62
Block 1	.55	.50	.41
Block 2	.63	.47	.46
Competing category			
Overall	.29	.33	.37
Block 1	.47	.47	.59
Block 2	.37	.53	.52

Table 2: Mean ratings. (High, medium, and low groups are based on the overall similarity estimates in Figure 1.)

Fit to models. Each of the models was first assessed on a block-by-block basis – basically, 40 regression analyses for each model – using the models specified at (1), (2) above. This was to allow for the most liberal fit of parameters for each model and was equivalent to 40 hypothetical experiments for which testing would simply occur at the n-th block after $n-1$ blocks of training. A comparison of the three models is shown in Figure 2 in terms of the residual error in the analyses done for each model at each block. The general result here is that all three models were quite close early on. After the first 5 blocks, the Cue Validity model began showing a clear disadvantage, and generally, the Exemplar model showed a slight advantage over the Cue + Rote model.

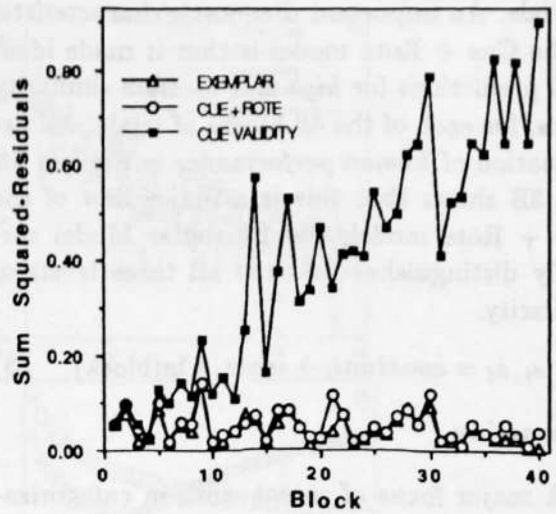

Figure 2: Fit of the three models.

To provide a uniform account of the learning that took place, we fit the data from all 40 blocks of trials by reinterpreting each s_i from the Exemplar model and each a_i from the Cue + Rote model as the logarithmic function in (3), with $constant_i$ defining the starting value for the redefined variable, s_i or a_i, and $lrate_i$ specifying how quickly it changes over the 40 blocks of trials. Figures 3A and 3B show the fitted Exemplar model, with (3) substituted for the s_is, superimposed on the human data. The overall fit of the model was excellent, with $R^2 = .90$. The overall fit of the Cue + Rote model (not shown) was similarly very good, with $R^2 = .89$. Figures 3C and 3D show how the reinterpreted s_i and a_i parameters change over the 40 blocks of trials. An important diagnostic characteristic of the Cue + Rote model is that it made identical predictions for *high* and *medium* similarity items, for each of the 40 blocks of trials. An examination of *human performance* in Figures 3A and 3B shows that this is a major flaw of the Cue + Rote model; the Exemplar Model correctly distinguishes between all three levels of similarity.

$$s_i, a_i = constant_i + lrate_i * \ln(\text{block}) \quad (3)$$

Discussion

A major focus of recent work in categorization has been on learning, and a compelling insight has concerned between-item similarity, as first described by Medin & Schaffer (1978). As learning proceeds, the s parameter in the Exemplar model goes to 0. This reflects a reduction in the contribution of stored items that are "similar" to the test item on the categorization outcome. In the limit, the influence of other items is nil. The Cue + Rote model helps us to distinguish between the process in the Exemplar model and the buildup of rote associations. If they were similar, we might expect the two models to converge at some point in learning, but they clearly do not when one uses the high- and medium-similarity items to monitor the behavior of the models.

A question that has interested us is how the three s values that we fit in Figure 3C contribute to the categorization rating. A cursory examination of the distribution of the letter values in the second and third positions shows the reliability (conditional probability) of these letter values to be 0.5 – i.e. they are distributed equally in both categories. The first letter position is the only informative one. Interestingly, when we computed the *predicted ratings* using only the fitted Exemplar model parameters for the second and third letter positions, they were uniformly 0.5 for each item in each block. This means that the work in the Exemplar model is being done by the first letter position. This is somewhat striking, since it shows that the Exemplar model is fully consistent with predictions about cue informativeness that would be made based on cue validities. Yet, it is not simply cue validities, as tested in the Cue Validity model, that are being computed. Rather, the Exemplar model goes deeper to uncover something about the human representations that cue validities cannot capture.

At this point, it is not clear how relevant these results will be to the Competition Model, which is meant to account for children's natural language learning. It could indeed be the case that children do tend to pick up independent cues and over time organize these into a dominance hierarchy, as suggested recently by McDonald & MacWhinney (1991). Given the present result, though, it would seem worthwhile to consider the notion of competition from the perspective presented here.

The Exemplar model provides a mathematical formulation for category learning. It provides some insight into the characteristics of a process model, however, nothing nearly as complete as a blueprint. At this point it would be important to look at available models that have in recent tests demonstrated an excellent ability to model category learning problems of the sort presented here. Two models that we have in mind are the "backpropagation" model of MacWhinney, et al. (1989) and Kruschke's ALCOVE (1992). From our current perspective we can only speculate that the ability of models in this class to effectively model human data may depend crucially on the characteristics of "hidden units"–i.e. that part of the model that plays a major role in internal representations that the model processes.

Acknowledgments

We are indebted to Jerry Myers and Steve Dopkins for some of the original ideas for this research and to Yiannis Vourtsanis, and Vir Phoha for helpful discussions. We would also like to thank Sandra Douglas, Chris McGee, Mukesh Rohatgi, and Mark Stephan for help in organizing and running the experiment and in analyzing the data. Finally, our thanks to Bob Bell, Brian MacWhinney, Janet McDonald, Jerry Myers, Glenn Nakamura, and two anonymous conference reviewers for helpful comments on an earlier draft of this paper.

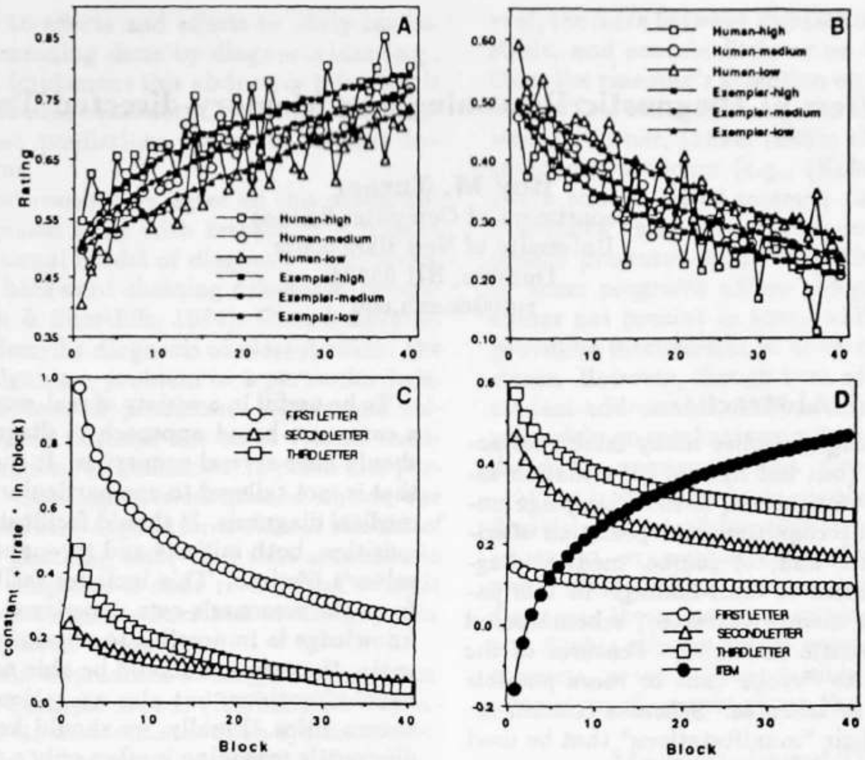

Figure 3: A: Data and model for correct category; B: data and model for competing category; C: plot of changes in parameter values for Exemplar Model; D: plot of parameters for Cue + Rote Model.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*, 409-429.

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art.* Cambridge, UK: Cambridge University Press.

Bates, E. & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition.* Hillsdale, NJ: Erlbaum.

Bates, E. & MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing.* New York: Cambridge University Press.

Beach, L. (1964). Cue probabilism and inference behavior. *Psychological Monographs, 78*, 21-37.

Kruschke, J. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22-44.

MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language, 28*, 255-277.

McDonald, J. & MacWhinney, B. (1991). Levels of learning: A comparison of concept formation and language acquisition. *Journal of Memory and Language, 30*, 407-430.

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

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

Reed, S. (1972). Pattern recognition and categorization. *Cognitive Psychology, 4*, 382-407.

Rosch, E. & Mervis, C. (1975). Family resemblance studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605

Shepard, R., Hovland, C., & Jenkins, H. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied, 75*, Whole No. 517.

Taraban, R., McDonald, J., & MacWhinney, B. (1989). Category learning in a connectionist model: Learning to decline the German definite article. In R. Corrigan, F. Eckman, & M. Noonan (Eds.), *Linguistic categorization* (pp. 163-193). Philadelphia: Benjamins.