Title
Abstraction versus Selective Attention in Classification Learning

Permalink
https://escholarship.org/uc/item/725271ft

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27)

ISSN
1069-7977

Author
Kurtz, Kenneth J.

Publication Date
2005

Peer reviewed

Abstraction versus Selective Attention in Classification Learning

Kenneth J. Kurtz (kkurtz@binghamton.edu)
Department of Psychology, PO Box 6000
Binghamton Unviersity (State University of New York)
Binghamton, NY 13902 USA

Abstract

The Shepard, Hovland, & Jenkins (1961) result that Type II (exclusive-or, XOR) category structures are easier to acquire than Type IV (family resemblance, FR) is shown not to replicate with less easily analyzable stimuli. By increasing the number of stimulus dimensions, the traditional XOR advantage emerges in terms of likelihood of reaching a demanding performance criterion, however more errors are actually made during study by XOR learners as compared to FR. Removing borderline instances from the training set promotes acquisition of the family resemblance structure and leads to a reversal: an actual learning advantage for FR-strong over XOR. Similarity ratings collected after learning reveal another failed prediction of the selective attention account. This set of findings upsets the foundations for the current pecking order among models of category learning.

Introduction

The seminal work of Shepard, Hovland and Jenkins (1961) stands as a leading benchmark for model testing and theory development in the field of categorization. Specifically, the explanatory constructs of selective attention and localist representation of exemplars (or exceptions) have become widely accepted on the basis of superior data fits. Only three published accounts (though see Kurtz, under review) have achieved successful fits to the SHJ data and, in fact, it is the only computational experiment common to all three papers. Otherwise influential and compelling theoretical approaches have suffered for their failure to show the observed ordering of ease of acquisition of the six types of category structures. The present research extends this classic paradigm with a goal of evaluating its place as the sine qua non for category learning models.

The original Shepard, et al. (1961) study tested a small sample of learners on the six general types of two-choice category structures which can be constructed using three binary-valued stimulus dimensions. The abstract form of the eight possible stimulus types (000, 001, 010, 100, 110, 101, 011, 111) can be visualized as vertices of a three dimensional cube. Each of the three dimensions was realized in terms of two possible values that were highly recognizable and analyzable: shape (square or triangle); color (white or black); and size (large or small). Shepard, et al. measured the number of errors made before a demanding learning criterion (4 consecutive perfect passes through the training set) was reached.

The six types of category structures can be summarized as follows. Type I is based on a unidimensional rule (UNI), so considering a single dimension is sufficient for perfect classification. Type II instantiates an exclusive-or (XOR) logical rule in which membership depends upon whether one, but critically not both, of two specific dimensions values are present. For example, white squares and black triangles might comprise one category, while white triangles and black squares comprise the other. In this case, size is an irrelevant dimension. Types III, IV, V are often grouped together as structures possessing regularities, but always with an exception to the rule. Among this group, Type IV conforms to a family resemblance (FR) structure. Examples 000 and 111 can be considered category prototypes with additional members consisting of those examples with a majority (two out of three) of prototype-consistent features. The final category structure, Type VI offers no regularities and can only be mastered by rote memorization.

The observed ordering in terms of ease of acquisition is as follows: Type I (UNI) faster than Type II (XOR); which is faster than Types III, IV (FR) and Vl which are faster than Type VI. Theorists have emphasized an interpretation of this pattern in terms of the number of dimensions requiring attention. The UNI case requires attention to only one dimension and it is the easiest to learn. The XOR case requires attention to two dimensions, and it is the second easiest. The remaining (and hardest) types each require attention to all three dimensions in order to be solved. These results were of early importance in ruling out classes of models based on principles of stimulus generalization. Contemporary theorists have found success in simulating the pattern in terms of: error-driven learning of attention weights (ALCOVE, Kruschke, 1992), discovery of regularities/rules plus exceptions (RULEX, Nosofsky, Palmeri, & McKinley, 1994) or a dynamic clustering approach that integrates aspects of both of these mechanisms (SUSTAIN, Love, Medin, & Gureckis, 2004).

In order to promote model comparison, Nosofsky, Gluck, Palmeri, McKinley, & Glauthier (1994) conducted a replication and extension of the SHJ study with several extensions. The researchers used solid versus dotted lines instead of color as one of the three dimensions. The same strict learning criterion was applied. Testing many more participants the originally observed pattern of acquisition was replicated. By examining the time course of learning, Nosofsky, et al. were able to conclude that the expected ordering of the types was in place from the initial learning blocks all the way to asymptote. (One procedural point is that each learner was tested on two out of the six possible types. While participants were instructed that the two training sets were independent, it remains true that half of the data for each type was collected after the participant had spent up to 400 trials studying the exact same stimuli under a different category structure.) Nosofsky, et al. conducted comprehensive model fits to conclude that ALCOVE produced a better account than competing models and that this success was attributable to selective attention.

The stimuli used in these studies are more or less equivalent to presenting participants with the lists of 0's and 1's used by experimenters. As Shepard put it, "The stimuli in that investigation [(Shepard, Hovland, & Jenkins, 1961)] were psychologically highly analyzable in the sense that they tended to be immediately perceived or described in terms of their values on a small number of perceptually isolated and salient dimensions" (Shepard & Chang, 1963, p. 95). It follows that two reversals of the SHJ results have been observed in the literature. The first case (Nosofsky & Palmeri, 1996) involved the use of non-analyzable or integral stimulus dimensions. Love (2002) tested participants on supervised and unsupervised versions of the SHJ task and found that Type IV (FR) was more easily acquired than Type II (XOR) under conditions of incidental learning (ratings of pleasantness were collected as opposed to classification decisions). In fact, the Type II (XOR) problem was as difficult to acquire as Type VI in this learning mode. In both the intentional unsupervised (a memorization task) and supervised learning conditions, performance on Type II (XOR) and Type IV (FR) was found not to differ. However, a significantly greater proportion of Type II (XOR) learners reached 95% accuracy.

A footnote in the paper articulates the point that this partial failure to replicate SHJ was due to the stimulus set rather than to the non-standard methodology that was required to allow comparisons across learning modes. The stimuli were patterned squares that varied in terms of: yellow vs. white border (representing category) and three of the following four features (with the remaining feature constant across items): slightly larger vs. slightly smaller size; purple vs. blue color; smooth or dotted texture; and presence or absence of a diagonal line. These dimensions were tested for independence and equivalent salience; making this a very useful set of materials. Importantly, the pairs of dimension values are not dramatically different from one another. The dimensions are certainly analyzable, but they are not the overt, overlearned features used in the previous studies. In fact, while not suggesting naturalism, these stimuli do have considerably more in common with the kind of category distinctions learners realistically make.

Experiment 1

This project began with an attempt to more fully explore Love's (2002) finding of no difference in mean accuracy between XOR and FR despite a greater number of XOR learners reaching near-perfect performance. Further reason to more closely investigate XOR vs. FR comes from Nosofsky, et al.'s (1994) time course results (using stimuli like SHJ) that show only a minimal Type II (XOR) advantage over Type IV (FR) for the first two passes through the eight training items; which is followed by a uniform and substantial Type II (XOR) advantage until asymptote. No analyses were reported with regard to a possible interaction between time course and learning condition.

The goals of the current study are: 1) to advance the argument that the traditional SHJ ordering of ease of learning is limited to the case of stimulus materials that are overtly analyzable along overlearned dimensions; and 2) to test whether ratings of item similarity collected after category learning are predicted by selective attention to diagnostic features during the learning phase. If the diagnostic features earn high attentional weighting, then a correlational analysis should show that diagnostic features predict more of the variation in similarity ratings than non-diagnostic features.

The logic to address the first goal is straightforward. Since the learning conditions of primary interest are the UNI, XOR, FR category structures, a between-Ss design was used manipulating learning condition in terms of the SHJ Types I, II, and IV. Regarding the second goal, it has been shown that same-category pairs are treated as more similar by participants who have acquired the category structure over the domain (see Goldstone, 1998; Goldstone, Lippa, & Shiffrin, 2001; Livingston, Andrews, & Harnad, 1998) for evidence of such top-down perceptual learning effects. Kurtz (1996; 1997; in preparation) offers a full treatment and account of category-based similarity (CBS) effects that occur at the conceptual encoding level rather than at the level of fine perceptual discrimination. One observed phenomenon is within-category compression (higher similarity) without a corresponding between-category differentiation (lower similarity). However, in some versions of learned categorical perception phenomena, the pattern differs. Therefore, for present purposes a non-controversial measure of CBS was used: difference scores between the set of same-category pairs and the set of different-category pairs. Comparing such scores for category learners to baseline scores produces a measure of the extent to which the difference between the similarity of same-category pairs and that of different-category pairs increases with category acquisition.

Selective attention predicts CBS to occur only for learning conditions in which attention is differentially allocated across features (i.e., UNI and XOR, but not FR with its three equally predictive features). The item pairs that become more psychologically similar after learning should be those which match on highly attended features. For example, after learning XOR on dimensions 1 and 2, the items (110 and 001) become a same-category pair. However, increased attention to dimensions 1 and 2 would increase the weight on mismatched features, thereby failing to explain any increase in perceived similarity of this same-category pair. Alternatively, pairings such as (111 and 110) would be expected to show CBS since the features matches underlying their similarity are on diagnostic dimensions.

Method

Subjects A total of 109 undergraduates at Binghamton University participated in the experiment in order to receive course credit. An additional 25 Ss participated, but these data were not analyzed due to the use of a data removal procedure pertaining to the similarity phase of the experiment (explained below). Each participant was randomly assigned to condition.

Materials The stimuli were eight examples of the patterned squares downloaded from the site indicated in Love (2002). In order to allow for interpretation of the similarity data, only one featural instantiation of each category structure was used. The texture dimension (smooth or dotted) was the definitional feature for UNI learning. Texture and Diagonal (presence or absence) were the diagnostic features for XOR. Border (yellow or white), Texture, and Diagonal were the three predictive features for FR. All eight items shared the same value for Color and Size. This restriction in feature assignments was made possible by the fact that these stimuli were tested for independence of dimensions and calibrated for equal salience (Love, 2002). In addition, a full range of feature assignments was utilized previously with no differences reported due to feature assignment (Love, 2002).

Procedure The experimental procedure was the same for each of the three between-subjects conditions (UNI, XOR, FR). Participants read a set of instructions explaining the category learning task under a minimal cover story about deciding which of the "examples of geometric images with subtle differences between them" are members of the Alpha or Beta category. The classification learning phase consisted of consecutive passes through the training set in randomized orders. On each trial, the stimulus appeared on a computer screen along with two radio buttons labeled Alpha and Beta. Participants responded via a mouse click. After each response, corrective feedback was provided followed by a self-paced interval for study prior to the beginning of the next trial. The maximum number of training trials was seventy-two (nine passes through the training set), and learning was stopped early according to a criterion of perfect performance on any of the six training blocks of twelve trials (1.5 passes through the training set).

After learning, a test phase was conducted. Participants were instructed to choose the correct category for each example without any further feedback (and to give ratings of the typicality of each example relative to its category). The final task in the experiment was to complete a set of all twenty-eight possible pairwise similarity ratings on a (1-7) scale with endpoints (1) "not at all similar" and (7) "highly similar." The stimuli were presented in random order and each pair was presented in randomized left-right order on the screen. Before beginning the similarity phase, participants were given explicit instructions to "keep in mind that examples from the same category may not be very similar and examples from different categories could be quite similar. Your judgments should reflect the similarity of the specific examples, NOT merely whether they belong in the same category." The reason for this instructional device was the threat of a potential task demand to produce high ratings for same-category pairs and low ratings for different-category pairs (i.e., to demonstrate having learning the categories). To further protect against the task demand problem, participants who showed low variability or high perseveration in responding (such as using only endpoints rather than the full scale) were removed from the analysis.

Results and Discussion

Learning performance was evaluated under the assumption that learners who reached criterion would have continued to produce errorless responding.

Table 1: Relative ease of acquisition of categories.

Condition	Study Accuracy	Test Accuracy	% of subjects reaching criterion
UNI	.95	.99	97%
XOR	.71	.82	44%
FR	.72	.82	26%

As can be seen in Table 1, UNI was vastly superior to the other conditions as expected. In accord with Love (2002), the difference in accuracy between the XOR and FR groups did not approach significance. Unlike Love's results, there was also no significant difference in the proportion of Ss reaching criterion, $\chi 2(1) = 2.38$, $p > .05$. The criterion measure shows a possible trend, but these data certainly defy expectations from SHJ and offer no basis to reject the null hypothesis of no difference between XOR and FR.

An examination of the time course of learning proved informative. On the last block of training, XOR accuracy (M=.83) is slightly higher than FR accuracy (M=.81). This could suggest that XOR acquisition would begin to show an advantage as more learners make the transition from moderately good performance to mastery. The logical nature of the XOR category structure in comparison with the probabilistic nature of the FR category structure is consistent with a prediction of fewer errors on XOR than FR among those learners who have developed a generally effective classification basis. It could be that this difference is what accounts for the SHJ pattern.

For the similarity data, correlation coefficients were computed to measure which feature matches accounted for variability in the mean rated similarity of the item pairs. Based on selective attention, if a feature was diagnostic during learning, then a match on that feature should matter more in rating similarity. Accordingly, the defining feature for the UNI condition showed r =.74, $p < .01$, while the other features failed to show significant correlations with similarity. In the FR condition, all three features only showed moderate, p*'s* $> .05$, correlations with similarity. These results are consistent with the selective attention account (in addition to others such as Kurtz, in preparation) since UNI has one diagnostic feature which is the best predictor and FR has three equally diagnostic features which are all moderate predictors.

In the XOR condition, all three features showed moderate (and narrowly significant, p <.05) correlations of r =.45, .40, .42 with the mean similarity ratings. The predictive power of a match on the diagnostic features was no more predictive than a match on the irrelevant feature. However, it must be recalled that in the XOR category structure on the first two dimensions, pairs such as 000 and 111 belong to the same category. A better predictor might be the match on both of the relevant features considered together. This paired match is a significant predictor r =.62, p < .01, but the other paired matches failed to reach significance. Therefore, it is evident that category knowledge influences similarity performance. However, the mechanism of applying greater weights to a similarity computation for matches on diagnostic features is inadequate. These data suggest instead that the mechanism used to learn the categories is a recoding of the input. Finally, as would be expected given the correlation analysis, UNI and XOR, but not FR, show significant category-based similarity effects using the difference score analysis described above (details omitted due to space restrictions).

Experiment 2

The essential design of Experiment 1 was preserved, but the training set was increased from eight to twenty-four items varying on five rather than three dimensions. Are the previously observed differences among UNI, XOR and FR preserved when the task is less of a toy problem (i.e., more examples comprised of more features)? This design offers an opportunity to replicate the learning and similarity findings with the patterned square stimuli which less dramatically caricature real-world categorization (than do the traditional materials).

In addition to the three learning groups in Experiment 1, a fourth condition (FR-strong) was introduced using a modified version of family resemblance structure. In the FR-strong condition, learners were only exposed to strong category members, i.e. exemplars possessing either four or five out of the five possibly prototype-consistent feature values. The SHJ findings have contributed to a largely doubtful view among theorists regarding any special status for family resemblance structure in learning. In the major modeling efforts, Type IV (FR) is talked about no differently from the non-descript Type III and Type V as category structures that require attention to all three dimensions (Kruschke, 1992) or as rule plus exception structures (Nosofsky, et al., 1994). This is despite the compelling and widely accepted evidence (e.g., Rosch & Mervis, 1975) for family resemblance as an organizing principle for natural categories with numerous demonstrable correlates in cognitive performance. The surprising difficulty inherent in getting participants to spontaneously organize a novel domain according to family resemblance in category construction tasks has contributed to this state of affairs (Medin, Wattenmaker, & Hampson, 1987). Participants tend to produce unidimensional sorts (it goes without saying that they do not produce XOR sorts).

The dissociation between the apparent family resemblance basis for real-world concepts and the nearly allergic response shown by learners to family resemblance structure requires explanation. Medin, et al., suggest that category cohesiveness (knowledge relating category features) may be the explanation. However, numerous results suggest that the noticing of correlations (as required to solve XOR) depends upon top-down knowledge; yet XOR is easily learned in the classic SHJ finding. A simpler hypothesis is that family resemblance is not very compelling to learners asked to consider very low-dimensional stimuli. With the use of (111 vs. 000) or (1111 vs. 0000) as prototypes – which is often the case in artificial category learning studies – the difference between members of the two 'families' is certainly meager. With three dimensions, as in the SHJ study, six out of the eight possible examples are borderline cases. With four dimensions, the situation improves slightly, although items that are literally halfway between the two prototypes are sometimes assigned membership to one category or the other in the training set (e.g., Medin & Schaffer, 1978).

With the use of five-dimensional stimuli, it is possible to test whether or not family resemblance has received the short end of the stick in learning studies. The FR-strong condition uses five-dimensional stimuli and remove borderline items (with only three prototype-consistent features) from the training set. The design addresses the question of whether either the small number of features or the proliferation of borderline examples in the training set has resulted in systematic underestimation of sensitivity to family resemblance in learning. An issue of direct interest is whether either the FR or FR-strong groups might surpass XOR in ease of acquisition. If so, theoretical accounts emphasizing selective attention would need to explain how a category structure that requires attention to five diagnostic features is easier to learn than a category structure determined by two features (XOR).

The design also allows for the evaluation of transfer performance on novel, untrained category examples. This has never been possible in the SHJ task since there are only eight training items and each is critical to realizing the six types. Throughout the literature on learning, the ability to extend a concept to new cases is considered as important as acquisition of the training set.

Method

Subjects A total of 190 undergraduates at Binghamton University participated in the experiment in order to receive course credit. A subset of 19 participants were removed from the analysis based on the data removal procedure pertaining to the similarity phase of the experiment. Participants were assigned randomly to condition.

Materials Twenty-four of the thirty-two possible five-dimensional patterned squares (Love, 2002) were used as training items such that each value on each feature occurred exactly half the time. The remaining eight items were used as novel transfer items in the test phase. The UNI and XOR learning conditions were unchanged in nature from Experiment 1 by the addition of two varying, non-predictive features (Color and Size). The FR training set consisted of the two prototypes (00000 and 11111), all ten of the strong category members (with four prototype-consistent features) and twelve borderline category members (three prototype-consistent features against two inconsistent features). By way of comparison, the FR condition in the original SHJ formulation used in Experiment 1 included two prototypes and six borderline category members (two consistent features against one inconsistent). For the FR-strong group, the training set consisted of only the two prototypes and the ten strong category members. In order to keep the learning phase balanced across conditions, the FR-strong group (with half as many training items as the other groups) received the same number of trials, but twice the exposures to each item.

The extension from three to five dimensions for the UNI and XOR conditions consisted simply of expanding each of the eight original items into four items based on each variation of the newly added non-diagnostic dimensions (00, 01, 10, 11). However, the same approach with the FR stimulus set leads to a violation of the family resemblance category structure. In exactly six of the thirty-two cases, the correct FR category of an item becomes switched by adding two additional feature values. For example, the item 001 is a member of the 0-based category, however the item 00111 is a member of the 1-based category according to family resemblance. The principle of family resemblance was given priority in the present design and category assignments were made accordingly. Of the six cases in which the category assignment would have gone the other way by rote extension of the three-dimensional set, all were borderline category members and two of these appeared in the training set. No unusual effects were observed for these items (none would be expected since participants in Experiment 2 are in no way exposed to the possibility of items and categories designed according to only the first three dimensions).

In the five-dimensional FR training set, the constraints were such that it was not possible to perfectly balance the predictive power of each of the five features. It was decided to assign the patterns as follows: for one of the features (size) the prototype-consistent value occurred in seven out of the twelve category members (58.33% predictive); for each of the remaining four features the prototype-consistent value occurred in nine out of the twelve category members (75% predictive). Across the five features, the prototype-consistent values were 71.67% predictive. This is a close approximation to the 67% predictive power of prototype-consistent features in the original three-dimensional version. In the five-dimensional FR-strong categories, each prototype-consistent feature occurred in five out of the six category members (83.33% predictive) as a result of the removal of borderline examples.

Procedure The experiment was conducted following the same procedure. One difference to note is that participants in this experiment receive fewer exposures to a greater number of items during the seventy-two trial study phase. Apart from the described conditions, an additional set of participants (N=82) was run in a version of the task without category learning in order to collect baseline similarity ratings. A study phase was included to control for item exposure (a generic instruction was given to study each example and click to continue). This was followed by the collection of pairwise similarity ratings.

Results and Discussion

The experiment yielded the results shown in Table 2.

Table 2: Relative ease of acquisition of categories.

Condition	Study Accuracy	% Ss to Criterion
UNI	.93	93%
XOR	.73	50%
FR	.68	11%
FR-strong	.81	31%

All pairwise $\chi2$ tests were significant (p's < .05) except for FR-strong versus XOR, $\chi2$ = 3.15, $p > .05$, indicating the following pattern for reaching criterion:

UNI > XOR = FR-strong > FR

Notably, a greater number of participants reached criterion in the FR-strong condition than in standard FR suggesting that removing the borderline cases promoted family resemblance acquisition. One caveat is that FR-strong learners saw twice the number of repetitions of each item in order to equate overall number of exposures. However, the time course data show that the FR-strong group performed better during the second block (M = .81) than the standard FR group in the final block (M = .70). Learning in the FR group was considerably more flat than what was observed in Experiment 1. To the contrary, the FR-strong group made rapid gains at a rate paralleling that of the UNI learners.

A four-level ANOVA on learning accuracy showed a significant main effect, F = 37.29, MSe = 69.89, p < .001. Pairwise t-tests using the Bonferroni correction showed significant differences between all pairs except for XOR (M = .73) versus FR (M = .68) which did not reliably differ, p > .2. While XOR learners were more likely to reach criterion than FR learners, the overall accuracy between these groups did not show a significant difference. The time course data shows evidence of XOR making steady gains over time relative to FR, but XOR (M = .79) is no closer to FR-strong (M = .88) in the last block than at earlier points in the learning phase.

Similarity performance matches the observed patterns in Experiment 1: only UNI and XOR showed category-based similarity and only the UNI condition showed any one feature accounting for a disproportionate amount of the variance in similarity ratings (details not included due to space restrictions).

Increasing the size of the training set and of the featural composition of the items had no great impact on the UNI, XOR, FR category structures. It is increasingly clear across these studies that XOR and FR generate very much the same learning trajectory – except that XOR learners more easily make the final leap to flawless performance. The superior learning by the FR-strong group suggests that under the right circumstances family resemblance is a privileged category structure (nearly rivaling UNI in study phase accuracy). Notably, FR-strong learners were still unlikely to reach criterial levels as quickly as UNI or XOR learners. In fact, a clear dissociation is seen: FR-strong learners were significantly more accurate than XOR learners during study, but half of the XOR participants reached a demanding learning criterion compared to less than one-third of the FR-strong learners. There are two possible interpretations. The first is a bimodal distribution of the XOR learners. The test phase data (not shown due to space restrictions) is informative in this regard. Exactly 80% of the XOR group performed at an accuracy level either above .90 or below .65 (chance is .50) and is evenly distributed between the two. Acquiring XOR with five-dimensions appears to be essentially an all-or-none proposition (as befits a logical rule). The five-dimensional family resemblance structure shows performance that is fuzzy like the category boundary.

General Discussion

There are two main conclusions to draw from these results. The first is that the famously easy acquisition of the XOR category structure and the notoriously poor acquisition of the FR category structure are limited to the case of stimuli based on three overtly analyzable dimensions. The second is that the selective attention account of category learning is dealt several critical blows: 1) a category structure requiring attention to five dimensions (FR-strong) shows reliably fewer errors during study than one that requires attention to two dimensions (XOR); and 2) XOR learners show compression, increased perceived similarity of same-category examples, relative to baseline, but the match between values for the diagnostic features does not predict the variability in rated similarity.

As the original researchers put it: "...the most serious shortcoming of the [stimulus] generalization theory is that it does not provide for a process of abstraction (or selective attention)" (Shepard, et al., 1961, p. 29). The present results encourage the exploration of abstraction, rather than selective attention, as the core explanatory principle.

Acknowledgments

Thanks to Allan Walker-Hodkin for helping to process the data, to Brad Love for making these stimulus materials available, and to the members of the Learning and Representation in Cognition (LaRC) Laboratory at Binghamton University. This project was partially supported by NIH award 1R03MH68412-1.

References

Goldstone, R. L. (1998). Perceptual Learning. *Annual Review of Psychology*, *49*, 585-612.

Goldstone, R. L, Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition, 78,* 27-43.

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

Kurtz, K.J (1997). The Influence of Category Learning on Similarity. *Unpublished doctoral dissertation.*

Kurtz, K.J. (1996). Category-based similarity. (Ed. G. W. Cottrell). *Proceedings of the 18th Annual Conference of the Cognitive Science Society,* 790.

Kurtz, K.J. (in preparation). Category learning as semantic recoding of instances.

Kurtz, K.J. (under review). The divergent autoencoder (DIVA) account of human category learning.

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 732-753.

Love, B.C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review, 9,* 829-835.

Love, B.C., Medin, D.L, & Gureckis, T.M (2004). SUSTAIN: A Network Model of Category Learning. *Psychological Review*, 111, 309-332.

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

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19,* 242-279.

Nosofsky, R.M., Gluck, M., Palmeri, T.J., McKinley, S.C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, *22*, 352-369.

Nosofsky, R.M., & Palmeri, T.J., (1996). Learning to classify integral-dimension stimuli. *Psychonomic Bulletin & Review, 3,* 222-226.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. K. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,55-79.*

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

Shepard, R. N. & Chang, J. J. (1963). Stimulus generalization in learning of classifications. *Journal of Experimental Psychology*, 65, 94-102.

Shepard, R.N., Hovland, C.L., & Jenkins, H.M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75* (13, Whole No. 517).