Modeling Classification and Inference Learning

Bradley C. Love and **Arthur B. Markman**

[*love,markman*]*@psy.utexas.edu*
Department of Psychology - MEZ 330
The University of Texas at Austin
Austin, TX 78712

Takashi Yamauchi

takashi+@pitt.edu
3939 O'hara St. Rm 815
LRDC, University of Pittsburgh
Pittsburgh, PA 15260

Abstract

Human categorization research is dominated by work in classification learning. The field may be in danger of equating the classification learning paradigm with the more general phenomenon of category learning. This paper compares classification and inference learning and finds that different patterns of behavior emerge depending on which learning mode is engaged. Inference learning tends to focus subjects on the internal structure of each category, while classification learning highlights information that discriminates between the categories. The data suggest that different learning modes lead to the formation of different internal representations. SUSTAIN successfully models inference and classification learning by developing different internal representations for different learning modes. Other models do not fare as well.

Introduction

Categorization is central to our mental lives. We use categories to order our experiences and to make predictions which in turn govern our behavior. This paper explores the relationship between classification and inference learning in order to increase the generality and applicability of models of human category learning. Unfortunately, the field seems to be moving in the opposite direction towards special purpose models of learning that are first and foremost models of classification learning (c.f., Schank, Collins, & Hunter, 1986).

In a typical classification learning experiment, the subject learns to assign simple geometric stimuli consisting of a few binary valued dimensions to one of two mutually exclusive artificial categories. On each trial, the subject infers the category membership of a stimulus item and then receives feedback indicating whether the category assignment was correct. Using the first four dimensions to represent size, texture, shape, and color respectively and the fifth dimension to represent the category label, a classification trial involving an object that is small, smooth, dark, and triangle-shaped

may be represented as (0 0 0 1 ?), while a trial involving an object that is large, rough, light, and square-shaped may be represented as (1 1 1 0 ?). Assuming the first item is in category "A" and the second item is in category "B", following feedback the item representations would be (0 0 0 1 0) and (1 1 1 0 1). In classification learning, subjects tend to focus on information that discriminates between the categories, receive feedback, search for rules and store exemplars that are exceptions to the rules, perform hypothesis testing, exhibit sharp drops in their error rates (all or none learning), and are consciously aware of the rules they are entertaining (c.f., Nosofsky, Palmeri, & McKinley, 1994). Clearly this task is not representative of the majority of real world learning situations. No single laboratory task can hope to address all facets of category learning.

Here, instead of focusing on classification learning, we examine the relationship between two closely related learning modes: classification and inference learning. Classification and inference learning are conceptually similar, but differ in that inference learning tends to focus subjects on the internal structure of each category as opposed to information that discriminates between the categories. In inference learning, subjects predict (i.e., infer) the value of an unknown stimulus dimension instead of predicting the category label as subjects do in classification learning (in inference learning, the category label is given on every trial). The unknown dimension varies from trial to trial such that on one trial a subject may infer the size of an item and on the next trial the subject may infer the shape of an item. Thus, a series of inference learning trials could be (0 1 1 ? 1), (? 0 0 1 0), (1 ? 1 0 0), (0 1 0 ? 0). Like classification learning, the subject is given feedback after making a response (i.e., both classification and inference learning are supervised; the "?" is replaced with a 0 or 1 following corrective feedback).

Yamauchi and Markman's data (1998), henceforth referred to as YM, suggest that inference and classification learning have different informational demands that result in different internal representations being acquired. YM argue that subjects performing classification are more likely to focus on stimulus dimensions that discriminate between categories, to perform hy-

Table 1: The stimulus structure for Yamauchi and Markman's (1998) studies. The stimulus dimension were form, size, color, and position.

Category A	Category B
1110	0001
1101	0010
1011	0100
0111	1000

Table 2: The stimulus structure for Yamauchi et al.'s (2000) studies. The stimulus dimension were form, size, color, and position.

Category A	Category B
1111	1101
1100	0110
0011	1000

pothesis testing, and to store exemplars than they are when engaging in inference learning. YM argue that subjects in an inference learning task focus on the prototype of each category, which should make inference learning easier than classification learning for problems that have well defined prototypes. YM found that subjects master the family resemblance problem illustrated in Table 1 (a linear problem in which each category has an underlying prototype that separates the two categories) faster as an inference learning problem than as a classification learning problem. Subjects were also more sensitive to the underlying prototypes in inference learning. Interestingly, subjects engaging in inference learning followed by classification learning made fewer errors than the reverse order.

Recent results with non-linear categories support YM's conclusion that inference learning focuses subjects on the prototypes of each category, while classification learning focuses subjects on discriminating stimulus dimensions. Yamauchi, Love, and Markman (2000) found that a classification learning advantage arises when non-linear categories are used (the logical structure is shown in Table 2). In the case of non-linear categories, the prototype of each category is not sufficient to separate the categories. Therefore, focusing on the category prototypes should be detrimental and inference learning performance should suffer.

Even though the exemplars are the same in both inference and category learning and the information content of the trials is the same (treating the category label as another stimulus dimension), different representations and radically different patterns of performance emerge. There is no strong a priori reason to favor inference learning or classification learning over the other and therefore it is important to be able to account for data from both learning modes within a single category learning model.

YM report that Generalized Context Model (Nosofsky 1986), a type of exemplar model (e.g., Hintzman, 1986; Nosofsky, 1986; Kruschke, 1992), and the rational model (Anderson 1991) cannot account for their results. In the remainder of this paper, we evaluate whether SUSTAIN (Supervised and Unsupervised STratified Adaptive Incremental Network) can model the results successively. SUSTAIN has successfully modeled classic studies of classification learning, learning at different levels of abstraction as expertise varies, sorting tasks, and unsupervised learning us-

ing the same set of parameters (Love & Medin 1998a; 1998b). If SUSTAIN can account for YM's findings, it would represent an important step towards a unifying model of human category learning that could be applied to a variety of learning modes (e.g., classification, inference, and unsupervised category learning). To foreshadow the results, SUSTAIN can capture inference learning using the same parameters used to model classification learning and its solution is consistent with YM's interpretation of their results. SUSTAIN also predicts the reversal that Yamauchi et al. (2000) observed with non-linear categories.

Overview of SUSTAIN

SUSTAIN is a clustering model that adaptively modifies its architecture during learning. When items are clustered together inappropriately (i.e., similar items from incompatible categories are placed in the same cluster), SUSTAIN adds a new cluster in memory to encode the misclassified item. For example, if SUSTAIN is applied to stimulus items and classifies them as members of the category mammal or the category bird it will develop one or more clusters (i.e., prototypes) for the bird category and one or more clusters for the mammal category. When SUSTAIN classifies a bat for the first time, the bat item will strongly activate a bird cluster because bats are similar to birds (both bats and birds are small, have wings, and fly). After incorrectly classifying the bat as a bird, SUSTAIN will create a new cluster to encode the misclassified bat item. The next time SUSTAIN classifies a bat, this new cluster will compete with the other clusters and will be the most strongly activated cluster (i.e., it will be more similar to the current stimulus than any other cluster), leading SUSTAIN to correctly classify the novel bat as a mammal and not as a bird. The new cluster would then become a bat prototype (a subcategory of mammal). Categories in SUSTAIN consist of one or more clusters (i.e., subcategories).

The method for adding units in SUSTAIN is psychologically motivated by the intuition that people ignore differences when they can (a bias towards simple solutions), but will note differences when forced to by environmental feedback (Medin, Wattenmaker, & Michalski, 1987; Ahn & Medin, 1992). At a more general level, SUSTAIN (like the ARTMAP model of Carpenter, Grossberg, and Reynolds, 1991) expands its architecture when observed inputs do not match top down expectancies.
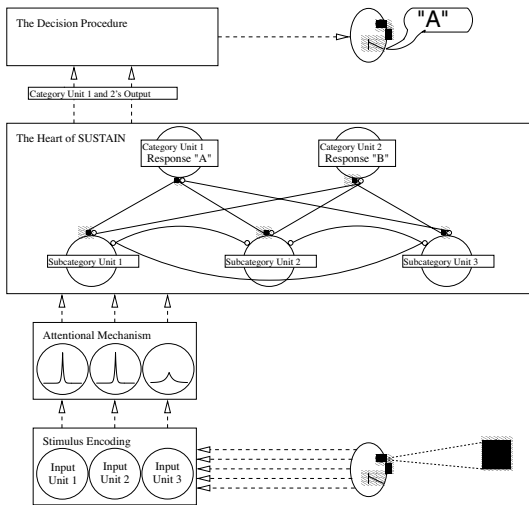
Figure 1: SUSTAIN's architecture is shown. Connections terminating in open circles are inhibitory connections while connections terminating in solid circles are excitatory. Arrows are intended to illustrate information flow.

SUSTAIN's Architecture

SUSTAIN consists of four layers: input, attention, subcategory, and category (see Figure 1). Input layer units take on real values to encode information about the environment (e.g., the encoding of a stimulus item that needs to be classified as a member of category "A" or "B"). In Figure 1, there are three binary valued stimulus dimensions represented by three input units. The three dimensions are dimension 1: size (small or large), dimension 2: shape (triangle or square), and dimension 3: luminance (light or dark). For example, the third input unit represents the luminance of a stimulus: a value of 0 denotes a light object, while a value of 1 denotes a dark object. The attention mechanism weights dimensions, making dimensions that are critical to classification more salient (SUSTAIN learns which dimensions to attend to). The implementation of the attentional mechanism is inspired by the operation of neuronal receptive fields. Each dimension has a receptive field. Dimensions that provide reliable information, and therefore are highly attended, develop peaked and narrow receptive fields (i.e., they develop a sharp tuning). In Figure 1, the first two dimensions (i.e., size and shape) are highly attended.

Units in the subcategory layer encode the prototypes and exceptions of the category units (i.e., the categories' clusters). SUSTAIN does not make a distinction between encoding exceptions and prototypes. A subcategory unit encoding a prototype is simply a unit that re-

sponds strongly to multiple items (i.e., input patterns) while a subcategory unit encoding an exception only responds strongly to one item. In Figure 2, two subcategory units are dedicated to representing category "A". These two units (subcategory units 1 and 2) have an excitatory connection to the category unit representing response "A" (each subcategory unit predicts response "A" when strongly activated). Only subcategory unit 3 is used to represent category "B". Subcategory units compete with one another to respond to patterns at the input layer (notice the inhibitory connections between subcategory units) with the winner being reinforced. The winning subcategory unit is the unit that is most highly activated by the current input pattern (i.e., the subcategory unit that is the most similar to the current stimulus). A subcategory unit is highly activated when an input pattern falls close to it in representational space. For example if a subcategory unit is centered at the point (.9, .8, .1) in three dimensional representational space, the majority of the clusters members would be large, square, and light. Therefore, a large lightly colored square would highly activate the cluster. When a subcategory unit is highly activated and "wins", it moves closer to the current input pattern (according to the Kohonen, 1984, unsupervised learning rule), minimizing the distance between its position and the input pattern. In effect, the correction makes the prototype more similar to the current input pattern (the cluster position is a running average of each member's position).

One novel aspect of SUSTAIN is that this unsupervised learning procedure is combined with a supervised procedure. When a subcategory unit responds strongly to an input pattern (i.e., it is the winner) and has an excitatory connection to the inappropriate category unit (e.g., the subcategory unit predicts "A" and the correct answer is "B"), the network shuts off the subcategory unit and recruits a new subcategory unit that responds maximally to the misclassified input pattern (i.e., the new unit is centered upon the input pattern). The process continues with the new unit competing with the other subcategory units to respond to input patterns. As previously stated, the winner's position is updated, as well as its connections to the category units by the one layer delta learning rule (Rumelhart, Hinton, & Williams, 1986). For example, if subcategory unit 1 is the winner, its connection to category unit 1 would be incremented, while its connection to category unit 2 would be decremented (i.e., it would become more negative). At a minimum, there must be as many subcategory units as category units when category responses are mutually exclusive.

In SUSTAIN, inference learning is assumed to engage the same processes as classification learning, though different internal representations (i.e., clusters) can emerge depending on which learning mode is engaged. A category unit is constructed for each dimension that is inferred in training (analogous to how a category unit is constructed for each category label that is inferred

in classification learning). When an incorrect prediction is made (think back to the bats/birds/mammals example), a new subcategory unit (i.e., cluster) is recruited in the same fashion as in classification learning. The unknown stimulus dimension is simply ignored by SUSTAIN for the purposes of subcategory unit activation. After feedback is provided, the missing stimulus information is filled in for the purposes of learning.

Mathematical Formulation

Receptive fields (which implement the attentional mechanism) have an exponential shape with a receptive field's response decreasing exponentially as distance from its center increases:

$$\alpha(\mu) = \lambda e^{-\lambda\mu} \tag{1}$$

where λ is the tuning of the receptive field, μ is the distance of the stimulus from the center of the field, and $\alpha(\mu)$ denotes the response of the receptive field to a stimulus falling μ units from the center of the field. The choice of exponentially shaped receptive fields is motivated by Shephard's (1987) work on stimulus generalization.

While receptive fields with different λ have different shapes, for any λ, the area "underneath" a receptive field is constant:

$$\int_0^\infty \alpha(\mu)d\mu = \int_0^\infty \lambda e^{-\lambda\mu}d\mu = 1. \tag{2}$$

For a given μ, the λ that maximizes $\alpha(\mu)$ can be computed by differentiating:

$$\frac{\partial\alpha}{\partial\lambda} = e^{-\lambda\mu}\left(1 - \lambda\mu\right). \tag{3}$$

These properties of exponentials prove useful in formulating SUSTAIN.

The activation of a subcategory unit is given by:

$$A_{H_j} = \frac{\sum_{i=1}^n (\lambda_i)^r e^{-\lambda_i\mu_{ij}}}{\sum_{i=1}^n (\lambda_i)^r} \tag{4}$$

where A_{H_j} is the activation of the jth subcategory unit, n is the number of input units, λ_i is the tuning of the receptive field for the ith input dimension, μ_{ij} is the distance between subcategory unit j's position in the ith dimension and the output of the ith input unit (distance is simply the absolute value of the difference of these two terms), and r is an attentional parameter (always nonnegative). When r is high, input units with tighter tunings (units that seem relevant) dominate the activation function. Dimensions that are highly attended to have larger λs and will have greater importance in determining the subcategory units' activation values. Increasing r simply accentuates this effect. If r is set to zero, every dimension receives equal attention. Equation 4 sums the responses of the receptive fields for each input dimension and normalizes the sum (again, highly attended dimensions weigh heavily). The activation of

a subcategory unit is bound between 0 (exclusive) and 1 (inclusive).

Subcategory units compete to respond to input patterns and in turn inhibit one another. When many subcategory units are strongly activated, the output of the winning unit is less. Units inhibit each other according to:

$$O_{H_j} = \frac{(A_{H_j})^\beta}{\sum_{i=1}^m (A_{H_i})^\beta} A_{H_j} \tag{5}$$

where β is the lateral inhibition parameter (always nonnegative) and m is the number of subcategory units. When β is small, competing units strongly inhibit the winner. When β is high the winner is weakly inhibited. Units other than the winner have their output set to zero. Equation 5 is a straightforward method for implementing lateral inhibition. It is a high level description of an iterative process where units send signals to each other across inhibitory connections. Psychologically, Equation 5 signifies that competing alternatives will reduce confidence in a choice (reflected in a lower output value).

Activation is spread from the winning subcategory unit to the category units:

$$A_{C_k} = O_{H_j}w_{jk} \tag{6}$$

where A_{C_k} is the activation of the kth category unit and O_{H_j} is the output of the winning subcategory unit. A winning subcategory unit (especially one that did not have many competitors and is similar to the current input pattern) that has a large positive connection to a category unit will strongly activate the category unit.

The output of a category unit is given by:

$$\text{if } (C_k \text{ is nominal and } |A_{C_k}| > 1), \text{ then } O_{C_k} = \frac{A_{C_k}}{|A_{C_k}|}$$
$$\text{else } O_{C_k} = A_{C_k} \tag{7}$$

where O_{C_k} is the output of the kth category unit. If the feedback given to subjects concerning C_k is nominal (e.g., the item is in category "A" not "B"), then C_k is nominal. Kruschke (1992) refers to this kind of teaching signal as a "humble teacher" and explains when its use is appropriate.

The following equation introduced by Ashby & Maddox (1993) determines the response probabilities (for nominal classifications):

$$Pr(k) = \frac{(O_{C_k} + 1)^d}{\sum_{i=1}^p (O_{C_i} + 1)^d} \tag{8}$$

where $Pr(k)$ is the probability of making the kth response, d is a response parameter (always nonnegative) and p is the number of category units. When d is high, accuracy is stressed and the category unit with the largest output is almost always chosen. In Equation 8, one is added to each category unit's output to avoid performing calculations over negative numbers. The Luce choice rule is a special case ($d = 1$) of this decision rule (Luce, 1959).

After feedback is provided by the "experimenter", if the winner predicts the wrong category, its output is set to zero and a new unit is recruited:

$$\text{for all } j \text{ and } k, \text{ if } (t_k w_{jk} < 0), \text{ then recruit a new unit} \quad (9)$$

where t_k is the target value for category unit k and w_{jk} is the weight from subcategory unit j to category unit k. For example, if the target value of category unit 1 is -1 (i.e., not present) and the winning subcategory unit has a positive connection to category unit 1, the target values times the weight will be negative and a new subcategory unit will be recruited. When a new unit is recruited it is centered on the misclassified input pattern and the subcategory units' activations and outputs are recalculated. The new unit then becomes the winner because it will be the most highly activated subcategory unit (it is centered upon the current input pattern).

The position of the winner is adjusted:

$$\Delta w_{ij} = \eta (O_{I_i} - w_{ij}) \quad (10)$$

where η is the learning rate, O_{I_i} is the output of input unit i. The centers of the winner's receptive fields move towards the input pattern according to the Kohonen learning rule. This learning rule centers the prototype (i.e., the cluster's center) amidst its members.

Using our result from Equation 3, receptive field tunings are updated according to:

$$\Delta \lambda_i = \eta e^{-\lambda_i \mu_{ij}} (1 - \lambda_i \mu_{ij}). \quad (11)$$

Only the winning subcategory unit updates the value of λ_i. Equation 11 adjusts the shape of the receptive field for each input so that each input can maximize its influence on subcategory units. Initially, λ_i is set to be broadly tuned. For example, if input unit i takes on values between 0 and 1, the maximum distance between the ith input unit's output and the position of a subcategory unit's on the ith dimension is 1, so λ_i is set to 1 because that is the optimal setting of λ_i for μ equal to 1 (i.e., Equation 11 equals zero). Under this scheme, λ cannot become negative.

When a subcategory unit is recruited, weights from the unit to the category units are set to zero. The one layer delta learning rule (Rumelhart et al., 1986) is used to adjust these weights:

$$\Delta w_{jk} = \eta (t_k - O_{C_k}) O_{H_j} \quad (12)$$

where t_k is the target value (i.e., the correct value) for category unit k. The target value is analogous to the feedback provided to human subjects. Note that only the winner will have its weights adjusted since it is the only subcategory unit with a nonzero output.

Table 3 lists all of SUSTAIN's parameters and the values used for the studies included in this paper and all cited studies. Unfortunately, it is unusual for a model of human learning to use the same set of parameters across a variety of studies. In this line of research, we focus on drawing conceptual links between diverse data sets and capturing qualitative patterns of performance.

Table 3: SUSTAIN's parameters.

function	name/value
learning rate	$\eta = .1$
cluster competition	$\beta = 1.0$
attentional focus	$r = 3.5$
decision consistency	$d = 8.0$

Modeling Results

As foreshadowed, SUSTAIN successfully fits YM's data on a family resemblance problem (see Table 1). In inference learning, human subjects required 7.9 learning blocks on average to reach the learning criterion compared to 12.5 blocks in classification learning.[1] SUSTAIN displayed the same qualitative pattern, requiring 10.8 learning blocks for inference learning and 16.8 learning blocks for classification learning.

SUSTAIN is an incremental clustering model and can come up with different solutions for different item orderings. SUSTAIN's modal solution (over 80% of simulations) in inference learning involved one cluster per category (i.e., one subcategory unit per category). The one cluster was the underlying prototype of the category. SUSTAIN's modal solution is in accord with YM's assertion that inference learning focuses subjects on the underlying prototype of each category. Attention was evenly spread across all four perceptual dimensions and was highest for the category label dimension (in inference learning the category label is presented with every stimulus). Other solutions involved between three and six clusters with the frequency of the solution decreasing with the number of clusters involved. These solutions arose when item ordering was not advantageous.

In classification learning, SUSTAIN's modal solution (over 60% of simulations) involved three clusters per category. In accord with YM's analysis of human subjects, SUSTAIN created imperfect "rule" clusters (i.e., a cluster that captures some regularity along one or two dimensions that helps discriminate between the two categories) and attention was focused along the "rule" relevant dimensions. Exceptions to these "rule" clusters were captured by "exception" clusters (i.e., a cluster that has one stimulus item as a member). The modal solution in classification learning is less efficient than the one cluster per category solution (the modal solution in inference learning) because with six total clusters there tend to be a large number of highly activated competing clusters (subcategory units inhibit one another in SUSTAIN). Interestingly, approximately 1% of classification learning simulations displayed the one cluster per category solution. When this rare solution occurred in classification learning (due to an advantageous ordering of items), classification learning was as

---

[1] A learning block involves each stimulus from Table 1 being presented once in a random order. The learning criterion was reached when average accuracy exceeded 90% for three consecutive learning blocks.

fast as the average inference learning simulation. This behavior allows SUSTAIN to successfully predict YM's finding that classification learning following inference learning should be easier than the reverse problem ordering. After completing inference learning and discovering the two underlying prototypes, classification learning is trivial.

SUSTAIN also predicts that a classification learning advantage results when the category structure is non-linear (i.e., a category structure in which the underlying prototypes do not separate the categories), as it is in Yamauchi et al. (see Table 2). In inference learning, human subjects required 27.4 learning blocks on average to reach the learning criterion compared to 10.4 blocks in classification learning. SUSTAIN requires 25.6 blocks for inference learning and 15.3 for classification learning. The modal solution (60% of simulations) in classification learning involved three clusters per category (i.e., every item was memorized). In inference learning, the model solution involved nine clusters with the number of cluster required roughly normally distributed (ranging from four to sixteen clusters). SUSTAIN's focus on the prototype leads to prediction failures, which leads to many clusters being recruited. With this non-linear category structure, classification learning performance is roughly the same as in YM's studies, while inference learning performance suffers due to subjects' (and SUSTAIN's) focus on the prototype of each category.

Conclusions

Different learning modes can lead to radically different internal representations on learning problems that involve the same stimulus set and where learning trials have the same information content. In the case of classification learning, subjects focus on a limited number of dimensions and store exceptions to their classification "rule". In contrast, inference learning promotes a focus on the underlying category prototypes. SUSTAIN successfully addresses this data, but other models that do not create different internal representations for different learning modes (such as exemplar models) cannot account for the results. The idea that different representations can emerge from different tasks that involve the same exemplars helped SUSTAIN address another data set in which face experts (i.e., adult humans) learned to identify photographs of faces more easily than they could learn to assign each face to one of two categories (Medin, Gerald, & Murphy, 1983; Love & Medin, 1998a). Exemplar models also have difficulty fitting this data set. By fitting human learning data from a variety of learning modes (classification, inference, and unsupervised category learning), SUSTAIN shows promise as a unifying model of human category learning.

References

Ahn, W. K., and Medin, D. L. 1992. A two-stage model of category construction. *Cognitive Science*

16(1):81–121.

Anderson, J. 1991. The adaptive nature of human categorization. *Psychological Review* 98:409–429.

Ashby, F. G., and Maddox, W. T. 1993. Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology* 37:372–400.

Carpenter, G. A.; Grossberg, S.; and Reynolds, J. H. 1991. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks* 4:565–588.

Hintzman, D. L. 1986. Schema abstraction in a multiple-trace memory model. *Psychological Review* 93(4):411–428.

Kohonen, T. 1984. *Self-Organization and Associative Memory*. Berlin, Heidelberg: Springer. 3rd ed. 1989.

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

Love, B. C., and Medin, D. L. 1998a. Modeling item and category learning. In *Proceedings of the 20th Annual Conference of the Cogntive Science Society*, 639–644. Mahwah, NJ: Lawrence Erlbaum Associates.

Love, B. C., and Medin, D. L. 1998b. SUSTAIN: A model of human category learning. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, 671–676. Cambridge, MA: MIT Press.

Luce, R. D. 1959. *Individual choice behavior: A theoretical analysis*. Westport, Conn.: Greenwood Press.

Medin, D. L.; Dewey, G. I.; and Murphy, T. D. 1983. Relationships between item and category learning: Evidence that abstraction is not automatic. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 9:607–625.

Medin, D. L.; Wattenmaker, W. D.; and Michalski, R. S. 1987. Constraints and preferences in inductive learning: An experimental study of human and machine performance. *Cognitive Science* 11(3):299–339.

Nosofsky, R. M.; Palmeri, T. J.; and McKinley, S. C. 1994. Rule-plus-exception model of classification learning. *Psychological Review* 101(1):53–79.

Nosofsky, R. M. 1986. Attention, similairty, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115:39–57.

Rumelhart, D. E.; Hinton, G. E.; and Williams, R. J. 1986. Learning representations by back-propagating errors. *Nature* 323:533–536.

Schank, R. C.; Collins, G. C.; and Hunter, L. E. 1986. Transcending inductive category formation in learning. *Behavioral and Brain Sciences* 9:639–686.

Shepard, R. N. 1987. Toward a universal law of generalization for psychological science. *Science* 237:1317–1323.

Yamauchi, T., and Markman, A. B. 1998. Category learning by inference and classification. *Journal of Memory and Language* 39:124–149.

Yamauchi, T.; Love, B. C.; and Markman, A. B. 2000. manuscript in preparation.