Using Interpolation Regions to Discriminate Models of Function Learning

Eric Dimperio (edimperi@indiana.edu)
Department of Psychological and Brain Sciences, 1101 East Tenth Street
Bloomington, IN 47405 USA

Abstract

This paper serves to compare existing models of function learning (EXAM & POLE) on a complex interpolation task. Previous comparisons of the models have focused primarily on extrapolation behaviors. Participants' mean responses suggested a simple linear interpolation from nearby points of reference. Both models were able to predict a similar response. Although POLE served as a better predictor of responses made during training, the EXAM model was a better predictor of interpolation responses.

Introduction

How people learn to categorize stimuli is an important topic in psychological research. Although in the past, the terms "category" and "concept" have been used interchangeably (Bourne, 1966; Smith & Medin, 1981) a distinction should be made. *Category* is generally thought of as a description of a group having a common label and a common set of attributes. *Concept* is a broader term that can encompass things other than just categories. Concepts can be distinguished by varying along some continuous scale. For instance, the concept of speed can be expressed either in discrete categories (fast and slow) or with a magnitude along a continuous dimension (115.4 km/hr). Relationships that exist between continuously variable concepts (e.g. income is correlated with intelligence) cannot be studied using traditional category learning studies.

In the area of mathematics, the term *function* is used to describe a relationship where each value from a set of input values (known as the domain of the function) is directly associated with one value from a set of output values (known as the range of the function). The expression "function learning" is used to refer to a situation where people learn a function that relates a set of input stimuli to particular responses where both responses and stimuli come from a continuous set of magnitudes. For instance, a doctor learns how much of a drug to administer based on the intensity of a symptom, or a driver learns how fast a car will go depending on how hard they press the gas pedal. This contrasts with category learning tasks where responses comprise discrete or nominal values.

Function learning certainly does seem related to category learning. They often share similar experimental procedures. In both types of experiments, it is common to show participants a stimulus referred to as a cue. The participant is asked to give a response (sometimes called the prediction). Participants are then shown feedback that indicates the correct response. Also, as mentioned earlier, both paradigms fall under the more general idea of concept learning. It is no surprise that many of the models proposed to describe how people perform function learning tasks are derived from existing models explained as category learning.

Early models of function learning assumed that people developed sophisticated rules from linear combinations of basis functions to make predictions (Brehmer, 1974; Koh & Meyer, 1991). In 1997, DeLosh, Busemeyer, & McDaniel identified limitations in the capacity for these models to generalize learned functions to new cues. McDaniel and Busemeyer (2005) did a more comprehensive comparison of rule-based and association-based models showing that association-based models outperform rule-based models on existing data sets from function learning tasks.

Extrapolation and Interpolation

In concept learning tasks, participants learn to identify the proper mapping of cue values to response values. Many analogous situations are presented in everyday life. However, in the real world, it is very common to be presented with novel situations. We use what we have previously learned about the world to make predictions about outcomes in new situations and act accordingly. For instance, a doctor may be presented with someone that has very mild or very strong symptoms that fall outside of the range observed in the past. That doctor must still determine how what dosage is appropriate for the new situation. Function learning tasks provide an excellent experimental paradigm for studying this phenomenon. Interpolation is a term that refers to the process of extending learned knowledge to make a prediction based on stimulus magnitudes that are between two learned values. The process of extending knowledge to make a prediction completely outside the range of trained cue stimuli is called extrapolation.

Early experimenters working with function learning observed that people were able to interpolate and extrapolate what they had learned. More specifically, people could accurately interpolate, but extrapolation had a much higher error (Carroll, 1963) rate. That is to say that extrapolation responses deviated from what would be predicted using the trained function more than interpolation responses. Attempts to study and model the details of how participants extrapolate in a function learning paradigm have only been quite recent (DeLosh et al., 1997; Kalis, Lewandowsky, & Kruschke, 2001; Griego, 2001; Bott & Heitt, 2004). Unfortunately, little has been done to investigate model interpolation tasks more complex than those reported by Carroll. Kalish et al. (2001) provided evidence that people use a strategy that relies on associating cues along a dimension to multiple different linear functions.

The general rule-based models were favored early on because of the mere fact that people could generalize learned knowledge. The thought was that if only exemplar based associations were being developed, then nothing can associate with cues that have not been presented. Later models either used hybrid association / rule mechanisms or used pure association by relying on the continuous nature of the stimuli in function learning tasks.

Model Comparison

Currently, two major models have been used to explain observations of general training, interpolation, and extrapolation behaviors. The Extrapolation-Association Model (EXAM) uses a simple associative neural network to associate cues to learned responses (Busemeyer, Byun, DeLosh, & McDaniel, 1997). Interpolation and extrapolation rely on using the nearest learned points to generate a linear rule used to calculate a response. The Population of Linear Experts (POLE) model relies on directly associating cues to linear 'experts' which are linear rules used to calculate responses. Interpolation and extrapolation rely on the associated experts of nearby learned cues as well as an overall bias for experts (Kalish et al., 2004). The POLE model can predict multi-modal distributions of responses observed when trained on discontinuous functions. See appendix for details regarding the implementation of the models.

The goal of this comparison is to identify behaviors during an interpolation task that distinguish unique aspects of the models. An inverted-V shaped function was used since it was anticipated that the models would make separate predictions when interpolating in a central region. It was anticipated that subjects trained on cue stimuli with low and high magnitude would properly learn then two functions representing the legs of the inverted-V. It was hypothesized that people would be able to used the learned relationships to produce responses based on the two linear functions. Initial tests of the models indicated that POLE would utilize the linear experts associated to the learned legs of the function to predict a continuation of each side up to a point (see figure 1). EXAM's generalization rule would make a simple horizontal, linear interpolation between the two nearest learned points. However, as explained in the results, both the participant responses and an optimized POLE model provided surprising predictions.

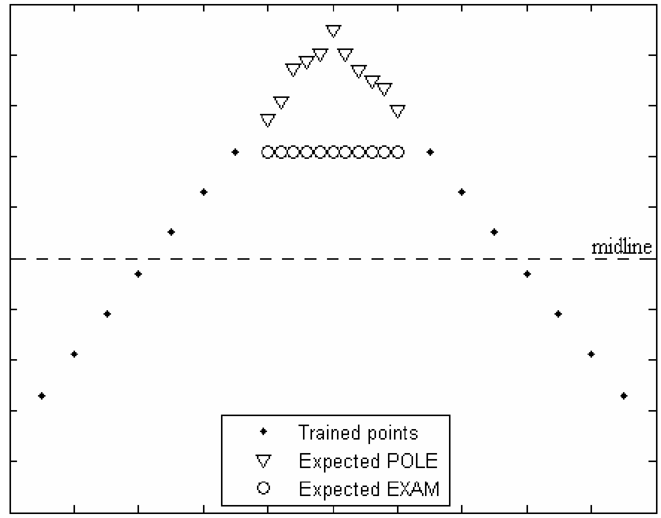

Figure 1: Expected Predictions of EXAM and POLE

Method

Participants Forty-five undergraduates participated in the study for monetary compensation. The data from one participant who did not complete the entire study not considered.

Design The function used to relate the cue stimulus and the response stimulus was an inverted-V shaped function. Our cue and response magnitudes were scaled to be between 0 and 100 at intervals of 10 units. The feedback was determined by the following formula:

$$f(x) = \begin{cases} 15 + 1.6x & if \quad x < 50 \\ 175 - 1.6x & if \quad x \geq 50 \end{cases} \qquad (1)$$

where x is the magnitude of some cue and f(x) is the expected magnitude of the response.

Procedure Participants read instructions from a computer. The instructions explained their task was to learn to predict the number of phone calls a retail business expected to receive based on the number of customers present in the store. The instructions explained the layout of the stimulus display and provided a practice trial.

The stimulus display consisted of three regions corresponding to the cue stimulus (upper left), the response stimulus (bottom center), and the feedback stimulus (bottom right). Figure 2 provides an example of the stimulus display when stimulus values are represented by a vertical bar that is filled by some percentage. Each trial displayed the cue stimulus at a specific value. The response stimulus was visible and its value varied via movement of the computer mouse. The feedback stimulus was not visible during this phase of the trial. The participant manipulated the value of the response stimuli and when the desired response was reached, the left mouse button was depressed to enter the response. Training trials immediately displayed the feedback stimulus as well as a numerical report of the accuracy (100 minus squared deviation) of that response. The feedback was displayed for two seconds. Interpolation trials paused for two seconds without displaying feedback. The next trial began immediately following this two second feedback phase.

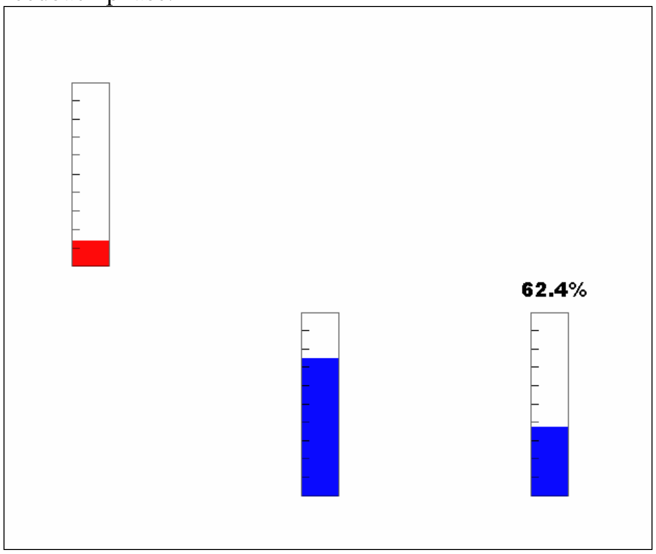

Figure 2: stimulus display featuring a cue stimulus (upper left), response stimulus (bottom center) and feedback stimulus (bottom right)

All sessions involved 207 trials organized as follows: The initial 10 cues were of magnitudes defined by a random permutation of 10, 20, 30, 40, 50, 60, 70, 80, & 90. The next 24 items contain 12 cues from range 5-35 and 12 from range 65-95 in random order. The following 3 blocks contain 12 cues from range 5-35 and 12 from range 65-95 and 4 from range 40-60 in random order. The final 3 blocks contain 12 cues from range 5-35 and 12 from range 65-95 and 6 from range 40-60 in random order.

This order was used to satisfy several goals. First, I wanted a general idea of the initial biases of the participant, so the first block of 10 trials were all given without providing feedback. Next, I wanted to provide some initial training that spanned the entire training range. Third, I wanted to investigate knowledge generalization throughout the learning process. The last two properties are satisfied by treating cue stimuli with a value in the range 40-60 as interpolation trials. Finally, the number of trials was limited to allow participants to complete the task in approximately one hour.

Two different stimuli were presented to each participant. The first stimulus type utilized vertical bars. Each bar had tick marks every 10 units with a maximum value of 100 shown. The second stimulus type involved fractional portions of circles (much like a pie with different sized pieces missing). No tick marks were present on the circles.

After the first session of 207 trials, participants received the same instructions and practice session using the second type of stimulus. They then saw the same 207 cue values in the same order. After both sessions were completed, a written and oral debriefing at the end of the experiment. The order of the stimulus type shown was counterbalanced.

Results

A 2x2 factorial ANOVA identified a main effect due to the session order, but not from the stimulus type. Additionally, a stimulus x order interaction was significant. Therefore, the means of the 207 responses across the first sessions of all 44 participants (collapsed across stimulus type) were used for the model comparison.

Parameters for each model were determined by maximizing the likelihood of the responses for the 167 trials where feedback was obtained[1]. The predictions made by the EXAM model were as expected and accurately reflected the pattern seen in the participants' responses (see figure 3). At first, it was thought that the behavior seen did not require a special interpolation rule and could be produced by the simple associative learning underlying the EXAM model. However, removing the linear interpolation rule led to predictions that regressed toward the mean response in the interpolation zone. Somewhat surprisingly, the optimal parameters for POLE lead to a very similar behavior as EXAM. The large parameter space makes it difficult to accurately fit parameters in the POLE model.

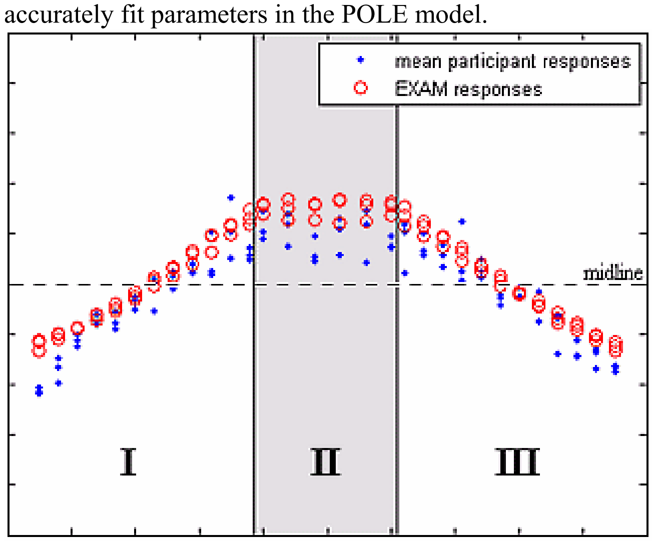

Figure 3: Responses from participants and EXAM to both training trials (regions I & III) and interpolation trials (region II).

My model fitting procedures originally identified parameters that yielded results similar to those expected (as seen in figure 1). Figure 4 shows the expected value of POLE's response for the final blocks of the session. POLE is defined as a probabilistic model, but since the distribution of responses for any cue changed after every training trial it is difficult to visualize those distributions. Moreover,

[1] Parameters were also obtained by maximizing likelihoods for all trial responses. No significant differences in the results were observed.

distribution of response was tightly surrounding the expected values.

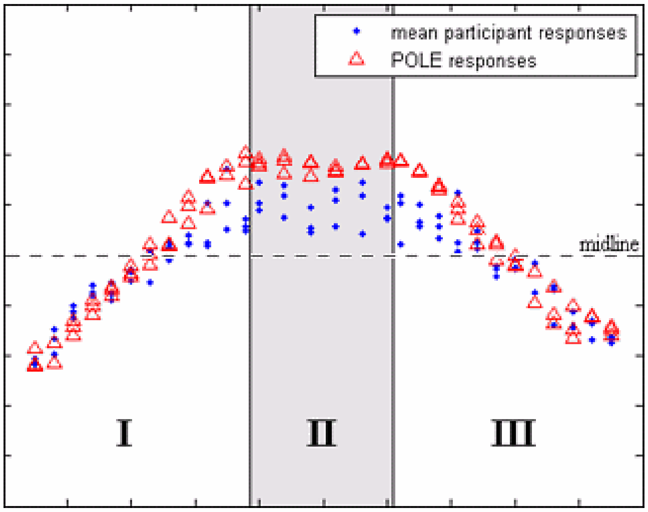

Figure 4: Responses from participants and POLE to both training trials (regions I & III) and interpolation trials (region II).

Since both models are at their cores learning models, I felt it important to compare each model's ability to predict training data when feedback is presented. A comparison of the Bayesian Information Criterion (BIC) was used to allow a penalty for the higher number of parameters used by POLE. POLE relies on 6 parameters while EXAM relies on 2 (see appendix for details). Even with the penalty, POLE was much more accurate (BIC_{EXAM}-BIC_{POLE}= 1368.75-1238.34=130.41) at predicting the data points in the training regions. This is also supported by a similar comparison of the learning abilities of EXAM and POLE by Griego et al. (unpublished manuscript).

The model parameters were only based on feedback trials so a BIC comparison is not appropriate for assessing performance in the interpolation region. A Wilcoxon sign-ranked test of the log-likelihoods of each model predicting the subject responses in the interpolation region demonstrated a significant difference ($p=0.002$). EXAM yielded a higher sum of the log-likelihoods and is therefore a better predictor of interpolation performance. A paired t-test of squared differences between expected value of model output and actual response yielded the same conclusion ($t(58)=4.10$, $p<.001$). EXAM explains a higher proportion of the variation ($r^2 = 0.95$) in the interpolation region than does POLE ($r^2=0.88$).

One interpretation of the results is that rather than using a linear interpretation rule like EXAM, participants are simply relying on a simple associative learning rule and using the associations of similar stimuli to make a prediction. To test this, I gathered predictions using EXAM's underlying associative learning mechanism (ALM) without the linear generalization. Because the ALM makes the same predictions as EXAM for all training points, the same model parameters can be used. A paired t-test demonstrated that the EXAM model performed significantly better than ALM alone ($t(58)=2.14$, $p=0.037$).

Discussion

The analyses show a general trend toward having the POLE model fit learning data more accurately while EXAM better fits interpolation data. This is a conclusion that is also supported by a model comparison of individual performance on extrapolation using similar functions (Griego et al., unpublished manuscript).

Observations of the mean responses of the final three blocks suggested that participants interpolated by using a linear interpolation rule between the two points marking the boundary of the interpolation region (see figures 3 & 4). The results suggesting a general lack of generalization of the simple linear function on the part of participants was a bit surprising. This may be due to the particular set up of this study. Previous investigations utilized stimuli that were all presented next to each other such that relative differences as well as absolute magnitudes could be used to identify relationships. Participants involved in preliminary tests of the experimental setup found stimuli which were not vertically or horizontally aligned much more difficult to work with. This setup was used because the reliance on only absolute magnitude in the absence of absolute differences better reflected function learning situations in everyday life.

Kalish et al. (2004) clearly found cases where interpolation seemed to rely on recalling disjoint functions associated to specific regions of the cue domain. This variation of the task does not require such unique functions to explain the participants' responses. An exploration of the conditions that lead to knowledge partitioning behaviors is needed. It may be that multiple interpolation and extrapolation techniques are used depending on conditions relating to the type of relationship being learned.

Although an analysis of individual subjects is not presented here, POLE's superior ability to match learning data is more prominent in individuals. In this setup, individuals provided extremely noisy responses. POLE is much better at capturing the fractured response distributions seen in individuals during learning.

Given that EXAM only utilizes two parameters and POLE requires six parameters, it seems as if EXAM may offer a better description of the interpolation process for now.

Acknowledgments

This paper was written while the author was supported by NIMH Cognition and Perception MH068346.

Appendix – Model Descriptions

EXAM In 1997, Busemeyer, Byun, DeLosh, & McDaniel described an Associative Learning Model (ALM). It was derived by extending Kruschke's attention learning covering map (ALCOVE) model for category learning. ALM was based on forming associations between cue and response values. This associative neural network was able to accurately describe learning, but was not successful for extrapolation. The EXtrapolation-Association Model (EXAM) was created to make up for this deficiency. It is a hybrid algorithm that uses the ALM learning model, but when it comes to extrapolation, the EXAM model switches to a rule-based linear extrapolation method.

The ALM model is a simple two-layer associative network that updates connections via the delta learning rule. Input nodes represent cue values and output nodes represent responses. When a stimulus of value X is presented, input node X_i is activated according to the Gaussian distribution described by the function

$$a_i(X) = e^{-\gamma \cdot (X - X_i)^2} \tag{1}$$

where γ is a scaling parameter. The output nodes take on the sum of the product of input node values and the weight. The activation of each output node is calculated by summing the products of the input nodes and the weights that connect them to a particular output node. This is shown in with the following formula

$$o_j(X) = \sum_{i=1}^{M} w_{ji} \cdot a_i(X) \tag{2}$$

where w_{ij} designates the strength of association between input node X_i and output node Y_i. The mean output to stimulus X as defined by the weighted average

$$m(X) = \sum_j Y_j \cdot \left[\frac{o_j(X)}{\sum_k o_k(X)} \right] \tag{3}$$

determines the response given.

As previously mentioned, the weights utilized in ALM are updated after every stimulus presentation according to the delta-learning rule. The rule utilizes a feedback signal described by the equation

$$f_j(Z) = e^{-\gamma \cdot (Z - Y_j)^2}$$

And is updated according to

$$\begin{aligned} &w_{ji}(t+1) = \\ &w_{ji}(t) + \alpha \cdot \{ f_j(Z(t)) - o_j(X(t)) \} \cdot a_i(X(t)) \end{aligned} \tag{4}$$

When new stimuli are presented, the model uses a rule based interpolation/extrapolation procedure to calculate a prediction. Two previously seen cues x_1 and x_2 and their associated outputs are used to make a prediction. In the case of interpolation, x_1 is the greatest learned cue smaller than x and x_2 is the smallest learned cue greater than x. In the case of extrapolation, x_1 and x_2 are the two learned cue values closest to x. A response is made using the following equation:

$$y(x) = \frac{m(x_1) - m(x_2)}{x_1 - x_2} \cdot (x_1 - x_2) + m(x_1) \tag{5}$$

POLE In response to evidence of context cues being associated with unique simple functions in a function learning task (Lewandowsky, Kalish, & Nang, 2002), Kalish et al. (2004) developed a new model to explain learning and extrapolation. Their Population Of Linear Experts (POLE) model was based on associating cues to specific rules or *linear experts* that could be used to calculate a response. POLE was able to explain past data sets as well as new data collected in experiments where target functions were not continuous.

The model's input utilizes the same structure as EXAM with nodes representing individual stimulus values. When a stimulus X is presented it activates each input node X_i according to the equation

$$a_i(X) = e^{-c \cdot |X - X_i|} \tag{6}$$

where c is a parameter determining the specificity of activation. This activation is multiplied by associative weights and combined with biased weights to calculate a strength value for each possible linear expert where the strength is defined by

$$s_k = w_{k0} \cdot e^{\sum_j w_{kj} \cdot a_j} \tag{7}$$

where w_{kj} represents the weight from input node j to expert k. It should be mentioned that POLE utilizes two parameters to determine the initial values of the bias weights. This represents the expectations about the functional relationship that a person has before receiving any feedback. Initial bias weights are set to be

$$w_{k0} = \omega \cdot e^{-\varepsilon \cdot |1 - m_k|} \tag{8}$$

such that ω and ε signify the maximum initial bias and the rate of decrease in bias respectively. The value m_k is the slope of expert k. The final prediction the model given in response to X is the weighted average

$$m(X) = \sum_k \hat{y}_{kX} \cdot \frac{s_k}{\sum_k s_k} \tag{9}$$

were $\hat{y}_{kX}$ is the prediction of expert k given input X.

The learning rule used by POLE calculates and error gradient and systematically descends it to minimize errors. the error for each expert k is calculated by $E_k = ½(y - \hat{y}_k)^2$ and is used to determine a weighted error over all experts $E_{mix} = \Sigma S_k E_k$ where S_k is an expert strength after normalizing the sum of strength to be equal to 1. The strengths will be adjusted based on the pre-normalized strengths using the equation

$$\Delta s_K = \eta_s \frac{(E_{mix} - E_K)}{\sum_k s_k} \tag{10}$$

where η_s is a free parameter for the shift rate. This shift is repeated 10 times to obtain a final strength value labeled s_k^{shift}.

Weights are adjusted to minimize the mixed error. The bias weights will be updates using the formula

$$\Delta w_{k0} = \lambda_b \cdot (s_k^{shift} - s_k) \cdot e^{\sum_j w_{kj} a_j} \tag{11}$$

while the weights used to associate inputs with experts are updated using the following formula:

$$\Delta w_{kj} = \lambda_w \cdot (s_k^{shift} - s_k) \cdot s_k \cdot a_j \tag{12}$$

The free parameters λ_b and λ_w are the bias and associative learning rates respectively.

References

Bott, L., & Heitt, E. (2004). Nonmonotonic extrapolation in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 38-50.

Bourne, L.E., Jr. (1966). *Human conceptual behavior.* Boston: Allyn & Bacon.

Brehmer, B. (1974). Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks. *Organizational Behavior and Human Performance, 11,* 1-27.

Busemeyer, J. R., Byun, E., DeLosh, E. L., & McDaniel, M. A. (1997). Learning functional relations based on experience with input-output pairs by humans and artificial neural networks. In K. Lamberts & D. Shanks (Eds.), *Knowledge concepts and categories* (pp. 405–437). Cambridge, MA: MIT Press.

Carroll, J. D. (1963). *Functional learning: the learning of continuous functional mappings relating stimulus and response continua.* Research Bulletin RB-63-26, Educational Testing Service, Princeton, New Jersey.

DeLosh, E.L., Busemeyer, J.R., & McDaniel, M.A. (1997). Extrapolation: The sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 968-986.

Griego, J.A. (2001). *The effects of transfer context on the stability of two types of conceptual structure in a function concept*. Unpublished Dissertation, University of New Mexico.

Griego, J.A., McDaniel, M.A., Dimperio, E.A. & Busemeyer, J.R. (unpublished manuscript). Generalization in function learning: Extrapolation-Association versus Knowledge Partitioning.

Kalish, M.L., Lewandowsky, S., & Kruschke, J.K. (2004) Population of linear experts: Knowledge partitioning and function learning. *Psychological Review, 111,* 1071-1099.

Koh, K., & Meyer, D.E., (1991). Function learning: Induction of continuous stimulus-response relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 811-836.

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

Lewandowsky, S., Kalish, M., & Nang, S.K. (2002). Simplified Learning in complex situations: Knowledge partitioning in function learning. *Journal of Experimental Psychology: General, 131,* 163-193.

McDaniel, M.A., & Busemeyer, J.R. (2005). The conceptual basis of function learning and extrapolation: Comparison of rule-based and associative-based models. *Psychonomic Bulletin & Review, 12*, 24-42.

Smith, E. E., Medin, D. L. (1981). *Categories and Concepts.* Cambridge Mass: Harvard University Press.

Trumpower, D.L. (2005, July*) Individual Differences in Extrapolation of Function Learning. Poster.* Poster presented at Annual Conference of the Cognitive Science Society in Stresa, Italy.