Title
Reinforcement Learning of Dimensional Attention for Categorization

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 26(26)

ISSN
1069-7977

Authors
Philips, Joshua L.
Noelle, David C.

Publication Date
2004

Peer reviewed

Reinforcement Learning of Dimensional Attention for Categorization

Joshua L. Phillips & David C. Noelle
({JOSHUA.L.PHILLIPS,DAVID.NOELLE}@VANDERBILT.EDU)
Vanderbilt University
Nashville, TN 37235 USA

Abstract

The ability to selectively focus attention on stimulus dimensions appears to play an important role in human category learning. This insight is embodied by learned dimensional attention weights in the ALCOVE model (Kruschke, 1992). The success of this psychological model suggests its use as a foundation for efforts to understand the neural basis of category learning. One obstacle to such an effort is ALCOVE's use of the biologically implausible backpropagation of error algorithm to adapt dimensional attention weights. This obstacle may be overcome by replacing this attention mechanism with one grounded in the reinforcement learning processes of the brain's dopamine system. In this paper, such a biologically-based mechanism for dimensional attention is proposed, and the fit of this mechanism to human performance is shown to be comparable to that of ALCOVE.

Introduction

Human category learning performance cannot be easily explained without recourse to a mechanism for selective dimensional attention (Shepard et al., 1961). Dimensional attention is the cognitive process which emphasizes task relevant stimulus dimensions while deemphasizing others. Thus, contemporary formal models of categorization, such as the Generalized Context Model (GCM) (Nosofsky, 1984), have incorporated adaptable dimensional attention parameters. By adjusting these parameters in a category-specific fashion, the GCM has repeatedly provided excellent fits to human data reflecting the frequency (or probability) with which each stimulus is recognized as an instance of a target category. When the GCM is applied to experimental results, dimensional attention parameters are freely varied to optimize the model fit. This means that, while the GCM provides a powerful account of learned categorization performance, it offers no explanation for how dimensional attention is adjusted over the course of learning.

This shortcoming of the GCM has been addressed by a connectionist model called ALCOVE (Kruschke, 1992). ALCOVE incorporates the GCM's formalization of category knowledge, but it also provides a precise algorithm for modifying the attentional "weight" assigned to each stimulus dimension, based on feedback provided to learners on their categorization judgments. In a typical category learning experiment, learners are presented with stimulus objects, one at a time, and are asked to make classification judgments for each. Immediately following each judgment, feedback is provided, typically informing the learner of the correct category label for the preceding stimulus. Once learning is complete, categorization judgments on transfer stimuli, for which no feedback is provided, can provide a window into the structure of the learned category knowledge. The ALCOVE model uses the feedback provided during training to calculate an "error signal", which is simply the difference between the category assignment made by the model and the specified "true" category. A variant of the backpropagation of error learning algorithm (Rumelhart et al., 1986) is used to communicate this error signal to an early stage of stimulus encoding, and this backpropagated error signal is used to adjust ALCOVE's *dimensional attention weights*. Like the GCM, ALCOVE provides good fits to human performance data on learned categories. Unlike the GCM, ALCOVE provides a detailed account of how dimensional attention is shaped by experience.

ALCOVE has been proposed as a model of *psychological* processes, with virtually no aspiration to explain the neural basis of human category learning. Despite this fact, the empirical successes of ALCOVE and its connectionist formalization make the model a tempting candidate for a coarse characterization of associated brain mechanisms. Perhaps ALCOVE can be refined, with each of its proposed psychological mechanisms mapped onto a corresponding detailed account of the underlying neural machinery. One feature of ALCOVE that stands in the way of such a theoretical reduction is its use of the backpropagation of error algorithm in order to learn dimensional attention weights. This powerful learning algorithm has long been criticized for its lack of biological plausibility (Crick, 1989), suggesting that the brain cannot be adapting dimensional attention based on such a gradient-based technique (c.f., O'Reilly (1996)).

As a first step toward a biological model of category learning, we replaced the backpropagation-based dimensional attention mechanism used by ALCOVE with a reinforcement learning mechanism intended to reflect the role of the brain's dopamine (DA) system in learning. This role for dopamine has been formalized by other researchers in terms of an algorithm called *temporal difference (TD) learning* (Sutton, 1988; Montague et al., 1996). Versions of ALCOVE which adapt dimensional attention weights using the biologically supported TD learning method, instead of the more computationally powerful but biologically implausible backpropagation method, were found to fit human performance data about as well as the original ALCOVE. Thus, this work offers a more biologically realistic model of the adaptation of dimensional attention without sacrificing accuracy in accounting for human categorization behavior. Also, the ability to capture human performance with the highly stochastic TD learning method suggests that cognitive mechanisms for adapting dimensional attention may not need to be particularly precise.

Background

ALCOVE Architecture

The ALCOVE (Kruschke, 1992) model of category learning is a feedforward connectionist model that involves three layers of processing units (see Figure 1(a)). The input layer consists of a set of units that each correspond to a single dimension in the stimulus psychological space. Explaining the structure of this perceptual representation is outside of ALCOVE's scope. When fitting ALCOVE to human data, multidimensional scaling (MDS) techniques are typically applied to collected stimulus similarity ratings in order to discern the psychological space used by human learners (Shepard, 1962a; Shepard, 1962b). Each input unit has its own dimensional attention weight, α_i. These weights are non-negative scalar values that modulate the amount of attention paid to the corresponding stimulus dimension. Higher α_i values magnify the differences between stimuli along the given dimension, making them easier to discriminate based on that dimension. As learning progresses, these weights are adjusted via the backpropagation of error algorithm.

The hidden layer in ALCOVE contains a set of units that are arranged in psychological space, one for each training *exemplar*. The activation level of each hidden unit is determined by the following equation:

$$a_j^{hid} \;=\; exp\left[-c\left(\sum_i \alpha_i \left|h_{ji} - a_i^{in}\right|^r\right)^{r/q}\right]$$

where a_j^{hid} is the activation of hidden unit j, c is the specificity of the hidden units, α_i is the attention weight for input unit i, h_{ji} is the preferred stimulus input for hidden unit j along stimulus dimension i, a_i^{in} is the activation value of input unit i, r is the psychological distance metric, and q is the similarity gradient. Hidden unit activity is at a maximum when the inputs match the preferred stimulus of the unit (i.e., a_i^{in} matches h_{ji}). This activation fades exponentially as the stimulus becomes more distant from the preferred exemplar in psychological space, with the c, r, and q parameters controlling exactly how activation decreases with psychological distance.

Finally the output layer contains a set of units receiving activation from the hidden layer via *association weights*. Each output unit corresponds to a category label that might be assigned to a stimulus. These units are standard linear units, with their activation levels, a_k^{out}, computed as the sum of exemplar unit activation levels, a_j^{hid}, weighted by the corresponding association weights, w_{kj}. Output unit activations are mapped onto response probabilities using an exponential Luce choice rule:

$$P(K) \;=\; exp\left(\phi a_K^{out}\right) / \sum_k exp\left(\phi a_k^{out}\right)$$

where $P(K)$ is the probability of selecting category K for the current stimulus, and ϕ is a gain term. These response probabilities may be used to compare network responses with human performance data.

After the presentation of each stimulus and the consequent outputs are produced, the output unit corresponding to the correct response is presented with a target activation level of $+1$, and other units are presented with targets of -1. An error signal consisting of the difference between a_k^{out} and these targets is used to adjust weight values (though output units that "overshoot" their target values are assigned zero error). The association weights are then adjusted using this error signal directly (i.e., using the delta rule), but the selective attention weights are adjusted based on a backpropagated error signal. The resulting weight update equations are:

$$\Delta w_{kj}^{out} \;=\; \lambda_w \left(t_k - a_k^{out}\right) a_j^{hid}$$

$$\Delta\alpha_i \;=\; \lambda_\alpha \sum_j \left[\sum_k \left(t_k - a_k^{out}\right) w_{kj}\right] a_j^{hid} c \left|h_{ji} - a_i^{in}\right|$$

where Δw_{kj}^{out} is the adjustment value for the association weight from hidden unit j to output unit k, $\Delta\alpha_i$ is the adjustment value for the attention weight for input unit i, λ_w and λ_α are the learning rate parameters for the association weights and attention weights, respectively, and t_k is the target value for output unit k.

Temporal Difference Learning

Electrophysiological studies of the dopamine neurons of the basal ganglia have suggested that the firing rates of these cells code for *changes in expected future reward* (Shultz et al., 1997). This is particularly interesting because a measure of change in expected reward is the key variable of a reinforcement learning method called *temporal difference (TD) learning* (Sutton, 1988). This has led a number of researchers to develop TD learning models of the role played by the midbrain dopamine system in learning (Barto, 1994; Montague et al., 1996).

In the TD framework, a continuous reward value (r) is delivered on each time step (t), with positive reward being desirable. A neural system called the *adaptive critic* learns to predict expected future reward (V), given features of the current situation. When future rewards are exponentially discounted by a factor, γ (between 0 and 1), with immediate rewards being valued more than temporally distant ones, the change in expected future reward between two consecutive time steps is given by:

$$\delta(t) \;=\; r(t) + \gamma V(t) - V(t-1)$$

This δ value is called the *temporal difference (TD) error*. The global TD error value can be used to drive learning in the adaptive critic, improving predictions of future reward, and it can also be used to adapt connection weights in neural networks which select actions, pushing those choices toward actions that regularly lead to reward. Models of this kind have been used to explain motor sequence learning in the striatum (Barto, 1994) as well as other forms of learning. We propose that this form of reinforcement learning may also be used to learn dimensional attention weights that lead to correct categorization responses and, thus, reward.

Modeling Approach

Applications of TD learning typically focus on choosing an action from a discrete set. There is currently no clear understanding of how to apply these methods to domains in which a continuous output is needed. Dimensional attention weights are continuous parameters, however, so some modification to standard TD learning is needed to apply this technique to the adaptation of dimensional attention. We have devised two novel connectionist architectures to accomplish this. Our strategy encodes attentional weight vectors (with one α_i weight per dimension) across a single layer of standard connectionist processing units, called the *attention map* layer. Each unit in this layer possesses a fixed preferred attentional weight vector, and activation of a unit encourages the use of that unit's preferred dimensional attention weights. The activation level of each unit is largely determined by its individual bias weight, and the TD learning method is used to adapt these bias weights so as to optimize reward.

At the start of each trial, each of these attention map units is activated, to some degree, by its bias weight. The units then compete to determine the set of dimensional attention parameters to be used by ALCOVE, and the result of this competition is a set of such attention weights. ALCOVE then processes the current stimulus in its usual fashion, producing a categorization judgment. ALCOVE's association weights are then modified in the usual way, using the delta rule, but the dimensional attention weights are handled differently. If ALCOVE confidently chooses the correct category, it is rewarded. Otherwise, it is not. The TD error, δ, is calculated based on this reward signal, and this error is used to modify the bias weights of all active attention map units.

Two different architectures for the attention map layer were investigated. The first of these used *conjunctive coding*, resulting in a *localist* representation of dimensional attention. Under this scheme, the preferred attentional weight vectors of processing units were distributed evenly throughout the weight vector space. Thus, each unit corresponded to a position in attention weight vector space, and the positions of all of the units in the attention map layer formed a uniform grid in this space. On each trial, a simple winner-take-all competition determined the one unit whose preferred weight vector would specify the distribution of attention for that trial. Learning occurred only for the winning unit, using the following weight update equation for its single bias weight:

$$\Delta w_i \;=\; \lambda_r \, (r - a_i)\, f'(net_i)$$

where λ_r is the attention map learning rate, r is the reward for the current trial, a_i is the activation value of the winning attention map unit, and $f'(net_i)$ is the derivative of the unit's activation function (which was the standard logistic sigmoid). Note that this is the standard method for updating weights based on TD error, under the condition of absorbing reward (i.e., we don't predict reward past the end of the trial). In this case, a_i acts as our reward prediction ($V(t-1)$), and we do not predict beyond this trial, so $V(t) = 0$ and $\delta = r - a_i$. A reward value (r) of $+1$ was delivered to the network on trials in which ALCOVE selected the correct category label and produced a confident response (i.e., all output units within 0.5 of their targets). A reward of 0 was delivered, otherwise.[1]

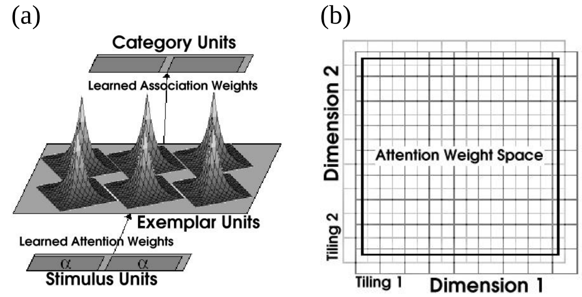

Figure 1: (a) ALCOVE Network Architecture. (b) Tile Coding Of The Attention Map Layer — A single unit is centered in each tile.

Our second attention map architecture used *tile coding*, resulting in a *distributed* representation of dimensional attention. In this case, the attention map layer was partitioned into disjoint *tilings*, where each tiling contained a set of units with preferred dimensional attention weight vectors that uniformly spanned the full weight space. The preferred weight vectors of the units in the various tilings were not identical, however, because each tiling was "offset" from the others, as shown in Figure 1(b). To precisely represent a position in the attention weight space, one unit in each tiling is activated, with the overlap in the *tiles* surrounding the positions of these units determining the dimensional attention weights to be used. This kind of distributed representation was originally used in the Cerebellar Model Articulation Controller (CMAC) (Albus, 1975), and improved generalization in TD learning systems has been

[1] An obvious alternative reward schedule involves stochastically making category judgments based on $P(K)$ and rewarding any correct judgment. While we are currently investigating this approach, it is likely that it will produce behavior that deviates substantially from that of standard ALCOVE. Dimensional attention weights do not change much in ALCOVE until the network starts to make strong responses. This is part of the "three-stage learning" profile that ALCOVE exhibits. Our reward schedule encourages this pattern of learning.

found to result from their use (Sutton, 1996). Previous models have used such representational schemes to encode network inputs, but here they have been used in a novel way to select dimensional attention weights. As in the conjunctive coding architecture, each attention map unit is activated by a bias weight, and a competition ensues between units. In the tile coding scheme, the most active unit across all tilings restricts activity in the other tilings to those that are close to the winning unit (i.e., units whose tiles overlap with that of the winning unit). This competitive process is recursively applied to tilings that do not contain the winning unit until one unit is active in each tiling, and the tiles corresponding to these units all overlap. The attention weight vector at the center of this overlapping region is then used by ALCOVE to process the current stimulus. Once feedback is provided, reward is calculated as in the conjunctive coding case, and TD learning is used to adjust the bias weights of all of the winning units in the attention map layer.

In standard ALCOVE, the initial attention weights are often set to be all equal and to sum to one. This effectively emphasizes all dimensions equally at the start of training. We selected initial bias weights in the attention map layer so as to form a similar initial bias in our models. The unit in the attention map whose preferred attention weight vector matched ALCOVE's standard initial attention weights was given a maximum bias weight (0.05), and the bias weights assigned to other units fell off in a Gaussian fashion as the distance from this peak increased (in attention weight space), bottoming out at -0.05. A small amount of uniformly sampled noise was then injected into each bias weight, and the result was clipped to the $[-0.05, 0.05]$ range. The variance of the Gaussian and the range of the injected noise were free parameters of the model.

Results

In order to assess the ability of our reinforcement-based dimensional attention mechanism to account for human performance, we applied our models to several previously reported category learning studies. The performance of our modified version of ALCOVE was compared to that of the standard version of ALCOVE and to the performance of the GCM. In all cases, the values of dimensional attention weights were bound between zero and one. (This was only a new upper bound for ALCOVE, which standardly forces these weights to be nonnegative.) In all of the learning models, weights were updated after every simulated trial.

Dimensional Attention & Learning Difficulty

Shepard et al. (1961) examined the effect of category structure on the relative speed with which a category is learned. Stimuli were composed of three easily separable binary dimensions, for a total of eight possible stimuli. Six category structures were examined, ordered approximately by increasing number of relevant dimensions. Thus, the Type 1 category structure requires attention to only one binary dimension to solve the task, the Type 2 structure requires that only two of the dimensions be attended, while Types 3, 4, 5, and 6 all require attention to all three, in order of increasing dimensional significance. The speed with which humans learn these categories matches this ordering of tasks, but models that lack a dimensional attention mechanism fail to learn Type 2 categories faster than some of the more difficult categories. Kruschke (1992) showed that ALCOVE, with its adaptive dimensional attention mechanism, learned Type 2 tasks at a relative rate comparable to human learners. We have replicated these simulations (using bounded attention weights and learning after every trial), and the results are shown in Figure 2.

We applied our reinforcement learning version of ALCOVE to these six categorization tasks. Since stimuli had three dimensions, the attention weight space was three-dimensional. The conjunctive coding model used a $15 \times 15 \times 15$ unit topology in its attention map layer (3375 units total), while the tile coding model used five tilings of $9 \times 9 \times 9$ units each (3645 units total). The results of these simulations are shown in Figure 2. Note that our models learn Type 2 categories faster than the higher numbered types, just as ALCOVE does. Model parameter values were manually selected to produce performance that matched the category learning times exhibited by ALCOVE. These results demonstrate that TD learning can adapt dimensional attention weights so as to speed category learning.

Categorization of Continuous Separable Stimuli

In order to demonstrate the ability of our models to quantitatively fit human performance on categorization tasks involving stimuli with continuous and separable dimensions, we applied these models to an experiment conducted by Nosofsky (1986) . The stimuli in this experiment consisted of semicircles that varied in size and contained a radial line oriented at different angles. These stimuli were to be categorized as members of one of two categories, and four different category structures were explored (see Figure 3). The frequency with which each of the sixteen possible stimuli were placed in a target category was measured after training, and the GCM was fit to these response probabilities.

We fit both standard ALCOVE and our reinforcement learning models to this data, as well. Since the stimulus space was two-dimensional, our models used a two-dimensional attention map layer. In the conjunctive coding case, a 15×15 unit topology was used (225 units total), and the tile coding model used 9 tilings of 5×5 units each (225 units total). While both schemes used the same number of units, the tile coding model discretized the space with a much greater resolution. Stimuli were presented to the models using the MDS code found by Nosofsky. Free parameters of the models were fit to Nosofsky's Subject 1 data for each category structure separately. A simple hill-climbing optimization algorithm on sum-squared error was used.

The quality of the resulting fits are summarized in Table 1. While the original ALCOVE model provided the

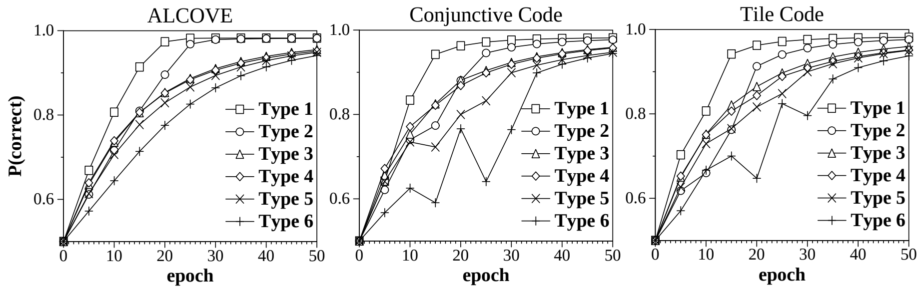

Figure 2: Model Learning Curves For Shepard's 6 Tasks — One epoch involves one trial with each distinct stimulus.

	Category Structure			
Model	*1*	*2*	*3*	*4*
GCM	99.93%	94.73%	84.52%	98.31%
ALCOVE	99.65%	96.45%	86.62%	98.61%
Conj. Code	99.51%	95.84%	86.01%	97.72%
Tile Code	99.54%	95.62%	83.55%	97.72%

Table 1: Model Fits To Nosofsky (1986) – Percent Variance Accounted For

	Category Structure					
Model	*1*	*2*	*3*	*4*	*5*	*6*
GCM	99.10%	98.30%	97.20%	99.80%	98.20%	99.20%
ALCOVE	98.56%	99.29%	93.53%	99.79%	98.34%	98.71%
Conj. Code	98.44%	98.16%	92.51%	99.57%	97.94%	98.53%
Tile Code	98.25%	97.27%	91.15%	99.15%	97.80%	97.30%

Table 2: Model Fits To Nosofsky (1987) – Percent Variance Accounted For

best overall fits, our models matched the data almost as well, and all general trends in the ALCOVE and GCM fits are present in our models. This suggests that our mechanisms for learning dimensional attention can quantitatively capture human performance on learning tasks that require selective attention to separable dimensions.

Categorization of Continuous Integral Stimuli

Integral stimulus dimensions often entail a difficulty in focusing attention on individual dimensions. Despite this fact, Nosofsky (1987) showed that models equipped with a dimensional attention mechanism fit human categorization performance on such stimuli slightly better than models that lacked such a mechanism. This study involved 12 different color chips which varied in saturation and brightness. Six different category structures were used, and these are shown in Figure 3. The frequency with which each of the 12 stimuli were placed in a target category was measured after training, and, once again, the GCM was fit to these response probabilities.

We applied both the original ALCOVE and our reinforcement learning versions to this human data. The same attention map layer sizes as used in the previous simulations were used here, and, as before, MDS representations of the stimuli were presented to the models. A summary of the model fits is shown in Table 2.

The GCM provides the best fits to the data in this study. It seems that the ALCOVE model and our models had trouble learning Category Structure 3. This is a difficult category structure which benefits little from selective attention to specific dimensions. Note, however, that the fits of our reinforcement learning models are close to the standard ALCOVE fits, and our models continue to exhibit the same trends in learning as ALCOVE.

Discussion

Our results show that established computational models of the brain's dopamine system can provide an adequate replacement for the biologically implausible backpropagation of error method for adapting dimensional attention during category learning. The new models were able to learn useful dimensional attention weights from their less-informative global reinforcement signal. This suggests that cognitive mechanisms for allocating dimensional attention may not be as precise as those posited by the original ALCOVE model.

One noteworthy feature of our reinforcement learning models was their tendency to exhibit fluctuations in performance over training, rather than smooth and monotonic learning as displayed by the original ALCOVE model. If each network model is to mirror the performance of an individual learner, these performance fluctuations may reflect stochasticity commonly observed in individual behavior. Also, if performance is averaged across multiple "simulated individuals", smooth learning curves, like those generated by ALCOVE, are produced.

Our models encoded dimensional attention weights in a fairly conjunctive fashion, with individual units in the attention map layer specifying levels of attention for all of the dimensions. This is needed because the appropriateness of attention to one dimension depends on how attention is allocated to the other dimensions. Such a conjunctive encoding requires very large attention map layers, however, and this may limit the scalability of this approach. In order to address this issue, we are currently exploring more compact distributed representations for dimensional attention weight vectors.

Eventually we hope to modify ALCOVE to make use of additional biologically plausible mechanisms of neural computation. This work represents the first step in this process, identifying a biologically realistic method for governing dimensional attention.

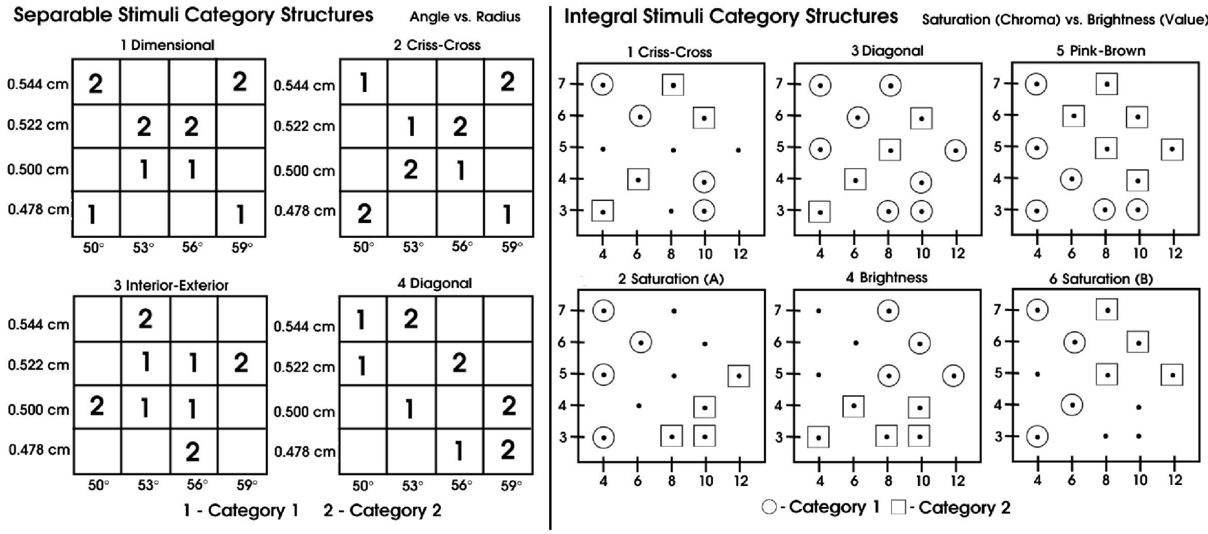

Figure 3: Category Structures Used In Nosofsky (1986) and Nosofsky (1987)

Acknowledgments

The authors extend their thanks for helpful comments on this work to Tom Palmeri, the members of the Vanderbilt University Computational Cognitive Neuroscience Laboratory, and three anonymous reviewers.

References

Albus, J. S. (1975). A new approach to manipulator control: The cerebellar model articulation controller CMAC. *Journal of Dynamic Systems, Meaasurement, and Control*, 97(3):220–227.

Barto, A. G. (1994). Adaptive critics and the basal ganglia. In Houk, J. C., Davis, J. L., and Beiser, D. G., editors, *Models of Information Processing in the Basal Ganglia*, pages 215–232. MIT Press.

Crick, F. M. C. (1989). The recent excitement about neural networks. *Nature*, 337:129–132.

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

Montague, P. R., Dayan, P., and Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of Neuroscience*, 16:1936–1947.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1):104–114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *JEP: General*, 115(1):39–57.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(1):87–108.

O'Reilly, R. C. (1996). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation*, 8(5):895–938.

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

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 27:125–140.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*, 27:219–246.

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

Shultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275:1593–1599.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, 3:9–44.

Sutton, R. S. (1996). Generalization in reinforcement learning: Successful examples using sparse coarse coding. *Advances in Neural Information Processing Systems*, 8:1038–1044.