COMPUTER SCIENCE

REINFORCEMENT LEARNING OF DIMENSIONAL ATTENTION FOR CATEGORIZATION

JOSHUA L. PHILLIPS

Thesis under the direction of Professor David C. Noelle

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.

Approved ______________________________ Date ______________________________

REINFORCEMENT LEARNING OF DIMENSIONAL ATTENTION FOR CATEGORIZATION

By

Joshua L. Phillips

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Computer Science

May, 2004

Nashville, Tennessee

Approved: Date:

ACKNOWLEDGMENTS

I would like to thank the Department of Electrical Engineering and Computer Science for providing academic and financial support for this work. I am especially indebted to my research advisor and mentor, Dr. David Noelle, Assistant Professor of Computer Science and Psychology, for his invaluable advice and willing assistance throughout my graduate school career. I would also like to thank Dr. Thomas Palmeri, Associate Professor of Psychology, for his support and assistance in reviewing this work. I am also grateful for all of the support and comments from the members of the Computational Cognitive Neuroscience Laboratory.

I would like to thank my parents for supporting my career goals and for providing me with the opportunities to reach those goals. Their love and guidance are always with me. Most importantly, I would like to thank my wife, Julie, for her patience and abiding love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

When I wake up in the morning and go into the kitchen, I am surrounded by an array of everyday kitchen utensils and appliances. Assuming I intend to make a cup of coffee, I need to find a coffee mug. The process of identifying which objects in the kitchen are coffee mugs and which are not is critical to my success in making the cup of coffee. This is a process called categorization. Other examples of tasks involving categorization include picking out all of the red trucks in a parking lot, or separating your winter clothing from your summer clothing. In both cases, various objects are placed under scrutiny and, based on their makeup, one makes a decision as to whether they belong in one category or another. Importantly, I had to learn what made something a coffee mug before I could find it in my kitchen. I also needed to learn the difference between cars, vans, and trucks to find them in the parking lot. Over time I had to learn what characteristics of these objects made them belong to these various categories. This process is known as category learning.

Category learning has been studies by psychologists for many years. 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. This sequence of events is done repetitively, often with the same stimuli being observed many times over. Once learning is complete, categorization judgments are made without any feedback, and this can provide a window into the structure of the learned category knowledge. During this part of the experiment, the frequency with which each stimulus is classified as a member of each category can be observed. This offers some insight into the criteria that subjects are using to categorize the stimuli. If the stimuli used in this period are novel, having never been the focus of a trial in which feedback was provided, then we can get an even better view of the general category knowledge that was acquired. This is possible because these stimuli will force the subjects to generalize their knowledge to the new stimuli, and their observable responses will communicate the nature of this general knowledge.

A coffee mug has various features that allow me to categorize it as such. For instance, it has a certain shape, size, color, or texture. However, the color of a coffee mug can vary significantly, and yet I do not stop calling it a coffee mug. This means that when I was learning to categorize coffee mugs, I had to learn that the dimension of color was not important to this decision. However, if the composition dimension of the mug was changed, so that it was now made of glass instead of ceramic, then I would no longer call it a coffee mug. These are examples of a commonly observed property of category learning known as dimensional attention. Dimensional attention is where dimensions used to categorize objects become more or less salient during the process of category learning.

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.

The GCM has been used to explain a variety of phenomena in the category learning domain. It explains why people strongly associate the ”average” stimulus belonging to a category more than stimuli outside of the category with its category label, even if that prototypical ”average” stimulus has never been viewed before. The GCM also captures differences in categorization performance related to the frequency with which individual stimulus items are observed. This is accomplished by relating information about individual stimuli in memory, allowing the GCM to simultaneously account for patterns of categorization and recognition memory behavior (Nosofsky, 1984).

When the GCM is applied to experimental results, dimensional attention parameters are freely varied to optimize the fit of the model to human data. 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.

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., 1986b) 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. It can explain the same phenomena as the GCM, and also many others. It explains differences in learning speed between different category structures, apparent base-rate neglect, and the pattern of "three-stage-learning" displayed by humans. Critically, 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 combination of 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, I replaced the backprop-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 our cognitive mechanisms for adapting dimensional attention may not be as precise as those proposed by ALCOVE.

CHAPTER II

BACKGROUND

Category Learning Paradigm

In standard category learning tasks, the participant begins by simply observing a stimulus and making a guess as to which category the stimulus belongs. Following this guess, the participant is provided with some form of feedback that indicates the actual category for the stimulus presented. The presentation of a stimulus, the participant response, and the presentation of feedback all together make up a trial. The participant then goes through several hundred to several thousand trials attempting to maximize the number of correct responses. This portion of the task is known as the training phase. After training is complete, several trials are done where the feedback portion of the trial is withheld. This is known as the transfer phase of the task. The participant responses are recorded during the transfer phase (also often called the testing phase.) These transfer phase responses are then analyzed to discern the probability of each stimulus being classified under each classification. The transfer phase provides insight into how subjects are categorizing the stimuli by providing statistical evidence for certain categorization strategies that subjects might be using.

The stimuli used in current category learning studies are often composed of features that vary across constituent stimulus dimensions. For instance, a set of stimuli might vary across the dimension of color where individual features along this dimension might be white, grey, or black. This characterization of a stimulus is in contrast to a more featural view, where the individual features of white, grey, and black are seen as being independent or having no inherent relationship to each other psychologically. The dimensional representation has at least an ordinal arrangement of stimulus features, allowing features to be compared in similarity by their relative “locations” along the stimulus dimensions. An example of such a representation can be seen in Figure 1 where there are four different stimuli that vary across two dimensions: size of the circle and orientation of the radial line. These stimuli are then mapped onto a dimensional space where the stimuli are represented as individual points in the space. In such a dimensional space representation scheme, individual stimulus items are encoded as points in the space. These points are selected so that the distance between two points in this *psychological space* is inversely related to the perceived similarity between the two corresponding stimuli. This similarity-distance requirement often cannot be met if the axes of the representation space are forced to correspond to arbitrarily chosen dimensions, using arbitrary units along each one. For example, the size of a circle might be represented by its diameter or by its area, and perceived similarity between sizes may map onto only one of these measures. In order to ensure that distance in the representation space is approximately related to perceived similarity, it is common to derive of discover the dimensions of this space by applying Multidimensional Scaling (MDS) techniques to explicit measures of perceived similarity (Shepard, 1957; Shepard, 1962a; Shepard, 1962b). For example, the degree to which stimuli are confusable can be assessed using an identification study (a category learning study where each stimulus is to be placed in its own separate category), and MDS can be applied to this confusability information in order to generate a dimensional psychological space which appropriately captures similarity between stimuli. All of the models discussed in this paper use a dimensional representation of this kind in order to encode stimuli.

Several hypotheses about the strategies that subjects might be employing to categorize stimuli have been proposed. At first, it was thought that subjects formed simple rules that defined category membership. In other words, the subjects were thought to be inventing a list of sufficient conditions (or stimulus features) that could be used to determine category membership. If a stimulus met all of the conditions for membership to a particular category then the subject would respond with this corresponding category label. Thus, category boundaries were generally taken to be sharp and absolute.

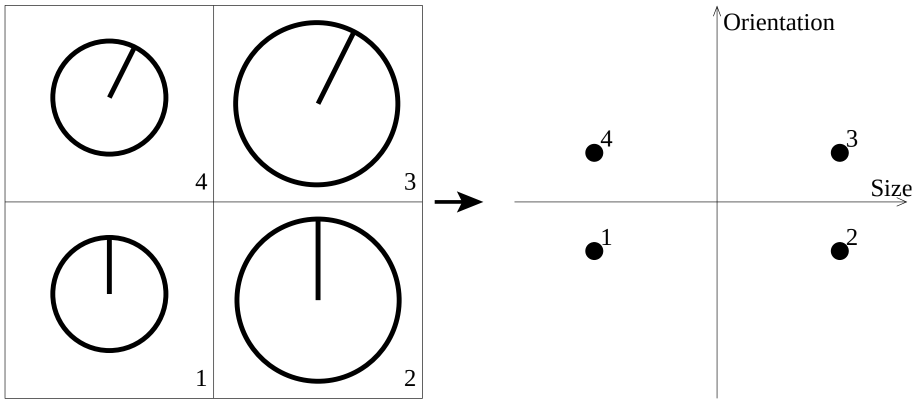

Figure 1: Four stimuli consisting of two dimensions (size and orientation) mapped to their corresponding representation in psychological space

Rule-based representations of categories allow for the discrimination of members from non-members, but they do not provide a way to separate "good" *exemplars* of a category from borderline cases. This is in conflict with the fact that human categorization behavior often displays a variable and graded response to different members of a category. For example, people often respond that a recliner is a chair faster than when they make the same categorization for a bar stool (that has a back portion). People are also less likely to categorize the bar stool as a chair than the recliner. Finally, people report that they consider the recliner more like a chair than the bar stool. These sorts of responses suggested to some researchers that there is some kind of *prototype* stimulus or an "average" of stimuli, that embodies knowledge of a category (Rosch and Mervis, 1975). To categorize a new stimulus, it is compared to the prototype, with similarity driving responding.

Even though the prototype view of categorization accounts for the graded nature of many categories, it does not account for stimulus frequency effects. If the frequency of seeing the various stimuli is not the same for all stimuli in the study, then performance will generally be better for high frequency items, even if these are fairly dissimilar to the category prototype. In general, subjects tend to be sensitive to the particular distribution of observed stimuli and not only to the mean (i.e., prototype) of this distribution. This has led researchers to postulate that many stimuli are retained in memory throughout training and used for categorization (Medin and Schaffer, 1978). Thus, instead of a prototype, a number of individual exemplars is compared to the current stimulus to ascertain category membership. Since certain exemplars are seen more frequently, they tend to be associated with category membership more than stimuli that are seen less frequently. Thus, exemplar-based models of categorization can account for this effect where simple prototype models fail.

The exemplar theory of categorization now dominates the category learning paradigm, and many mathematical or computational models of categorization have been developed based on this theory. Two of the most prevalent and successful models are examined in detail later in this paper, the Generalized Context Model (Nosofsky, 1984) and the Attention Learning Covering Map (Kruschke, 1992). However, hybrid models of categorization have been developed recently that are influencing current theories of category learning as well. Some of these models attempt to combine the exemplar-based and rule-based strategies into a cohesive framework. For example, ATRIUM (Erickson and Kruschke, 1998) consists of two interacting systems: one that accomplishes rule-based learning and another that accomplishes exemplar-based learning. RULEX (Nosofsky et al., 1994) is a model of rule-based learning that incorporates the learning of exceptions, which are particular stimuli that seem to fit the rule but belong in some other category. Even more interesting behaviors result when subjects are actually provided with a categorization rule before training. This is known as instructed category learning. Exemplar information actually seems to interfere with subjects' abilities to properly use the given rule, and these interference effects can be explained using a hybrid model incorporating rule representations and exemplar similarity information (Noelle and Cottrell, 2000).

Of particular interest here is the notion of selective dimensional attention in category learning theory. Dimensional attention is the ability of humans to selectively emphasize or deemphasize dimensions to aid categorization. For some tasks, some dimensions might be of no relevance when categorizing the stimuli. Reducing attention to these dimensions would improve performance. This makes sense intuitively because people seem to have limited attentional resources. If attentional resources are allocated more appropriately, then it will take less time and effort to solve a task. Conversely, attending to more relevant dimensions will improve categorization performance. Humans learn to do just that when categorizing stimuli. Even though exemplar models of category learning can explain human performance well, they fail to do so without incorporating some mechanism for selective dimensional attention (Shepard et al., 1961).

Generalized Context Model

The Generalized Context Model (GCM) (Nosofsky, 1984) is a mathematical, exemplar-based model of categorization. It was developed to account for human performance in category learning based on the Context Theory of Classification (Medin and Schaffer, 1978), a previous exemplar model of categorization. The GCM extends Context Theory in two ways: by incorporating a choice model for stimulus identification (Luce, 1963) and by incorporating a mechanism for selective dimensional attention. The GCM assumes that perceived similarity between stimuli is best represented in terms of the distance between stimulus representations in a multidimensional psychological space, as described earlier. However, in the GCM, as two stimuli become more distant in psychological space, their measure of similarity exponentially decreases (Shepard, 1987). The choice rule is applied to the calculated stimulus similarity ratings to generate response probabilities which are fit to the human response probabilities from identification and category learning studies.

The exponential similarity measure and choice rule incorporated by the GCM allows the model to fit human data from stimulus identification studies well. In these studies, all stimuli are classified as belonging to their own individual categories. All other aspects of the task are the same as in typical category learning. Thus, the objective of identification tasks is to identify the stimuli individually. However, stimuli close together in psychological similarity space tend to be easily confused. This is evident in the human performance

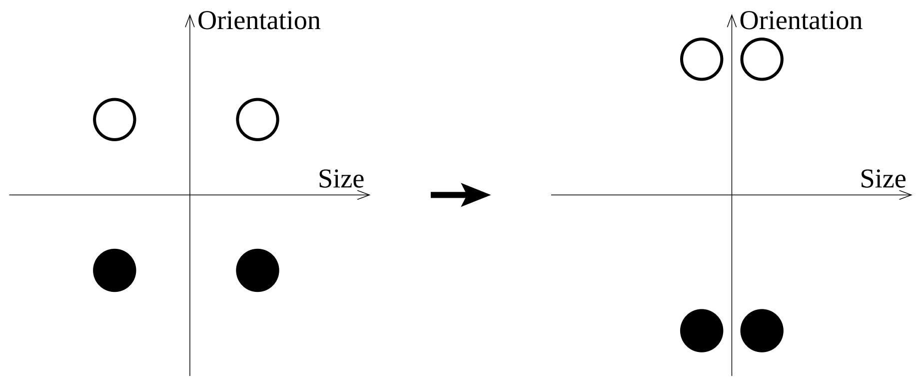

Figure 2: Stimuli of the two categories (filled or blank) have lower between-category similarity and greater within-category similarity as a result of shrinking the horizontal axis and stretching the vertical axis.

data from the transfer phases of these tasks, where the response probabilities tell us something about how often close stimuli are mistaken for each other. Empirically, the response probabilities show that similarity does indeed decrease exponentially with increasing psychological distance. Also, the data from such studies can be used to discern the actual positions of the stimuli in the psychological space using MDS techniques, as described earlier. The MDS values are later provided to the GCM when fitting the model to response probabilities from the transfer phases of category learning studies.

However, the exponential similarity measure by itself is not enough to account for human category learning data. Because people learn to pay attention to more relevant dimensions and ignore irrelevant ones, the psychological space seems to change during these studies. The GCM formally models this by incorporating scalar dimensional attention weight parameters, one for each stimulus dimension. These weights effectively “shrink” the psychological space along irrelevant dimensions and “stretch” the psychological space along more relevant dimensions. This makes stimuli seem highly similar along irrelevant dimensions so that no distinctions can be made using that dimension, and the stimuli become more dissimilar along relevant ones, making stimuli along these dimensions easier to discriminate. Figure 2 shows how these dimensional attention effects can aid discrimination.

Due to the incorporation of this dimensional attention mechanism, the GCM does an excellent job of fitting human response probabilities from category learning studies. The dimensional attention weights do indeed morph the original psychological space in the correct manner to account for the human data. However, these dimensional attention weights are free parameters in the GCM. So the model provides no indication of how people might be learning to morph the psychological space for better categorization performance. To better understand how people might be learning to do this, other models have been proposed that model the training phase of category learning studies.

Connectionist Modeling

A common approach to modeling various psychological and neurobiological phenomena is the connectionist framework. In this framework, a set of simple processing elements called *units* are arranged in groups called *layers*. Units are connected to other units via connections. Often all of the units in one layer connect to all of the units in another layer for computational simplicity. This is called a full *projection* of connections and information flows across these connections in one direction. Each connection also has a scalar *weight* value associated with it. Each unit computes an *activation value* that is some function of the activation values of the units connected to it, modulated by the weight of the connection between the units. This function is referred to as the unit's *activation function*.

A connectionist network consists of multiple layers of units connected by projections of connection weights. One layer of units is typically called an *input layer* because it does not receive a projection from any other layer. Therefore, the activations of the these units are fixed to some set of values, known as an *input pattern*, to begin processing. Activations are computed in each successive layer until reaching the *output layer* which has no projections to other layers. This layer then contains the network output for the input pattern that was provided to the input layer. If there are no cycles in processing from the input to the output layers, then the network is called a *feed-forward* network. All other layers in a feed-forward network are called *hidden layers*. Various *learning algorithms* exist for these connectionist models that adapt the connection weights over time, resulting in improved model performance.

The connectionist framework is interesting because it can roughly approximate the processing done by the brain, if it is carefully designed to do so. However, connectionist models are not necessarily biologically realistic and may include extra computational mechanisms to account for various psychological phenomena. These additions might play an important scientific role if the actual underlying neural mechanisms are not well understood.

ALCOVE Architecture

The ALCOVE (Kruschke, 1992) model of category learning is a feed-forward connectionist model based on the GCM that involves three layers of processing units as shown in Figure 3. The input layer consists of a set of units that each correspond to a single dimension in the stimulus psychological space. MDS representations of the stimuli are provided as input patterns to this layer. Each input unit has its own dimensional attention weight, α. These weights are non-negative scalar values that modulate the amount of attention paid to the corresponding stimulus dimension in the same manner as the GCM. Higher α values result in more attention being paid to the corresponding stimulus dimension by increasing the separation between stimuli along that dimension, making discrimination easier along that dimension. The opposite effect can be achieved with lower α values, which reduce the separation between stimuli along the dimension, making discrimination of stimuli more difficult along the dimension.

The hidden layer in ALCOVE contains a set of units that are arranged in psychological space, one for each training *exemplar*. Each of these exemplar units has a preferred level on each stimulus dimension. The activation value of each hidden unit is determined by the

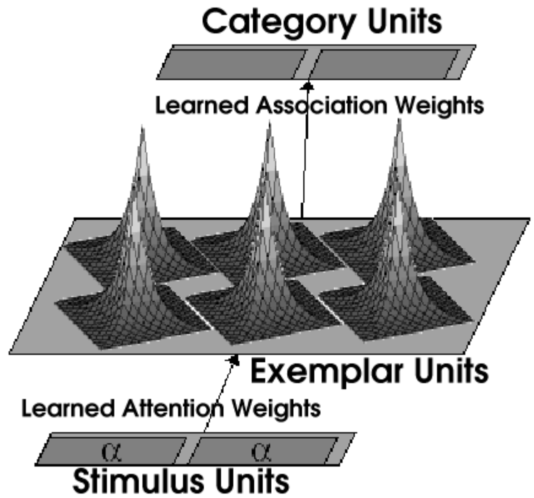

Figure 3: The ALCOVE Architecture

following equation which is also the similarity computation performed by the GCM:

$$a_j^{hid} = exp\left[-c\left(\sum_i \alpha_i \left|h_{ji} - a_i^{in}\right|^r\right)^{q/r}\right] \tag{1}$$

where a_j^{hid} is the activation of hidden unit j, c is the specificity of the hidden unit, α_i is the attention weight for input unit i, h_{ji} is the preferred stimulus level of the hidden unit j along stimulus dimension i, a_i^{in} is the activation value of dimension i, r is the psychological distance metric, and q is the similarity gradient. Hidden unit activity is at at a maximum when the inputs match the preferred stimulus input 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. The rate at which this fading occurs is controlled by the specificity parameter c. A large c value causes activation to drop off more quickly as stimuli become more distant from the preferred exemplar. The r and q parameters define the distance metric used to examine similarity between the preferred stimulus and the current stimulus being presented to the network. With $q = 1$ and $r = 2$, the common Euclidean distance metric is used. However, sometimes it is more appropriate to use a city-block distance metric: $q = r = 1$. It has been shown that the city-block metric creates a better fit to human data on categorization tasks involving stimuli with separable dimensions than the Euclidean metric which produces better fits for integral stimuli (Nosofsky, 1987). The differences between integral and separable stimuli will be discussed later.

Finally the output layer contains a set of units receiving activation from the hidden layer units via *association weights*. Each output unit corresponds to a category to which an input stimulus might be assigned. These units are standard linear units with their activation computed by the following equation:

$$a_k^{out} = \sum_j \left(w_{kj} a_j^{hid} \right) \tag{2}$$

where a_k^{out} is the activation value of output unit k, w_{kj} is the value of the association weight from unit j to unit k, and a_j^{hid} is the activation value of hidden unit j.

The activations of the output units are mapped onto response probabilities using an exponential Luce choice rule:

$$P(K) = exp\left(\phi a_K^{out}\right) / \sum_k exp\left(\phi a_k^{out}\right) \tag{3}$$

where $P(K)$ is the probability of responding that this stimulus belongs in category K, a_k^{out} is the activation value of output unit k, and ϕ is the gain term (this term is a free parameter that is used to fit ALCOVE to human data.) 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).

There have been a number of connectionist algorithms that are capable of adjusting the weights of the network based on gradient-descent in network error. A common biologically plausible learning algorithm of this kind is the *delta rule* which has been used in other connectionist models of category learning (Gluck and Bower, 1988; Gluck et al., 1989). This algorithm was generalized to be used in multilayer networks and is therefore known as the *generalized delta rule* (Rumelhart et al., 1986a; Rumelhart et al., 1986b). Both of these learning mechanisms are computationally powerful ways to learn connection weights. They adjust weights incrementally by a small amount (Δw) based on network error. However, the delta rule can only adjust connections running into the output layer. Therefore, the generalized delta rule must be used to adjust connection weights deeper in the network. The association weights are 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:

$$\Delta w_{kj}^{out} = \lambda_w \left(t_k - a_k^{out} \right) a_j^{hid} \tag{4}$$

$$\Delta \alpha_i = \lambda_\alpha \sum_j \left[\sum_k \left(t_k - a_k^{out} \right) w_{kj} \right] a_j^{hid} c \left| h_{ji} - a_i^{in} \right| \tag{5}$$

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, t_k is the target value for output unit k, and all other symbols have already been defined. The learning rate parameters are part of the standard delta rule and generalized delta rule and they are often kept very small ($[0, 1]$) to encourage gradual weight changes. The only constraint in applying these update equations is that ALCOVE restricts dimension attention weights to be non-negative because negative attention weights do not represent anything of psychological value.

The generalized delta rule (also known as the backpropagation of error learning algorithm or, more simply, backprop) (Rumelhart et al., 1986a; Rumelhart et al., 1986b) changes weights in the network by adjusting values in a slow, incremental fashion to minimize a formal network error function which is a function of the network's output. An error function called *sum-squared error* is used by ALCOVE, just as it is in many other connectionist architectures. The more the network's responses differ from the target response to the provided input stimulus, the greater the error. An error value of zero represents a perfectly correct response. The network error is "backpropagated" from the output layer by passing error information back across connections to previous layers. The dimensional attention weight update equation in ALCOVE incorporates this distant error information to properly learn these weights.

ALCOVE's combines error-driven learning and the GCM's mathematical formalisms into a powerful connectionist architecture, allowing it to fit human performance during category learning quite well. It has been used to explain selective dimensional attention learning, how category structure effects learning speed, apparent base-rate neglect phenomena, and "three-stage learning" of rules and exceptions (Kruschke, 1992). The hidden layer architecture of ALCOVE also allows it to overcome the effects of *catastrophic interference* that are often observed in other connectionist models of category learning (McCloskey and Cohen, 1989).

Like many other connectionist architectures, ALCOVE makes no assumptions about the underlying neural mechanisms that are responsible for category learning in the brain. It is, instead, simply a psychological model of categorization performance and category learning. However, its success in describing category learning phenomena in a connectionist framework make it a promising approach to creating a biologically plausible model of category learning in the brain. It might be possible to make incremental changes to this architecture by replacing biologically implausible mechanisms with more feasible neural mechanisms to gain some insights into how the brain might be performing categorization.

The goal of this work is to begin the process of making ALCOVE biologically plausible through a change in ALCOVE's mechanism for learning dimensional attention. The

ALCOVE model uses the backpropagation of error algorithm to adjust dimensional attention weights, but this algorithm has been criticized for its biological implausibility (Crick, 1989; O'Reilly, 1996). In brief, the backpropagation of error learning algorithm requires error signals to be sent backwards along the same weighted connections that control the flow of activity through the network. To the degree that weighted connections are seen as abstractions of the information processing that takes place at neural synapses in the brain, such a backward communication of error is impossible. In the brain, such information cannot be propagated back across synaptic connections. Therefore, if the basic structure of ALCOVE is correct, some other mechanism must be used by the brain to update dimensional attention. Thus, some other biologically plausible mechanism of learning needs to be found that properly learns the dimensional attention weights in ALCOVE.

Temporal Difference Learning

Recent studies indicate that the firing rates of dopamine neurons in the basal ganglia encode a signal about *changes in expected future reward* (Shultz et al., 1997) that is globally available throughout much of the brain. Schultz et.al. (1997) reported single cell recordings from a population of dopamine neurons in the substantia nigra of a monkey, taken during a simple classical conditioning study. The results are shown in Figure 4. The top graph represents the situation where a monkey is given a reward (a small sip of juice). Notice that the dopamine neurons fire above their baseline firing rate just after this reward is given. Therefore, we can easily see that these neurons are influenced by rewards. If a conditioning stimulus (a tone or sound) is presented to the monkey shortly before giving the reward, and this is repeated for several trials, these neurons then begin to fire for the conditioned stimulus as shown in the graph on the left. However, they do not fire for the actual reward. Now, it would appear that these neurons have learned to predict the upcoming reward. In other words, the monkey has learned to expect a reward after the conditioned stimulus. Interestingly, if the conditioned stimulus is presented, but then the reward is withheld as in the graph on the right, the cells actually suppress below their baseline firing rates. This

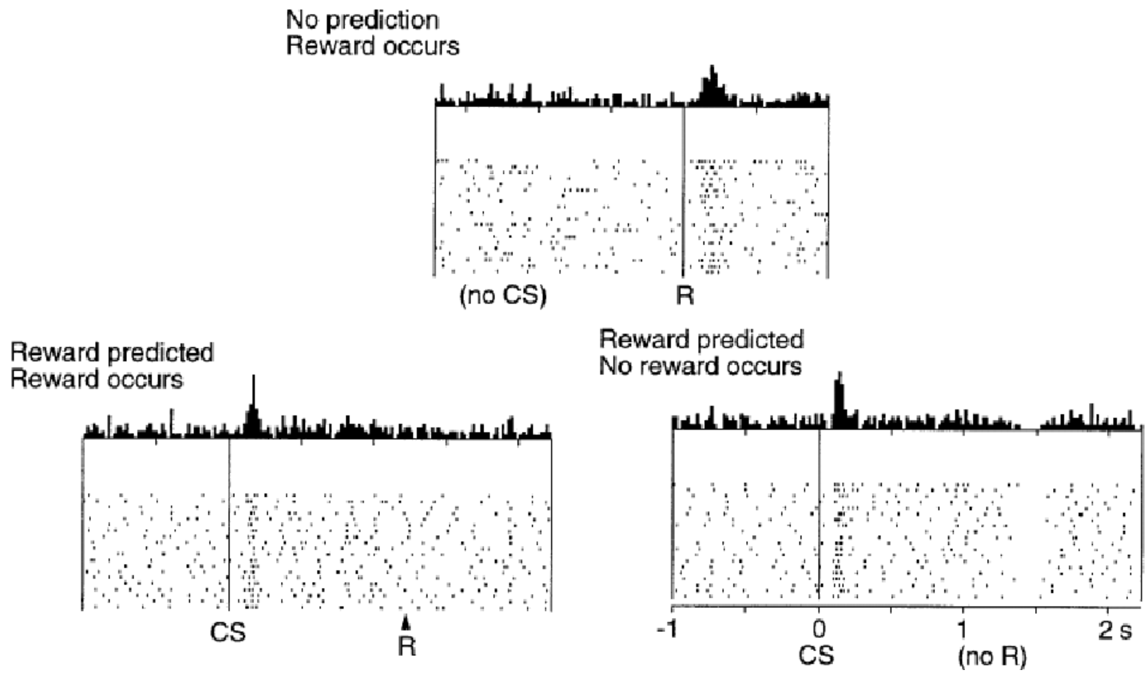

Figure 4: Firing rates of a population of dopamine neurons in the midbrain during classical conditioning (Adapted from ”A Neural Substrate of Prediction and Reward” by W. Schultz, P. Dayan, & P. R. Montague, 1997, *Science*, *275*, pp.1593–1599.)

indicates that the dopamine neurons are encoding for changes in expected future reward. They fire more strongly when the animal perceives a situation where it expects to soon be rewarded. However, they also fire less strongly when no rewards are provided at a time that they are expected. This is 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 realization has led a number of researchers to develop models of the dopamine system using TD 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, commonly referred to as the current *state*. The typical architecture of the adaptive critic is shown in Figure 5 where a_i is the activation value of the output unit i, V is the value of the current state, w_{ij} is the connection weight for the connection from unit j to unit i, and features

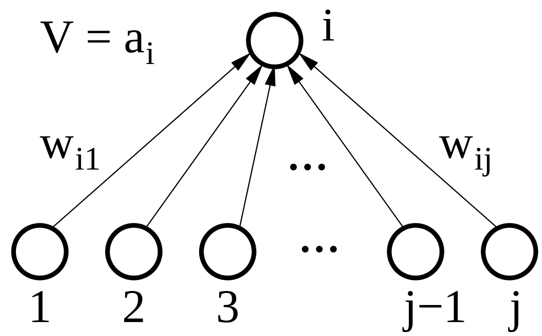

Figure 5: A typical adaptive critic architecture

associated with the current state are encoded across all of the input units (all j). Since the idea is to learn to predict how rewarding a state is, ideally we want the following to be true:

$$V(t) \;=\; r(t) + r(t+1) + r(t+2) + \ldots + r(t+n) \tag{6}$$

This would make the our predictions of future rewards V at time t be the sum of all reward we will receive as a result of being in our current state. However, we might look at rewards that are received soon as being more important than rewards received more distantly in the future. (This is not always the case, but this method has proved reliable for learning (Sutton, 1988).) We can discount future rewards by multiplying subsequent future rewards by a factor γ which is between 0 and 1. Thus, the prediction of reward changes to the following:

$$V(t) \;=\; \gamma^0 r(t) + \gamma^1 r(t+1) + \gamma^2 r(t+2) + \ldots + \gamma^n r(t+n) \tag{7}$$

However, we don't know these future reward values and, therefore, cannot calculate V. However, it is not necessary to have all of these reward values to begin making predictions. One can instead use the prediction of reward for the next state, where one ends up, $V(t{+}1)$, in order to approximate our prediction of future rewards for the current state $V(t)$. Here is how this occurs: Since,

$$V(t) \;=\; \gamma^0 r(t) + \gamma^1 r(t+1) + \gamma^2 r(t+2) + \ldots + \gamma^n r(t+n) \tag{8}$$

and

$$V(t+1) \;=\; \gamma^0 r(t+1) + \gamma^1 r(t+2) + \ldots + \gamma^{n-1} r(t+n) \tag{9}$$

then it follows that:

$$V(t) \;=\; \gamma^0 r(t) + \gamma^1 V(t+1) \tag{10}$$

This means that if our predictions of V are reliable, then we only need the reward received at the current time step and our prediction for the next time step to accurately predict expected future rewards. It is up to the adaptive critic to approximate this function, $V(t)$, in order to accurately predict changes in expected future reward. The critic needs to incrementally adjust its predictions based on reward information from each time step in order learn this approximation. In order to accomplish this, the critic calculates a change in expected future rewards:

$$\delta \;=\; (r(t) + \gamma V(t+1)) - V(t) \tag{11}$$

Note that this difference is derived from the above equations governing expected future rewards. In particular, δ is zero when the appropriate relationship holds between $V(t)$ and $V(t+1)$. Weights in the adaptive-critic network that participated in the computation of V are then adjusted in the following manner:

$$\Delta w_{ij} \;=\; \lambda_{td}\, \delta\, a_i\, f'(net_i) \tag{12}$$

where Δw_{ij} is the amount to change the weight leading from unit j to unit i, λ_{td} is the learning rate, and $f'(net_i)$ is the derivative of the activation function of unit i evaluated at the net input for unit i. Also note that $V = a_i$ where a_i is the activation value of unit i. The δ term is called the *temporal difference (TD) error*. This global error value can be used to drive learning in the adaptive critic, as shown in the weight update equation above, improving its ability to predict expected future reward. This error term can also be used to drive learning in other neural systems that select actions, driving these systems to choose actions that regularly lead to reward. This is how learning is accomplished in the *actor-critic framework* which has been used to explain motor sequence learning in the striatum as well as other forms of learning (Barto et al., 1990; Barto, 1994; O'Reilly et al., 2002). The adaptive critic can be seen as computing what is called the *value function* in the reinforcement learning literature while the actor computes what is known as the *policy function*. While the value function is an estimate of the total expected future reward for any state encountered, the policy function determines an action that proceeds to the next state. However, the adaptive critic can be used to choose actions without the aid of an actor network by predicting rewards for all future states that result from taking every action that is available to the system and choosing to take the action that leads to the most rewarding state (i.e. the state that has the highest estimate of total future reward.)

TD has been used extensively to learn sequences of overt actions, which are commonly motor skills. However, in some models, TD has been used to explain the coordination of covert cognitive activities like updating working memory contents (Braver and Cohen, 2000; O'Reilly et al., 2002). These models are good examples of how TD can control covert cognitive activities through reinforcement learning. The allocation of attention can also be thought of as a covert cognitive function. Therefore, I 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.

CHAPTER III

METHODS

Conjunctive Coding

The attention weights in ALCOVE take on continuous values to represent varying levels of attention for the various stimulus dimensions. However, most applications of temporal difference learning focus on choosing from among a set of discrete actions, while there is no clear understanding of how to apply these methods to domains where continuous output is needed (Sutton, 2001). Therefore, some modification to standard TD learning is needed to apply these techniques to the problem of learning dimensional attention. I devised two novel connectionist architectures to accomplish this. I chose to encode 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 preferred attentional weight vector, and the activation of a unit encourages the use of that unit's preferred dimensional attention weights. The activation level of each unit is mainly determined by its individual bias weight. These bias weights are adapted using the TD learning method to optimize reward. This very simple, single-layer network forms the backbone of selective attention learning in my models.

At the start of each trial, each of the attention map units computes its activation value based on its bias weight. The attention map units then compete in such a way so as to effectively select a vector of attention weight values to be used by ALCOVE. (This competition varies across models and is described below.) With the new set of attention weights in place, ALCOVE processes the current trial in the usual fashion and produces a categorization judgment. In response to feedback, ALCOVE's association weights are adjusted in the usual manner using the delta rule. However, the adjustment of the dimensional attention weights is handled differently. If ALCOVE confidently chooses the correct category, it is rewarded. Otherwise, it is not. The TD error, δ, computed as a function of this reward signal is used to modify the bias weights of all active attention map units. Over multiple trials, the bias weights are strengthened for attention map units whose associated dimensional attention weights regularly lead to reward.

Two different architectures for the attention map were investigated. The first of these employed *conjunctive coding*, resulting in a *localist* representation of dimensional attention. In this scheme, the preferred dimensional attention weight vectors of the attention map units were distributed evenly throughout the attention weight vector space. Therefore, each unit corresponded to a position in weight vector space and the attention map layer consisted of a grid of units that spanned this space. On each trial, a simple winner-take-all competition determined the single attention map unit whose preferred weight vector would specify dimensional attention for that trial. Learning occurred only for the winning unit, where the weight update equation for its bias weight was as follows:

$$\Delta w_i \ = \ \lambda_r \, (r - a_i) \, f'(net_i) \tag{13}$$

where Δw_i is the weight adjustment for the bias weight for attention map unit i, λ_r is the attention map learning rate parameter, r is the reward for the current trial, a_i is the activation value of the winning attention map unit (i), and $f'(net_i)$ is the derivative of the activation function (the standard logistic sigmoid). This is the standard TD weight update equation under the condition of *absorbing rewards*, where we do not predict reward past the end of the current trial. In this case, a_i acts as our reward prediction ($V(t)$), and we do not predict beyond this trial, so $\gamma V(t+1) = 0$, so $\delta = r - a_i$. A reward value (r) of $+1$ was delivered to the network on trials where ALCOVE selected the correct category label with a confident response. In other words, this reward was delivered when all output unit activations were within 0.5 of their respective targets. Otherwise, a reward of 0 was delivered.

Tile Coding

Our second attention map architecture employed *tile coding*, resulting in a *distributed* representation of dimensional attention. Here the the units in the attention map layer were partitioned into *tilings*, where each disjoint tiling contained a set of units with preferred dimensional attention weight vectors that uniformly spanned the entire attention weight space. However, the preferred dimensional attention vectors were not identical because each tiling was "offset" from the others as shown in Figure 6. In order to precisely represent a position in attention weight space, exactly one unit from each tiling needs to be active and the *tiles* surrounding the unit positions from each tiling need to overlap. This kind of distributed representation was originally used in the Cerebellar Model Articulation Controller (CMAC) (Albus, 1975), and its use in other TD learning systems has been found to result in improved generalization (Sutton, 1996). Just as in the conjunctive coding architecture, each attention map unit is activated by a bias weight and a competition between units ensues. However, in the tile coding scheme the most active unit restricts activity in other tilings. First, the most active unit across all of the tilings in identified. Then, the most active unit in each tiling whose corresponding tile overlaps with the first winning unit's tile remains active while the activity of the other units in the tiling are suppressed. This restriction is applied recursively to all tilings, allowing only one active unit per tiling, whose tile overlaps with all other active tiles. The attention weight vector corresponding to the center of this overlapping region is used by ALCOVE to process the current stimulus. Once feedback has been 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 (active) units in the attention map layer.

In standard ALCOVE, the initial attention weights are often set to all be equal and sum to 1. This effectively emphasizes all dimensions equally from the start of training. We selected initial bias weights in the attention map layer in order to form a similar initial bias in our models. The unit in the attention map whose preferred dimensional attention vector

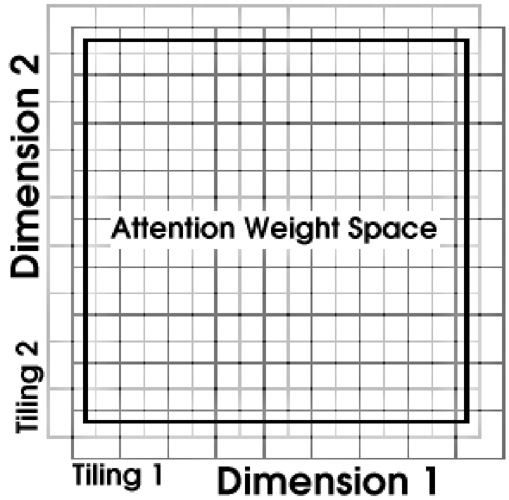

Figure 6: Tile coding of the attention map layer – A single unit is centered in each tile

matched the initial vector used in standard ALCOVE was assigned an initial bias weight value of 0.05. Bias weights assigned to to the other units in the attention map fell off in a Gaussian fashion as distance from this position increased, with a lower bound of -0.05. A small amount of uniform noise was then added to each bias weight and the resulting weights were clipped to remain in the $[-0.05, 0.05]$ range. The variance of the Gaussian and the range of the uniform noise were free parameters of the model.

CHAPTER IV

RESULTS

In order to verify that these reinforcement learning mechanisms could indeed account for the learning of dimensional attention and human performance, the models were applied to several category learning tasks previously studied in the category learning literature. The performance of the models was compared to standard ALCOVE and also ALCOVE without attention learning. If the reinforcement learning mechanisms can learn dimensional attention appropriately, then the new models should closely match the performance of standard ALCOVE. However, if the conjunctive and tile code models deviate from the performance of standard ALCOVE, and behave in a manner similar to ALCOVE without attention learning, then we can say that the reinforcement mechanisms are not producing appropriate dimensional attention weights. Tasks in which standard ALCOVE and ALCOVE without attention learning perform equally well show that dimensional attention does not play a significant role in these tasks. Thus, any degradation in performance on these tasks for the conjunctive and tile code models indicates that the reinforcement learning mechanisms are actually interfering with the learning of these tasks. The following experiments show that the new models learn useful dimensional attention weights in tasks that require dimensional attention to match human performance, and they do not interfere with the learning of tasks where dimensional attention does not aid performance. In summary, the models can account for differences in learning speed as a function of category structure, the utility of dimensional attention when stimulus dimensions are separable, and the performance consistency of dimensional attention when stimulus dimensions are integral.

Dimensional Attention & Learning Difficulty

Shepard, Hovland, and Jenkins examined the effect of category structure on the relative speed with which a category structure is learned (Shepard et al., 1961). Stimuli were composed of three easily separable binary dimensions, for a total of eight possible stimuli. Six category structures where examined as shown in Figure 7. Each of the eight resulting stimuli was assigned to one of two categories which are represented in the figure as either a filled (black) corner or a blank (white) corner. The resulting structures were then ordered based the relative difficulty in learning the structures. The Type I category structure requires only information about the first dimension (dim 1) in order to correctly categorize the stimuli. Thus, it is the easiest to learn. However, the Type 2 category structure requires information about two dimensions (dim 1 and dim 2) to make a correct decision and should be more difficult to learn. The remaining category structures require attention to all three dimensions to make correct decisions, but certain dimensions are more informative in some cases (i.e., most of the stimuli for Type 4 can be classified based solely on information about the third dimension (dim 3), but there are two "exceptions" that require attention to either of the other two dimensions for their classification). Therefore, the remaining category structures are ordered based on their relative dimensional usage across all dimensions.

The subjects in this study were trained on a stimulus set consisting of three separable, binary dimensions: shape (triangle or square), size (large or small), and color (filled or blank). The number of trials taken to reach categorization proficiency was recorded in each task. The subjects learned the Type I category structure the fastest, the Type II structure next, and so on for all six structures. However, there was not a significant difference in the learning difficulty of Types III, IV, and V, which only differ slightly (i.e. their relative dimensional usage across all dimensions is the same, but their stimuli assignments are slightly different). Shepard et.al. argued that category learning models based on reinforcement learning could not account for this learning order without incorporating some mechanism for selective dimensional attention. Therefore, this study provides a framework for testing the biologically-plausible dimensional attention learning mechanisms of the conjunctive code and tile code models against ALCOVE's backpropagation mechanism for learning dimensional attention.

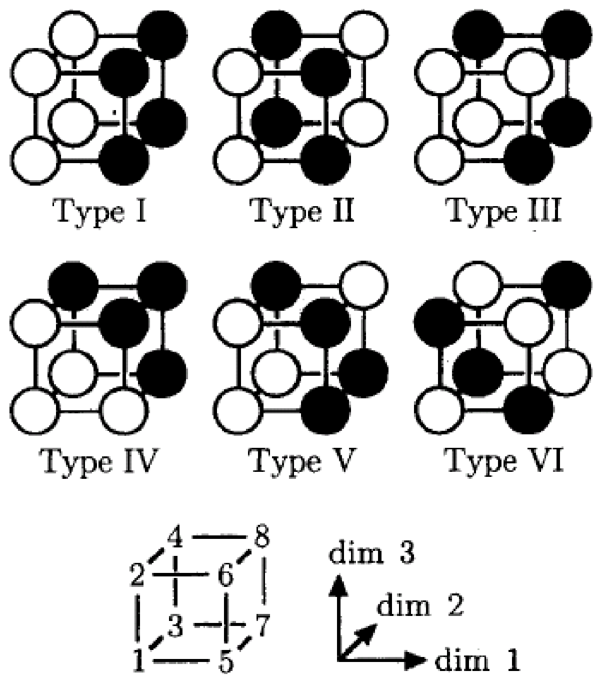

Figure 7: Category structures used by Shepard et.al., (1961). (From "Learning and Memorization of Classifications" by R. N. Shepard, C. L. Hovland, & H. M. Jenkins, 1961, *Psychological Monographs*, *75*, 13, Whole No. 517, p. 4. In the public domain.)

Table 1: Parameter values for attention learning demonstration using the Six Type data (Shepard et al., 1961).

Parameter	Parameter Values	
	ALCOVE - Standard	ALCOVE - No α Learning
Specificity	6.5	6.5
Association Weight Learning Rate	0.03	0.03
Attention Weight Learning Rate	0.0033	0.0
Luce Choice Gain	2.0	2.0

ALCOVE is clearly able to account for this learning order through its attention learning mechanism, but not without it. Figure 8 shows how ALCOVE performs on these category structures involving binary stimuli when attention learning is either enabled or disabled. Standard ALCOVE is able to account for the proper learning order using its attention learning mechanism. Also, Shepard et.al. could not find a significant difference in the relative learning difficulty for Types III, IV, and V, and ALCOVE also displays this behavior, though the type V difference is slightly more pronounced. (A difference was found in the empirical study, but it was not statistically significant.) However, without its dimensional attention learning mechanism, ALCOVE fails to learn these tasks in the proper order. (In particular, the Type II task is learned much too slowly.) Indeed, no matter what other parameter values are chosen, if the attention learning mechanism is disabled, the learning order remains improper. The learning curves in Figure 8 were generated using online learning and all eight stimuli were seen in random order on each epoch (permuted) with parameter values as shown in Table 1.[1]

The challenge for the conjunctive code and tile code models is to account for this learning difficulty as well as standard ALCOVE. Therefore, the models were applied to the Six

[1]This is a replication of a study from Kruschke (1992) except I used online learning instead of batch learning. In batch learning, weight updates that result from the presentation of all eight stimuli are not applied until all of the stimuli have been presented. An *epoch* consists of one complete pass through all of the training patterns (stimuli). Therefore, in batch learning, weights are updated once per epoch. However, in the online learning scheme, weights are updated after each stimulus presentation. ALCOVE typically uses batch learning.

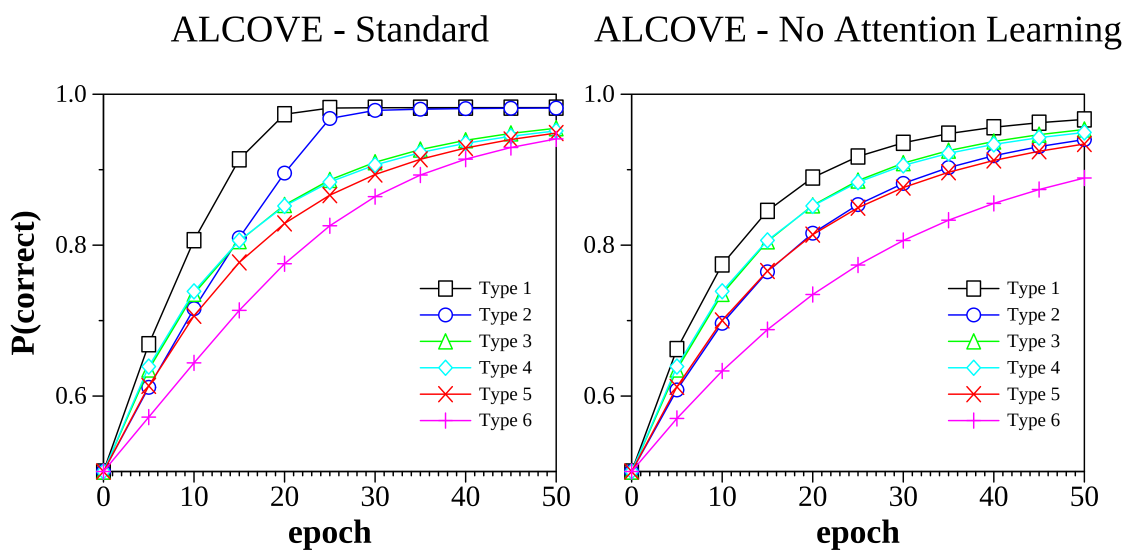

Figure 8: Performance on Six Type category structures for ALCOVE with attention learning either enabled (left) or disabled (right).

Type categories for comparison. Since the stimuli were composed of three dimensions, the attention weight space was three-dimensional. The conjunctive coding model used a $15 \times 15 \times 15$ unit topology for 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). Note that while these two models contained a comparable number of units in their attention map layers, the use of distributed representations in the tile coding case both increased the precision with which weight vectors could be specified and offers the promise of improved generalization. The results of these simulations are shown in Figure 9 and Figure 10. Learning was done online and all eight stimuli were seen in random order on each epoch. All parameters were manually selected and are shown in Table 2.

The graphs in Figure 9 show a representative individual run for both the conjunctive code and tile code models. Both models display the desired learning traits, similar to standard ALCOVE. In particular, the Type II structure is learned faster than the Type III, IV, V, and VI structures, but still slower than the Type I structure. This behavior is critical to the validation of the conjunctive and tile coding approaches to learning dimensional attention.

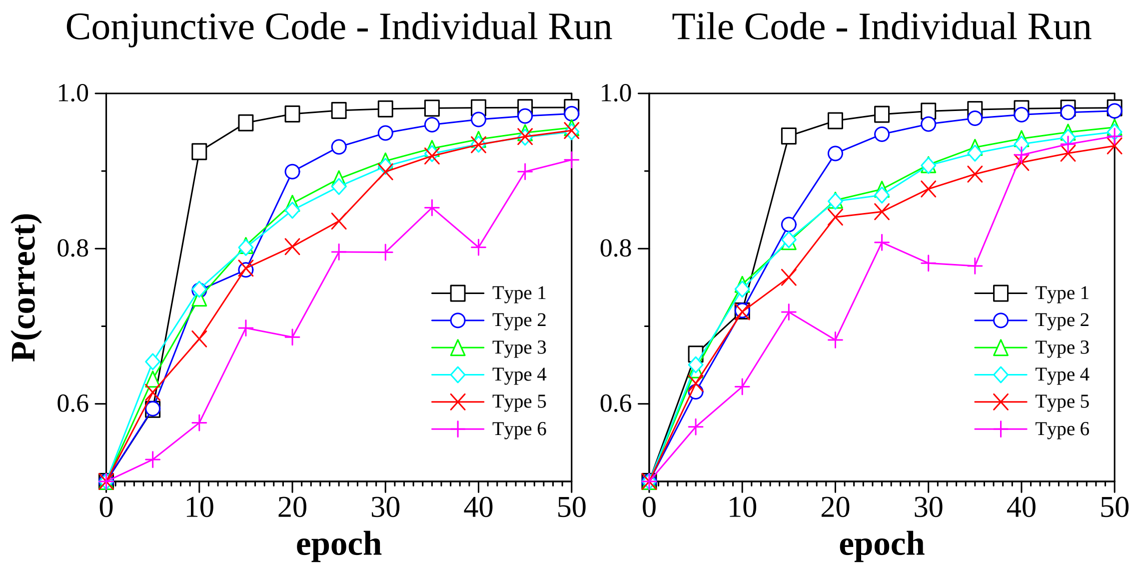

Figure 9: Performance on Six Type category structures for conjunctive and tile coding models – results for an individual representative run

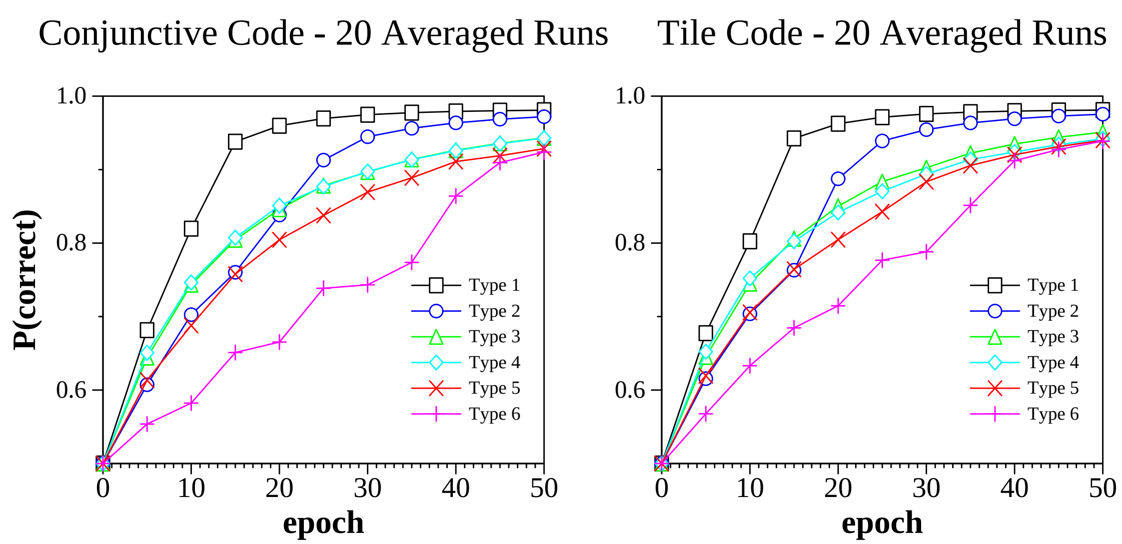

Figure 10: Performance on Six Type category structures for conjunctive and tile coding models – results averaged across 20 individual runs

Table 2: Parameter values for conjunctive code and tile code models using the Six Type data (Shepard et al., 1961)

	Parameter Values	
Parameter	Conjunctive Code	Tile Code
Specificity	6.5	6.5
Association Weight Learning Rate	0.033	0.033
Reinforcement Learning Rate	0.001	0.001
Luce Choice Gain	2.0	2.0
Gaussian Variance	0.8	0.8
Uniform Noise Variance	0.008	0.01

If the models could not account for the empirical results from Shepard et.al. then we would know for sure that they were not performing attention learning correctly. Note, however, that there are obvious differences in the conjunctive and tile code models' performance compared to standard ALCOVE, even though they learn the tasks correctly. ALCOVE displays a smooth learning curve, indicative of the gradient-descent learning performed by ALCOVE. The conjunctive code and tile code models do not display this smooth learning phenomenon due to the stochastic nature of the winner-take-all mechanism combined with TD learning that determines dimensional attention weights for the models. This result is interesting because ALCOVE assumes that people update dimensional attention in a graded, incremental fashion in order to solve category learning tasks. Since the conjunctive code and tile code models can account for the same performance using a stochastic mechanism, people might possibly be learning dimensional attention in this stochastic fashion, instead. Indeed, ALCOVE might be thought of as matching the average performance across a population of individual subjects. This can be seen in comparing the results in Figure 10 (where the average learning curve from twenty individual runs of each model is shown) to the standard ALCOVE learning curves from Figure 8. ALCOVE could easily be fit to human performance data from individual subjects, but it could not show stochastic shifts in dimensional attention and learning like the conjunctive and tile code models. However, we do not have empirical data for individual subjects to compare with our models. Therefore, while we do not know how stochastic the empirical data really is, ALCOVE predicts it is smooth while the conjunctive code and tile code models predict this more stochastic learning. This would be an interesting result to examine empirically.

Even though the Six Type tasks provide a good framework for testing the qualitative performance of the conjunctive code and tile code models, they have several limitations that need to be addressed. First, the stimuli in the tasks are binary in each dimension (i.e. they only take on one of two possible values for that dimension.) However, the stimuli in category learning tasks are often continuous in each dimension. That is, they can take on one of several possible values for each dimension. Size could range from small to medium to large or color could range from blank (white) to grey to filled (black.) It could be the case that the reinforcement learning mechanisms employed by the models could fail to properly learn dimensional attention for tasks involving continuous stimuli. Also, the empirical data on the Six Type tasks only allows for an analysis of model learning speed for these structures and does not provide a framework for analyzing how well these models capture human generalization performance. Finally, the Six Type tasks involve separable stimuli, where a change in the value of one dimension does not affect the perception of the other dimensions. If integral stimuli are used, where perception of other dimensions is affected by a change in any single dimension, the importance of dimensional attention decreases to a large degree. But it is important that category learning models that incorporate dimensional attention match human performance even when integral stimuli are used – that dimensional attention mechanisms do not hinder the ability of the models to fit human performance. Again, the conjunctive code and tile code models might fail to properly learn dimensional attention (possibly causing the models to fail to account for performance as well as standard ALCOVE does), even though the weights are not as important in this case.

In order to address these concerns, two more experiments were conducted to examine the performance of the conjunctive code and tile code models on stimuli of these kinds. Both experiments involve continuous stimuli to address the binary limitation of the Six

Type tasks. Also, both experiments involve fitting the models to actual human response probabilities at arbitrary times during training. Finally, one of the experiments uses stimuli with separable dimensions and the other with integral dimensions.

Categorization of Separable Dimensioned Stimuli

Nosofsky conducted a category learning study involving stimuli with separable dimensions and then fit the GCM to the resulting human performance data in order to show the effectiveness of the GCM (Nosofsky, 1986). The stimuli consisted of two continuous dimensions: a semicircle inscribed with a radial line. The semicircles varied across four radius lengths, and the radial lines varied across four angles for a total of sixteen possible stimuli. The four different category structures in Figure 11 were designed using these stimuli. For each category structure, eight of the stimuli were explicitly assigned to one of two categories. These were the stimuli used in the training phase of the study, while the remaining eight were used to assess generalization. Subjects underwent an approximately 1,200 trial training phase on a category structure and then an a 3,500 trial transfer phase on the same structure. (During the transfer phase in this study, feedback was provided for stimuli that were explicitly assigned to a category.) Each category structure was learned by the subjects in this manner. The response probabilities for all sixteen stimuli were calculated based on the results from the transfer phases for each category structure. The GCM was then fit to these response probabilities.

In order to assess the performance of the conjunctive code and tile code models on these category learning tasks, a slightly different strategy from that used with the Six Type data was taken. ALCOVE with attention learning and without attention learning was fit to the response probabilities in the transfer phase after being trained for a total of 1,200 trials (as in Nosofsky's study). The same was also done for the conjunctive code and tile code models. A simple hill-climbing optimization algorithm minimizing the sum-squared error between network generated response probabilities and the Subject 1 response probabilities was used to fit the free parameters of the models (4 for ALCOVE as listed in Table 1, 6 for

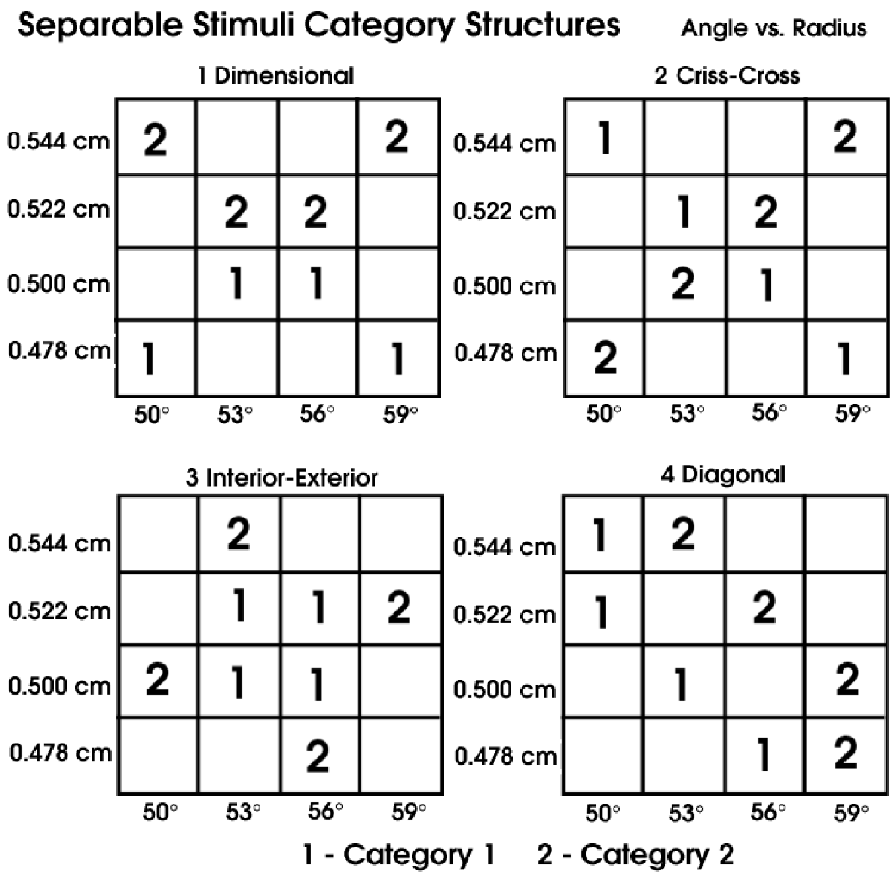

Figure 11: Category structures for the sixteen stimuli with separable constituent dimensions (Nosofsky, 1986)

conjunctive code and tile code models as listed in Table 2) for each category structure. The conjunctive code attention map layer was arranged in a 15×15 unit topology (225 units total), while the tile code model used 9 tilings of 5×5 units (225 units total.) The stimuli were presented to the models using the multidimensional scaling code found by Nosofsky.

The quality of the fits are summarized in Figure 12. The standard ALCOVE model provided the best overall fits to the data, but the conjunctive code and tile code models performed well at matching the subject response probabilities as well. The only noticeable difference in performance occurred using the tile code model on category structure 3. This is a very difficult category structure to learn and the tile code model does not even perform as well as ALCOVE without attention learning. Thus, even though the tile code model has several computational benefits over the conjunctive code model, it does not account for human performance as well. The dimensional attention weight values that might actually aid the model in solving this (relatively difficult) task might be represented in a very small area of weight space. Thus, the tile code model's generalization method might actually be hindering its performance because surrounding areas of weight space, which fail to produce reward, might make the area containing the appropriate values less likely to be discovered. This could more than likely be overcome by using more units in a tiling and fewer total tilings. However, for category structures 1 and 2, the conjunctive code and tile code models show significant performance benefits over ALCOVE without attention learning. The conjunctive and tile code models do almost as good a job as ALCOVE. It would be unreasonable to expect them to do better since the gradient-descent learning methods employed by standard ALCOVE are extremely powerful. Instead, the goal of these models is to perform almost as well as standard ALCOVE while using biologically-based reinforcement learning mechanisms. Overall, the models seem to account for human performance and learn adequate dimensional attention in tasks involving continuous stimuli.[2]

[2]The GCM fits are from Nosofsky (1986). I provided fits for the other four models.

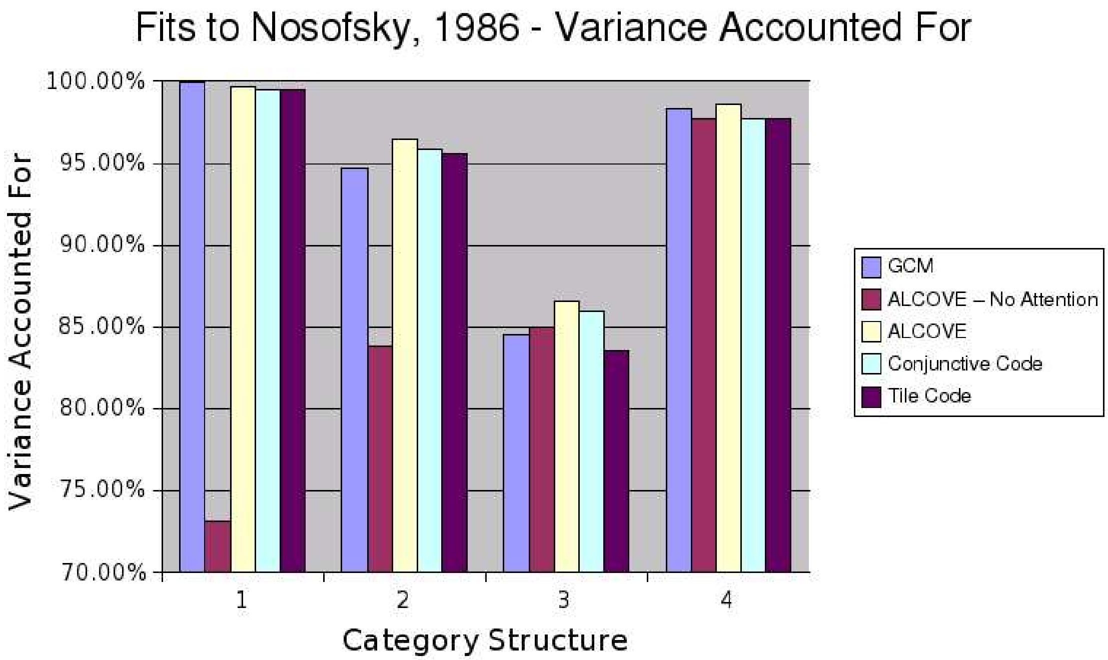

Figure 12: Fits to human performance on category structures involving continuous, separable stimuli

Categorization of Integral Dimensioned Stimuli

Nosofsky also conducted a category learning study involving stimuli with integral dimensions and then fit the GCM to the resulting human performance data in order to show the effectiveness of the GCM in capturing human performance even when stimulus dimensions were not separable (Nosofsky, 1986). The stimuli consisted of twelve colored chips from the popular Munsell chip set manufactured by the Munsell Color Company. The colors on these chips vary along two dimensions: saturation and brightness. These dimensions are continuous and any change in one dimension affects human perception of the other dimensions. Twelve such chips were partitioned into six different category structures as shown in Figure 13. Some or all of the stimuli were assigned to categories for use in the training phase and all twelve stimuli were used during the transfer phase. Each of the category structures were studied for a fixed number of trials as shown in Table 3.

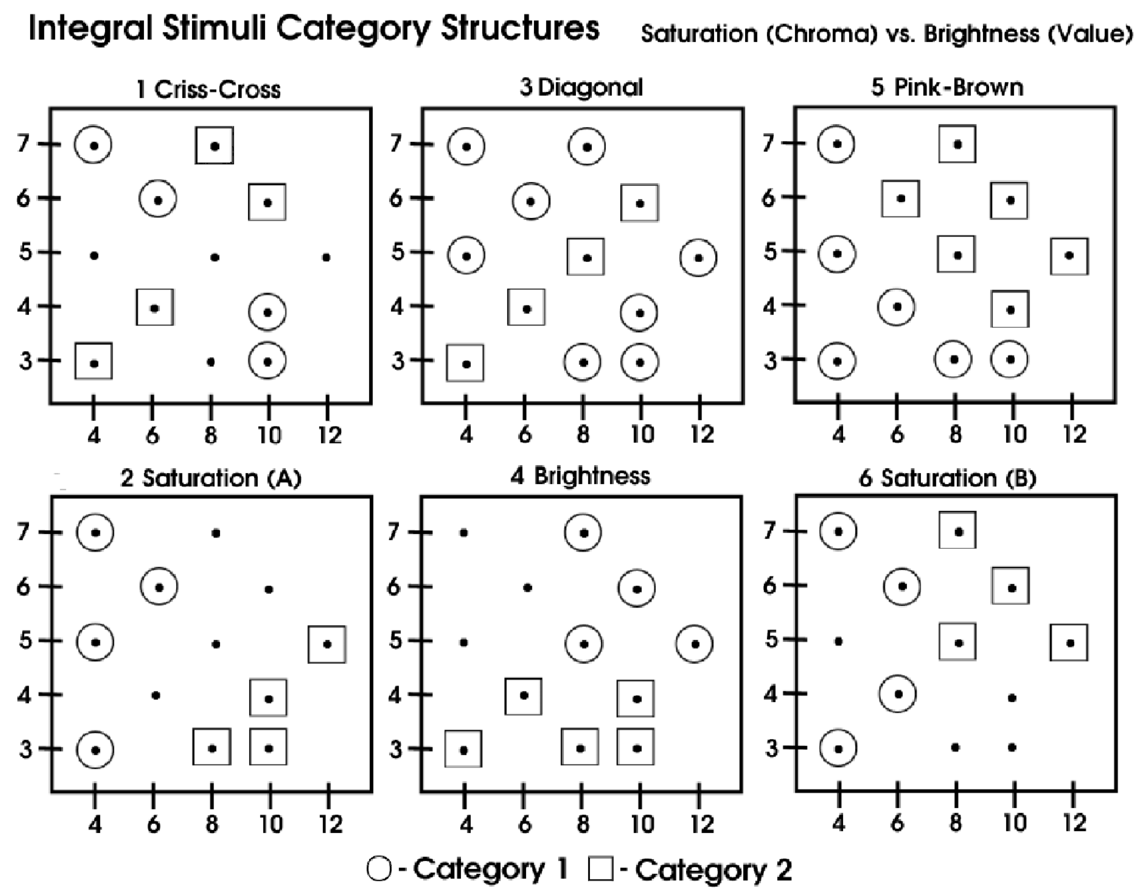

Figure 13: Category structures for the twelve stimuli with integral constituent dimensions (Nosofsky, 1987)

The responses of multiple subjects for all twelve stimuli during the transfer phase were aggregated to produce the response probabilities fit by the GCM.

The same fitting process explained earlier was used to fit ALCOVE, the conjunctive code, and the tile code models to the response probabilities for all twelve stimuli from the transfer phase associated with each category structure. Since the attention weight space was two-dimensional, as in the separable stimuli study, the attention map layers were identical to those used in the previous simulations. Also, the number of training trials done by each model before fitting the model to the data for each category structure was equal to those used in the human study (shown in Table 3.) Again, the stimuli were presented to the models using the MDS code found by Nosofsky.

Table 3: Training trials for integral stimuli category structures

Category Structure	Name	Number of Training Trials
1	Criss-Cross	240
2	Saturation (A)	180
3	Diagonal	380
4	Brightness	490
5	Saturation (B)	400
6	Pink-Brown	400

The quality of the fits are summarized in Figure 14. The standard GCM model provided the best overall fits to the data, but ALCOVE, the conjunctive code, and the tile code models performed well at matching the subject response probabilities as well. However, all other models showed significantly poorer performance than the GCM for category structure 3. This structure has roughly the same layout as the category 3 structure from the separable stimuli study so, again, problem difficulty may be to blame. In the GCM, the attention weights are free parameters discovered by an extensive fitting process and it may just be that this problem requires a strange combination of attention weight parameters that the other models find difficult to learn. Again, the tile code model suffers more than the others which only strengthens this hypothesis. The standard ALCOVE fit for category structure 2 is excellent. The conjunctive code and tile code models had to learn this task rather quickly and so the rewards might not have been strong long enough to push the dimensional attention weights into the same area of attention weight space as the gradient-descent learning of ALCOVE. Even with the structure 3 learning limitation and slightly worse fits on structure 2, the conjunctive code and tile code models fit the data roughly as well as ALCOVE and do not fail to do so when integral stimuli are used.[3]

[3]The GCM fits are from Nosofsky (1987). I provided fits for the other four models.

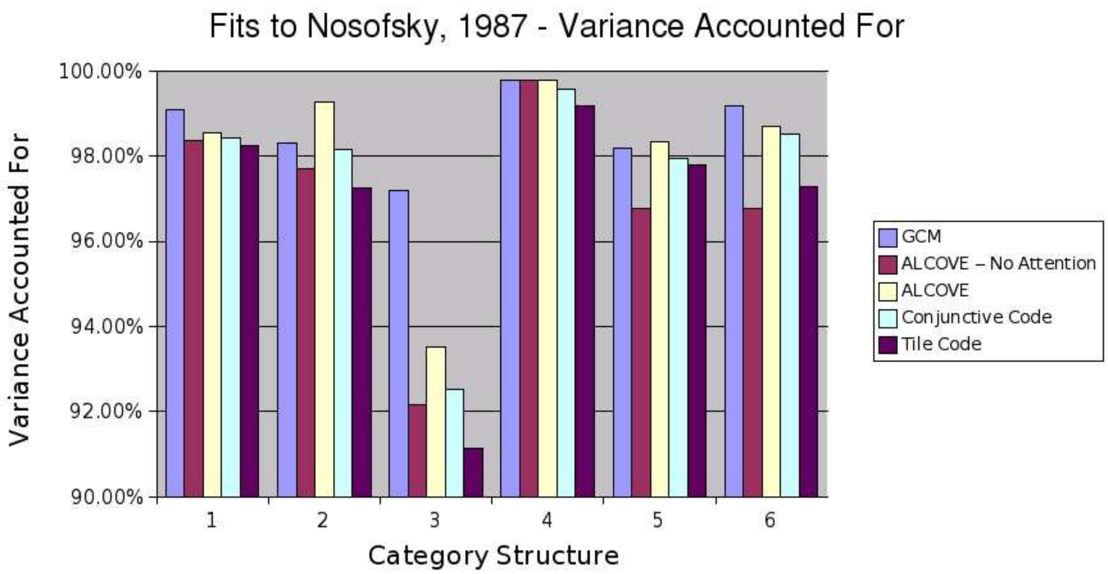

Figure 14: Fits to human performance on category structures involving continuous, separable stimuli

CHAPTER V

DISCUSSION & FUTURE WORK

The 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 two new models actually fit the human performance data quite well. Even though ALCOVE generally performed the best of all of the models examined, the new models were able to learn useful dimensional attention weights from their less-informative global reinforcement signal. The TD error signal is less-informative in that is does not supply the network with information about the direction in which the attention weight parameters should be changed. The backpropagation of error algorithm provides this dimension-specific information when updating attention weights (for example, the attention weight for orientation needs to be increased and the attention weight for size needs to be decreased) through adapting weights based on gradient-descent in network error. However, the TD learning algorithm has to use the underlying architecture to discern how to change attention weight values. Thus, it strikes a balance between exploring new attention weight values and exploiting the attention weight values it has learned for producing rewarding responses.

In contrast to ALCOVE, the mechanisms described here are stochastic in nature. That is, attention weights can sometimes change drastically in an attempt to explore better attention weight combinations while other times they may stay close to a particular area of weight space. This stochastic behavior comes from the initialization methods described above, where uniform random noise is injected into the bias weights of the attention learning networks. If a particular combination of values does not get rewarded over several trials, the biases of the units encoding that combination may become low enough that another unit on the other side of the space could become more highly active and begin to win the winner-take-all competitions between attention map units. Thus, the attention learning produced by these TD learning mechanisms does not lend itself to slow, incremental changes in dimensional attention. Instead, these models are extremely dynamic, but still capable of learning useful dimensional attention representations. Other models of categorization like RULEX (Nosofsky et al., 1994) as well as other studies in category learning (Rehder and Hoffman, 2003) see the slow, incremental changes in dimensional attention as an artifact of averaging data across multiple subjects. The models presented here show how dimensional attention that is highly stochastic can be leveraged properly for good categorization performance. If it is determined that the individual learners display rapid, stochastic shifts in dimensional attention, the conjunctive code and tile code models may provide a better fit to individual performance than the necessarily smooth-shifting ALCOVE.

In all of the proposed TD models, individual attention map units encode *conjunctions* of attention weights – one weight value for each dimension. This encoding strategy was adopted for a good reason. It turns out that if attention weight space is not represented in this conjunctive fashion, then good attention weight values can often get unduly penalized. This occurs because dimensional attention weights do not operate independently from one another. Raising a single dimensional attention weight effectively lowers the amount of attention paid to all other dimensions as well as increasing attention for that dimension. A disjunctive coding of attention weight space, where the attention weight parameters are selected independently, does not capture this effect and leads to a general failure to learn when combined with TD learning. The conjunctive code overcomes this problem by not penalizing the network for producing bad attention weight values, but by penalizing bad conjunctions of attention weight values. The opposite consequence of rewarding the network for choosing poor attention weight values occurs for disjunctive encodings, as well, and is overcome by the conjunctive code approach. These advantages are achieved with the tile code model which uses a distributed, conjunctive encoding of attention weight values, as well as the conjunctive code model which uses a localist, conjunctive encoding.

Disjunctive representations make TD learning extremely unstable and some form of conjunctive coding seems to be necessary for proper learning. This is true in other reinforcement learning schemes as well (Sutton, 1996). However, the conjunctive code model has one major drawback common to all localist conjunctive representations. The number of attention map units required to adequately cover the attention weight space grows exponentially in the number of stimulus dimensions. This limits the scalability of conjunctive coding approaches in connectionist modeling. Also, conjunctive coding results in a localist representation of attention weight space. Therefore, conjunctive coding lacks the ability to generalize without some additional mechanisms to aid in this process. Therefore, the tile code attention map architecture was examined, since it does not suffer as much in terms of scalability and also has the ability to generalize across conjunctions of dimensional attention weights. Even though the number of units needed to fully cover the attention weight space still increases exponentially with stimulus dimensionality, the tile code model uses significantly fewer units to achieve the same level of space discretization. Also, the tiles in the tile code model have overlapping receptive fields which provide for generalization between similar attention weight vectors. Thus, the tile code model has several clear computational advantages over the conjunctive code model. However, the conjunctive code model might still be the best representation for learning dimensional attention even with its apparent weaknesses. Other forms of distributed codes could be explored in the future. The use of *CPCA Hebbian learning* combined with *k-winners-take-all* inhibition has been proposed as a biologically plausible method of learning sparse distributed representations in an unsupervised or self-organizing fashion (O'Reilly and Munakata, 2000), where no target signal is needed to learn these representations. Other forms of unsupervised learning that use winner-take-all inhibition like the models explored here could also be investigated, like competitive learning (Rumelhart and Zipser, 1986) or Kohonen learning (Kohonen, 1984).

The use of conjunctive coding to learn the continuous attention weight parameters in

ALCOVE has other implications in the general reinforcement learning domain. There is no clear understanding of how to apply reinforcement learning methods to learning continuous parameters, but this work provides a plausible framework in which to accomplish this type of learning. The conjunctive coding approaches explored in this work overcome the limitations of disjunctive coding in a biologically plausible manner and the unsupervised methods mentioned earlier for learning sparse, distributed representations might also be incorporated into the framework used here. Thus, future research in applying the methods explored here to more general reinforcement learning problems that involve continuous parameters is promising.

The way in which reward was provided to the new models also plays a critical role in successfully learning dimensional attention. The current models employ a reward schedule where reward is given to the network for correct, confident responses as described earlier. This is in contrast to the more obvious reward strategy of stochastically making category judgments based on $P(K)$ and rewarding any correct judgment. The reward scheme that was used here was motivated by the "three-stage learning" profile exhibited by ALCOVE. In ALCOVE, the association weights are all initialized to zero and the backpropagated error signal is multiplied by these weights before it influences dimensional attention learning. Thus, the dimensional attention weights do not change much until the network has begun to generate strong responses. This initial conservatism with regard to attention weights is not exhibited if reward is delivered for correct responses with low confidence during the early parts of training. Initial attempts at using a reward scheme based only on correctness caused network behavior to deviate substantially from standard ALCOVE. So, even though the simpler reward schedule makes more sense intuitively, because it is more indicative of the actual reward from the environment that subjects receive on a trial by trial basis, it seems that some other neural mechanism may be influencing the effective reward that drives learning. Biological and computational mechanisms that could be responsible for this effect are currently being investigated.

Eventually I 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. In particular, the modeling of ALCOVE's exemplar layer in terms of the underlying neural mechanisms that the brain could be employing to account for distance between exemplars in psychological similarity space is an enticing area of future research.

BIBLIOGRAPHY

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*, chapter 11, pages 215–232. MIT Press.

Barto, A. G., Sutton, R. S., and Watkins, C. J. C. H. (1990). Learning and sequential decision making. In Gabriel, M. and Moore, J., editors, *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, pages 497–537. MIT Press.

Braver, T. S. and Cohen, J. D. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In Monsell, S. and Driver, J., editors, *Control of Cognitive Processes*, volume 18 of *Attention and Performance*, chapter 31, pages 713–737. MIT Press.

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

Erickson, M. A. and Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, 127(2):107–140.

Gluck, M. A. and Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117:227–247.

Gluck, M. A., Bower, G. H., and Hee, M. R. (1989). A configural-cue network model of animal and human associative learning. In *Proceedings of the 11th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ. Lawrence Erlbaum.

Kohonen, T. (1984). *Self-Organization and Associative Memory*. Berlin: Springer Verlag.

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

Luce, R. D. (1963). In Luce, R. D., Bush, R. R., and Galanter, E., editors, *Handbook of Mathematical Psychology*, pages 103–189. New York : Wiley.

McCloskey, M. and Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In Bower, G., editor, *The Psychology of Learning and Motivation*, volume 24, pages 109–165. Academic Press.

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

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.

Noelle, D. C. and Cottrell, G. W. (2000). Individual differences in exemplar-based interference during instructed category learning. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, Philadelphia, PA. Lawrence Erlbaum.

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. *Journal of Experimental Psychology: 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.

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

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

O'Reilly, R. C. and Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience*. MIT Press.

O'Reilly, R. C., Noelle, D. C., Braver, T. S., and Cohen, J. D. (2002). Prefrontal cortex and dynamic categorization tasks: Representational organization and neuromodulatory control. *Cerebral Cortex*, 12:246–257.

Rehder, B. and Hoffman, A. B. (2003). Eyetracking and selective attention in category learning. In Alternman, R. and Kirsh, D., editors, *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, Boston, MA. Lawrence Erlbaum.

Rosch, E. H. and Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7:573–605.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986a). Learning internal representations by error propagation. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1, chapter 8, pages 318–362. MIT Press.

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

Rumelhart, D. E. and Zipser, D. (1986). Feature discovery by competitive learning. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1, chapter 5, pages 151–193. MIT Press.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 22:325–345.

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. (1987). Towards a universal law of generalization of psychological science. *Science*, 237:1317–1323.

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.

Sutton, R. S. (2001). Reinforcement learning faq: Frequently asked questions about reinforcement learning. Retrieved March 22, 2004 from http://www.cs.ualberta.ca/ sutton/RL-FAQ.html.