COMPUTATIONAL ACCOUNTS OF ATTENTIONAL BIAS: NEURAL NETWORK AND BAYESIAN NETWORK MODELS OF THE DOT PROBE PARADIGM

By

AMITOJ SINGH LIKHARI

A THESIS PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN EXERCISE AND SPORT SCIENCES

UNIVERSITY OF FLORIDA

2005
Copyright 2005

by

Amitoj Singh Likhari
This thesis is dedicated to my family, especially my Dad, Sarab Jit Singh, who always pushed me to follow my heart and gave me the reason to pursue this Master’s. My Mom, Sukhjinder, who patiently supported me through every step of the way. My sisters, Tanvir and Kamalpreet, and my brothers-in-law, Parneet and Dhananjay, for helping and guiding me find what I truly believe in. Last but not least, my three nephews, Sukhsahegj, Sarguun and Manu, for making me realize that things were not that bad after all.
ACKNOWLEDGMENTS

Countless hours have gone into the preparation of this thesis. However, I could not have made it if I did not have people pushing and prodding me to do better. I would like to thank Dr. Christopher Janelle, my advisor, for helping and supporting the idea of my thesis even when it was not as concrete, and for letting me down easy on the numerous occasions that I presented him with what surely must be the worst writing possible. I would also like to thank my friends and peers at the Motor Behavior Laboratory at the University of Florida, Steve Coombes, without whose data and revisions to text, this work not have been possible, and Sarah Huie, for proofreading my first drafts and going a long way in improving the quality of this work. Special thanks go to Dr. Anand Rangarajan, Associate Professor at Computer and Information Science and Engineering, at the University of Florida, for helping me solidify the concepts and asking me questions to force me to think clearly. Over the last few weeks, I have spent a significant amount of time at the library, and I thank the one person who agreed to study there with me, Angela Duke. Finally, none of this would not have been possible if Aaron Duley, also of the Motor Behavior Laboratory, had not first suggested the idea of developing a computer simulation.
TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................ iv
LIST OF TABLES .................................................................................................................. ix
LIST OF FIGURES .............................................................................................................. xi
ABSTRACT ......................................................................................................................... xii

CHAPTER

1 INTRODUCTION ........................................................................................................ 1
  1.1 Cognitive Models of Anxiety ................................................................................. 2
  1.2 Computer Tools to Investigate Mechanisms of Attentional Bias ....................... 4
    1.2.1 Neural Networks ........................................................................................... 4
    1.2.2 Bayesian Networks ....................................................................................... 5
  1.3 Previous Research ................................................................................................... 5
  1.4 Limitations .............................................................................................................. 9
  1.5 Statement of Problem ........................................................................................... 10
  1.6 Statement of Purpose ............................................................................................ 10
  1.7 Current Study ........................................................................................................ 10
    1.7.1 Objectives of the Neural Network Model .................................................. 11
    1.7.2 Objectives of the Bayesian Network Model ............................................... 12
  1.8 Hypotheses ............................................................................................................ 12

2 REVIEW OF LITERATURE ....................................................................................... 14
  2.1 Anxiety ................................................................................................................. 15
  2.2 Attentional Bias .................................................................................................... 22
    2.2.1 Dichotic Listening Paradigm ...................................................................... 23
    2.2.2 Stroop Task ................................................................................................... 24
    2.2.3 Dot Probe Task .............................................................................................. 29
      2.2.3.1 Initial studies (basic dot probe task) ......................................................... 30
      2.2.3.2 Manipulation of stimulus duration ......................................................... 37
      2.2.3.3 Backward masking ................................................................................ 38
      2.2.3.4 Pictorial dot probe task ....................................................................... 39
      2.2.3.5 Social anxiety ....................................................................................... 42

v
2.5.4 Conditional Probability ................................................................. 80
2.5.5 Chain Rule .................................................................................. 80
2.5.6 Bayes’ Theorem ........................................................................ 81
2.5.7 Conditional Independence ............................................................ 82
2.5.8 Graphical Notation ..................................................................... 83
2.5.9 Causal Networks and d-separation .............................................. 84
2.5.10 Bayesian Networks ................................................................. 85
2.5.11 An Example of a Bayesian Network ........................................... 90
2.6 Summary ......................................................................................... 94

3 METHODS .......................................................................................... 95
3.1 The Task ......................................................................................... 96
3.2 Neural Network Model of the Dot Probe task .................................. 96
   3.2.1 Mechanisms ............................................................................... 97
      3.2.1.1 Baseline condition ................................................................. 99
      3.2.1.2 Exposure mechanism .......................................................... 99
      3.2.1.3 Interaction mechanism ..................................................... 100
      3.2.1.4 Intensity condition ............................................................ 101
   3.2.2 Simulations ............................................................................... 102
   3.2.3 Structure ................................................................................... 102
   3.2.4 Net Input .................................................................................. 103
   3.2.5 Activation ............................................................................... 103
   3.2.6 Output ..................................................................................... 104
   3.2.7 Initialization ............................................................................ 105
   3.2.8 Training .................................................................................. 105
   3.2.9 Testing ................................................................................... 105
3.3 Bayesian Network ........................................................................... 105

4 RESULTS ............................................................................................. 110
4.1 Neural Network Model .................................................................. 110
   4.1.1 Results for Simulation 1: RT .................................................... 113
   4.1.2 Results for Simulation 2: Activation .......................................... 117
4.2 Bayesian Network .......................................................................... 119

5 DISCUSSION ....................................................................................... 124
5.1 Neural Network Model .................................................................. 124
   5.1.1 Weights of the Network ............................................................. 125
   5.1.2 Performance of the Training Mechanisms ................................... 126
5.2 Bayesian Network .......................................................................... 129
   5.2.1 The Conditional Probability Tables .......................................... 129
   5.2.2 Interpretation of Probability Values .......................................... 132
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Input patterns and corresponding outputs used for training the network by Cohen et al. (1990)</td>
<td>64</td>
</tr>
<tr>
<td>2.2 Training patterns and number of times each condition was presented to the network to train for the emotional Stroop task.</td>
<td>71</td>
</tr>
<tr>
<td>3.1 Training patterns used to train the NN for baseline, exposure 3x and exposure 5x conditions</td>
<td>98</td>
</tr>
<tr>
<td>3.2 Training patterns for the attention mechanism for baseline, exposure 3x and exposure 5x conditions.</td>
<td>98</td>
</tr>
<tr>
<td>3.3 Training patterns used to train the network for the interaction mechanism.</td>
<td>101</td>
</tr>
<tr>
<td>4.1 Number of training iterations and MSE for each training mechanism.</td>
<td>110</td>
</tr>
<tr>
<td>4.2 Basic test patterns for the neural network model.</td>
<td>112</td>
</tr>
<tr>
<td>4.3 Test patterns for the intensity condition.</td>
<td>112</td>
</tr>
<tr>
<td>4.4 Results of Simulation 1 under conditions of high anxiety.</td>
<td>114</td>
</tr>
<tr>
<td>4.5 Results of Simulation 1 under condition of low anxiety.</td>
<td>115</td>
</tr>
<tr>
<td>4.6 Simulation 2: Output activations for high and low anxiety.</td>
<td>118</td>
</tr>
<tr>
<td>4.7 Conditional probability tables for variables Arousal Rating (AR), Anxiety (Anx), Probe Side (PS), and Reaction Time (RT).</td>
<td>120</td>
</tr>
<tr>
<td>4.8 Conditional probability tables for variable AR for the Bayesian network.</td>
<td>120</td>
</tr>
<tr>
<td>4.9 Posterior probabilities of various variables computed given the evidence (in bold) using the CPT derived from all data.</td>
<td>121</td>
</tr>
<tr>
<td>4.10 Posterior probabilities of various variables computed given the evidence (in bold) using the CPT derived from data using pictures appearing on the left only.</td>
<td>121</td>
</tr>
<tr>
<td>4.11 Posterior probabilities of various variables computed given the evidence (in bold) using the CPT derived from data using pictures appearing on the left only.</td>
<td>122</td>
</tr>
</tbody>
</table>
4.12 Prior and posterior probabilities for various prior probability values of $AR=\text{Neg}$. 122

4.13 Prior and posterior probabilities for various prior probability values of $Anx=\text{high}$. 122

4.14 Prior and posterior probabilities of $Anx=\text{high}$ for various prior probability values of $AR=\text{Neg}$ given $AD=\text{neg}$. .................................................................123

4.15 Prior and posterior probabilities of $AR=\text{Neg}$ for various prior probability values of $Anx=\text{high}$ given $AD=\text{neg}$. .................................................................123

5.1 Conditional Probability Tables for variable AR for the Bayesian network..............130

5.2 Posterior probabilities of various variables computed given the evidence (in bold) using the CPT derived from all data...................................................................................132

A-1 Weights between the input and hidden units after training for baseline mechanism137

A-2 Weights between the input and hidden units after training for exposure 3x mechanism...........................................................................................................................................137

A-3 Weights between the input and hidden units after training for exposure 5x mechanism...........................................................................................................................................137

A-4 Weights between the input and hidden units after training for interaction mechanism137

A-5 Weights layer 2 (between the hidden and the output units) after training for baseline mechanism...........................................................................................................................................138

A-6 Weights layer 2 (between the hidden and the output units) after training for exposure 3x mechanism...........................................................................................................................................138

A-7 Weights layer 2 (between the hidden and the output units) after training for exposure 5x mechanism...........................................................................................................................................138

A-8 Weights layer 2 (between the hidden and the output units) after training for interaction mechanism...........................................................................................................................................138

A-9 Biases for hidden units for all training mechanisms...............................................138

A-10 Biases for the output units for all training mechanisms. ........................................138
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Cognitive mechanisms underlying biases in initial orienting to threat in anxiety</td>
<td>19</td>
</tr>
<tr>
<td>2.2</td>
<td>Versions of the Stroop task</td>
<td>25</td>
</tr>
<tr>
<td>2.4</td>
<td>A general multi-layer backpropagation neural network</td>
<td>53</td>
</tr>
<tr>
<td>2.5</td>
<td>Details of a simple processing unit of a neural network</td>
<td>55</td>
</tr>
<tr>
<td>2.6</td>
<td>Graph of the logistic sigmoid function</td>
<td>56</td>
</tr>
<tr>
<td>2.7</td>
<td>Flow of activation (solid lines) and error (dotted lines) in a multi-layer</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>backpropagation neural network</td>
<td></td>
</tr>
<tr>
<td>2.8</td>
<td>Neural network model for simulation of the Stroop task</td>
<td>61</td>
</tr>
<tr>
<td>2.9</td>
<td>Matthews and Harley Model (a) the first two models. The dotted lines were</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>connected in Model 2 while non-existent in model 1, (b) model 3 shared the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>same connections for the 2\textsuperscript{nd} weight layer with model 1.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Connections that differ in layer 1 are shown as solid lines while those</td>
<td></td>
</tr>
<tr>
<td></td>
<td>carrying over from 1 and 2 are shown in dotted lines.</td>
<td></td>
</tr>
<tr>
<td>2.10</td>
<td>d-separation in (a)serial, (b) diverging and (c) converging connections.</td>
<td>87</td>
</tr>
<tr>
<td>2.11</td>
<td>Illustration of conditional independence relationships</td>
<td>89</td>
</tr>
<tr>
<td>2.12</td>
<td>A Sample Bayesian network to determine model the probability of the grass</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>being wet given states of cloudiness (C), Rain (R) and Sprinkler (S).</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>Neural network model for simulating dot probe task</td>
<td>104</td>
</tr>
<tr>
<td>3.2</td>
<td>Bayesian network model of the dot probe task</td>
<td>108</td>
</tr>
</tbody>
</table>
Anxiety disorders afflict roughly 19 million American adults and their treatment costs upwards of 40 billion dollars annually. Attentional bias is believed to play a critical role in the etiology and maintenance of such disorders. The dot probe paradigm is used to measure attentional bias. In order to develop better treatment protocols, it is essential to understand the mechanisms of attentional bias.

The current study attempted to simulate human performance on the dot probe task using a neural network (NN) and compare three potential mechanisms of attentional bias. The NN accurately simulated performance for one of the mechanisms called the exposure mechanism. The mechanism successfully produced an attentional bias in the network by repeatedly applying negative inputs to it under conditions of high anxiety. The other two mechanisms tested were based on the interaction hypothesis and intensity mechanism. The latter explains occurrence of attentional bias through an increased in the perceived
threat value of salient stimuli by high anxiety individuals. The model also indicated a need to create a mechanism to simulate deliberate attention.

The second part of the study consisted of building a probabilistic model of attentional bias in the dot probe task using a Bayesian network (BN) to uncover probabilistic relationships among the variables. The network was able to partially model the relationships among the variables. However, it proved to be unfit in its current form for the task; finer divisions are required to model data more accurately in the BN. On the whole, the model met with limited success but offered important insights and lessons that can be applied to building better models in the future.
CHAPTER 1
INTRODUCTION

Anxiety is typically considered a negative emotion that adversely affects the ability to attend to salient information required to complete tasks at hand (Woodman & Hardy, 2001). Distraction due to anxiety can affect the information processing system to an extent that the affected individual cannot perform tasks efficiently. Such a condition characterizes a wide spectrum of anxiety disorders. Anxiety disorders are currently the most common mental illness in the United States today. According to the latest information available on the website of the Anxiety Disorder Association of America (ADAA, 2004), over 19 million adult Americans currently suffer from some form of anxiety disorder. Over 80% of these adults are afflicted by two specific disorders, namely, Generalized Anxiety Disorder (GAD) and phobias. These two disorders are twice as likely to occur in women than men. On the whole, treatment costs for anxiety and related disorders approximate to $42 billion dollars a year. Clearly, a strong need exists to understand the nature and mechanisms of these disorders to devise better treatment protocols.

Anxiety affects an individual’s ability to attend to task relevant cues in requisite detail by diminishing attentional resources available. Levels of trait anxiety reflect the propensity of a person to experience anxiety in a wide range of contexts, whereas state anxiety refers to the susceptibility to experience higher levels of anxiety in a given situation. Typically, individuals high in trait anxiety experience higher levels of state anxiety.
Anxiety influences the direction of attentional allocation and how one processes information. Specifically, anxiety influences *attentional bias*, which is defined as a discrete shift in attention to some change in the environment that is brought about either voluntarily or involuntarily (though typically the former) (Williams, Watts, MacLeod & Matthews, 1997). According to some cognitive models, an attentional bias towards threat related information plays an important role in etiology and maintenance of anxiety disorders.

### 1.1 Cognitive Models of Anxiety

Two such cognitive models are Beck’s schema theory (Beck, 1976; Beck, Emery & Greenberg, 1986; Beck et al. 1979) and associative network model of Bower (Bower, 1981). Both were among the most popular models until the middle of the 1980’s. Beck’s schema theory proposed that all information was processed according to a set schema. In *anxiety disorder*, the schema related to the processing of negative information became dysfunctional, thereby leading to selective processing of only negative information. This formed a cycle, with attention to *(negative)* threat related information strengthening the schema, thereby rendering the individual unable to avoid attending to such information.

Bower (1981) explained the same using an associative network. He posited that information was stored in an associative network, with memories of events linked to emotions they evoke and vice versa. According to Bower, anxiety caused an attentional bias towards threat. That is, nodes linked to threat information were activated more strongly and more frequently than others, leading to strengthening of the connections between nodes representing that class of information. The increase in connection strength resulted in even small activation of a node having a large overall effect. In other words,
cues having a higher level of perceived threat demanded a larger share of attentional resources.

Both models suggested attentional bias as the primary driving force responsible for causing and maintaining anxiety disorders. Both models however, incorrectly predicted an attentional bias toward information concerning loss or failure among depressed individuals. Williams, Watts, MacLeod and Matthews (1987) attempted to rectify this erroneous model by suggesting that anxiety is linked to an attentional bias towards threat information while depression is similarly biased towards recall of information related to loss or failure and as such is unaffected by the level of anxiety of the individual. They claimed that attentional bias was a product of the level of trait anxiety and the perceived threat value of the information. High trait anxiety individuals orient toward the threat stimulus while low trait anxiety individuals attend away from it.

The main emphasis of research to date has been to understand causes and mechanisms of attentional bias. Causes include such attributes as the situation in which bias occurs, the threat value of the stimulus, trait and state anxiety levels of the person. Investigations into the causes are typically carried out by experimentation, using paradigms such as the dichotic listening paradigm, the Stroop task, the dot probe and the visual search paradigm to observe and understand the nature of attentional bias in various situations. Mechanisms, on the other hand, refer to how the bias occurs and the location in the information processing system on which it acts (Williams et al., 1997). Typically, mechanisms are determined by formulating theories and models of attentional bias and verifying their correctness.
1.2 Computer Tools to Investigate Mechanisms of Attentional Bias

1.2.1 Neural Networks

A promising method that has been used to better understand the mechanisms is by modeling tasks that measure attention. A popular tool for constructing such models is a neural network. A neural network (NN) (or Parallel Distributed Processing (PDP) network) is a modeling method conceptually based on the functioning of the brain. NN models attempt to computationally mimic the massive parallelism inherent in the structure of the brain. A NN model essentially consists of a number of processing units connected with each other with either excitatory or inhibitory connections, thereby either increasing or decreasing activation levels of other processing units with which they articulate. The processing units themselves are constant, in that each unit performs the same computation. The critical element influencing the output of the network is the input to each unit. The input to a unit depends upon the weights on the connections between the unit in question and other units. As such, changes made to weights in the network affect the input to each unit, and ultimately the output of the network. Learning in the context of NN consists of determining the correct weight for each of the connections to accurately model the observed data from the problem domain.

Learning in NN is essentially dichotomized into *supervised* and *unsupervised* learning. The model used in the present study follows supervised learning. Specifically, the network has to be “trained” with known data. Training in this context consists of supplying input with known output to the network. Weights on the connections between units are then adjusted until the network yields the correct output for the given input pattern. This cycle is repeated on a large number of input patterns, until the network
produces the correct output to all training patterns. Details of this process are provided in the Chapter 2.

One main advantage of using NNs to model cognitive phenomena is that they offer a precise, computational account of the observed phenomenon consistent with the parallelism in the brain. Another advantage is in the ability of NN to handle unknown data and situations using data from known situations and data, thus allowing statistical regularities to emerge without requiring explicit coding (Matthews & Harley, 1996).

1.2.2 Bayesian Networks

A second method of modeling that has been gaining popularity in artificial intelligence and other industrial and statistical modeling settings but has not been used in anxiety research, is constructing probabilistic models of the variables involved in a task. The present study marks the first attempt to develop a probabilistic model of attentional bias using Bayesian networks (BN). BN allow intuitive probabilistic modeling of problem domains in which the relationships between variables are clearly understood. Such models rely on the laws of probability coupled with subjective probabilistic relationships to perform probabilistic inference, from a quantitative standpoint. Although the study of attentional bias has matured to the point where these relationships have been empirically delineated, no attempt has been made to quantify these relationships.

1.3 Previous Research

The Stroop task (Stroop, 1935) has traditionally been the most popular choice for studies investigating attentional bias. Words are presented in different colored inks, with the participant being required to perform one of two tasks, color naming (naming the color of the ink) or word reading (reading the word out loud). Typically, color naming is slower than word reading in conflicting conditions (i.e., when the word represents the
name of a color different than the ink (Cohen, Dunbar & McClelland, 1990)). A generally accepted explanation is that because word reading is a more automatic task than color naming, the content of the word interferes with processing information regarding color, and so causes the delay (Dyer, 1973; Glaser & Glaser, 1982). This robust finding has become commonly known as Stroop interference.

The emotional version of the Stroop task involves presentation of an emotional word instead of neutral words or color names. Color naming is slower for emotional words than for neutral words in this version. Interference in the modified Stroop task is explained as occurring due to the amount of effort required to shut-out the content of the word, leaving fewer processing resources to perform color naming (Mogg & Bradley, 1998a). Otherwise stated, emotional words arguably carry greater information load than neutrally valenced words, thereby yielding higher reaction times.

One shortcoming of the Stroop is in interpreting the results; it is not clear whether the response latency is due to interference by the word content or diversion of attention from the word. This shortcoming of the Stroop precludes the ability to locate where attentional biases occur over the course of information processing. In the absence of a valid alternative, the Stroop remained the mainstay of attentional bias researchers for over five decades.

Recognizing the significant limitations of the emotional Stroop task, MacLeod, Matthews, and Tata (1986) developed an attractive alternative to the Stroop that removed many of its shortcomings: The dot probe task. The dot probe task evaluates attentional draining rather than interference to measure attentional bias. The basic task consists of displaying a pair of emotional stimuli (words or images) simultaneously for a fixed
duration. A dot (probe) appears at the spatial location of one of the stimuli following stimulus-offset. Participants are instructed to respond as quickly as possible to probe onset by pressing a button. The task was developed based on the hypothesis that high anxiety individuals oriented toward threat stimuli while low anxiety individuals divert attention from the same. This hypothesis was supported by shorter response latencies to dot probes appearing in place of the threat stimuli for the high anxiety group, and neutral stimuli in the low anxiety group.

The dot probe has been used to uncover bias in various disorders, including eating disorders, drug addiction, smoking, and trauma to name just a few by presenting cues related to the respective disorders to patients suffering from those. The task was the first to explain attentional bias without the confounding effects of interference, and has been very influential in the formulation of the model of attention by Williams et al. (1987) mentioned above. Further, the task presented a much clearer picture of the relationship between trait anxiety and attentional bias, though lacking clear explanations of different roles of state and trait anxiety.

The NN model of the classic Stroop task by Cohen et al. (1990) marked the first model of an attentional task using NN. With this simulation, they were able to replicate the major findings of the Stroop task better than any other existing model. Up until that time, interpreting Stroop results was marred by its inherent shortcomings As a result all explanations of the results were open to discussion and debate. The model explained the results on the basis of the training utilized to get the network to produce the desired results, using a concept called strength of processing (SOP). SOP refers to higher activation levels for the units for particular inputs, which occurred due either of two
reasons; (1) an increase in strength of connections between the processing units, and (2), higher resting activation levels of some input units. In other words, observed interference was explained building on the mechanisms used to produce the same interference in the network.

Essentially, the model put forward three different possible mechanisms of attentional bias which the authors refer to as (1) exposure (involving repeated exposure to the stimuli), (2) intensity (involving superactivating certain input nodes) and (3) attentional (involving implementing a specialized unit that simulated monitoring for threat and therefore influenced the activation levels of the other units) mechanisms. The structure of the network consisted of two distinct pathways for the two tasks (explained in greater detail in Chapter 2). As such, it eliminated the possibility of ascribing Stroop effects to response interference, thereby providing a direct insight into the mechanisms without the confounds that mar interpreting results of the Stroop task.

Matthews and Harley (1994) attempted to replicate the success of the Cohen model by building a model of the emotional Stroop task. There were two main differences between their model and that of the classic Stroop (Cohen et al., 1990); firstly, the Cohen model simulated the time course of the psychological process (Cohen et al. derived a relationship between the number of repetitions required to compute the output and the RT typical for the condition being simulated, and then presented results in terms of the computed RT) while the Matthews and Harley model did not. The second difference was that Matthews and his Harley sought to investigate specific mechanisms that had emerged from the first simulation rather than build the simulation to let the relationships emerge. However, the authors later acknowledged that choosing not to simulate the time course in
the model limited its effectiveness. Still, the model was able to successfully simulate Stroop interference for each mechanism.

Despite its vast applications and popularity in studies of attentive disorders, no such model of the dot probe was developed until the current study. Although the dot probe has been used extensively to study attentional bias, it remains a method primarily to investigate the causes of bias and not the mechanisms that underlie these biases. Although the collective empirical results are relatively coherent and consistent, conflicting results have emerged. For example, some studies of social anxiety have uncovered a bias towards socially threatening stimuli while others have not, leaving confusion about the mechanisms of this particular bias. Our hope is that modeling the dot probe task will yield potential answers to the questions while generating definite considerations for future research in this area.

1.4 Limitations

Studies with the dot probe have yielded a vast database on various aspects of attentional biases. However, conclusions regarding the underlying mechanisms that drive these biases have been inferred from the data rather than directly investigated using models. As a result, although research with the dot probe paradigm has yielded robust relationships between the variables that comprise the task, these relationships have not been clearly quantified for any of the samples tested. Consider the relatively well-established relationship between trait anxiety, and valence and arousal level of the stimuli used in the dot probe task. Specifically, individuals with high levels of trait anxiety react faster to probes replacing the negatively valenced stimuli. However, the specific values of trait anxiety, picture valence and arousal rating are unknown. These and other variables of the dot probe task (e.g., stimulus duration, probe type, etc) cannot be easily quantified
by variations in task methodology, but potential answers can be generated by manipulating the variables of the task in a computer simulation.

1.5 Statement of Problem

Prior to the current study, research to study the mechanisms of attentional bias largely stemmed from studies of causes of attentional biases using paradigms such as the dot probe task. Little work had been done to directly investigate the mechanisms by developing models that mimicked human behavior when performing the dot probe task. Although studies using different modifications of the dot probe and similar paradigms yielded a huge amount of data, and significantly advanced understanding of attentional bias, a model was needed that could be updated and verified (or challenged) on the basis of new data. Further, although empirical work in this area is substantial, only a limited amount of data is collected from these studies, thereby constraining the analyses that can be performed.

1.6 Statement of Purpose

The purpose of this study was twofold; first, I constructed a NN model to simulate performance on the dot probe task so as to investigate the underlying mechanisms of the task. Second, a probabilistic model of the paradigm was constructed using Bayesian networks to develop probabilistic relationships between the variables of the model. The Bayesian model could also be viewed as a causal model and used to investigate the causal impact of one variable on the others.

1.7 Current Study

No models had been developed to simulate human performance on the dot probe task despite the multitude of studies performed to investigate the causes of attentional bias using variations of the task. As such, the purpose of the investigation was to model
the dot probe task and its findings using a NN. The study further developed a
probabilistic model (i.e., a BN) to define discrete probabilistic relationships between the
variables of the task.

1.7.1 Objectives of the Neural Network Model

Multiple simulations with the NN model will be performed to investigate issues
related to the dot probe similar to the earlier works using the Stroop task (Cohen et al
1990; Matthews & Harley, 1994).

Matthews and Harley (1996) investigated three potential causal mechanisms of
attentional bias: exposure, intensity and threat monitoring (explained above). Simulations
in the current study investigated the first two of those causes vis-à-vis the dot probe task
and an additional mechanism consistent with the interaction hypothesis. Consistent with
the hypotheses of Matthews and Harley, the current study proposed that repeated
exposure to threat stimuli lead to an attentional bias towards such stimuli for the exposure
condition. The intensity mechanism posited that individuals assigned a higher negative
valence to a stimulus owing to higher levels of trait and state anxiety. The simulation
attempted to model this mechanism in a NN. Simulation 1 was identical to the preceding
simulation except that results will be presented in terms of RT rather than activation
levels and error of the network. In doing so, I hoped to overcome a significant limitation
of the Matthews and Harley model. Specifically, they did not simulate the timecourse of
psychological process. As such, the current study represents the first work to attempt to
do so. The first simulation was aimed at replicating the main empirical findings of the dot
probe task for each of the three mechanisms. Results were to be presented in terms of RT
in ms computed from equations derived from the number of iterations required to
produce the output, and the typical RT for the condition. The main assumption in this
case was that a linear relationship exists between the number of iterations required to produce the output and the typical RT specific to each condition. As such, we made a similar assumption to that of Cohen et al. (1990).

1.7.2 Objectives of the Bayesian Network Model

The Bayesian network (BN) model served as a “proof of concept” of the advantages of probabilistic modeling in analyzing data. The Bayesian model of the dot probe consisted of the following variables:

1. **Anxiety level** of the individual, entered as normalized scores on the STAI.

2. The **arousal rating of the negative stimulus**. This parameter reflected the arousal value of the emotional stimulus presented to the individual. After the parameters were set, the arousal value was computed using the probabilistic relationships established for different values of the other variables.

3. The **side on which the dot probe appeared**, specified simply as “same” or “opposite” for dots replacing the negative and neutral stimuli, respectively.

4. The **direction of attention**, specified as either toward the negative stimulus or away from it.

5. The **reaction time**, classified as fast and slow.

The model was used to perform inference, that is, find the probabilities of one or more of the variables being in a given state, given the knowledge of the states of all or some of the remaining variables.

1.8 Hypotheses

A hypothesis was associated with each of the two models to be developed in the study:

Empirical findings from existing of the dot probe studies can be simulated using a NN. Furthermore, a relationship does exist between the number of iterations required by the network to compute the output and the RT for the particular condition of the dot probe task. Finally, the NN will be able to correctly simulate the various mechanisms of attentional bias.
6. Quantifiable and discrete probabilistic relationships exist among the variables in the dot probe task that can be uncovered in the BN model.
CHAPTER 2
REVIE W OF LITERATURE

According to information on the website of the Anxiety Disorder Association of America (ADAA [ADAA, 2004]), anxiety disorders (Generalized Anxiety Disorder (GAD), Posttraumatic Stress Disorder (PTSD), Panic Disorder, Obsessive Compulsive Disorder (OCD), Social Anxiety Disorder (SAD), and specific phobia affects) are the most common mental illnesses in the United States, affecting 19.1 million adults (ages 18-54). Treatment of these disorders costs the U.S. more than $42 billion annually, twice the amount spent on treatments for non-anxiety related disorders, including physical illnesses (Simon, Ormel, Von Korff, & Barlow, 1995). Indeed, prescription drugs for treatment of these illnesses are among the most commonly used in the world (Barlow, 2000). Of the various anxiety disorders, GAD, SAD and specific phobia affects together afflict about 15.3 million individuals, or 10.9% of adult Americans. Further, GAD and phobia affects are twice as likely to afflict women than men. People suffering from some form of anxiety disorder are six times more likely to be hospitalized for a psychiatric treatment than non-sufferers. Clearly, an understanding of the causes and mechanisms of how anxiety can lead to anxiety disorders is required so as to devise better treatment protocols.

For ages, philosophers have linked anxiety to the very essence of being human (see Barlow, 2000), which is quite an ironic observation given that humans spend billions of dollars every year to rid themselves of the same. Anxiety and anxiety disorders have been
observed in various cultures, from the Eskimo hunters of Greenland in the early part of
the last century, who experienced a sudden and extreme panic attack on their hunting
trips (Danish travelers to the region recorded it as “kayak angst”) and not be able to
venture far out of the village again to a “sore neck” that affected the Khymer refugees in
more recent times. Another consistency of these disorders is their prevalence among
women. According to a WHO report, the odds of women being affected by some form of
anxiety disorders are 1.63 (95% confidence interval) (Barlow, 2000).

The aim of the current study is to develop a Neural network (NN) and Bayesian
network (BN) model of the dot probe task to investigate the underlying mechanisms. This
chapter provides a review of literature related to attention, attentional biases, and PDP
models developed to investigate the mechanisms of attention. The chapter reviews the
two PDP models of an attention task developed so far, specifically a model of the classic
Stroop task (Cohen, Dunbar, & McClelland, 1990) and a model of investigating the
mechanisms of the emotional Stroop task (Matthews & Harley, 1996). Literature related
to the dot probe paradigm is covered in significant detail to highlight methodology,
variables, and salient characteristics of the task.

2.1 Anxiety

Humans have a limited attentional capacity; implying that in order to efficiently
perform multiple tasks simultaneously, it is crucial to identify the task relevant cues and
the level of detail to which each cue must be processed. The information processing
system then must process these relevant cues in requisite detail and ignore the others; a
job delegated to the attentional system. As a result, the system plays a key role in survival

1 Panic attack experienced by Eskimo seal-hunters while hunting alone for days in their kayak. After the
attack, the afflicted hunter could not venture but a few miles out of the village.
and evolution (Mogg & Bradley, 1998a). However, the amount of attentional capacity available is affected by the emotional state of the person.

Research over the last two decades has led to several robust findings linking anxiety to performance decrement on a broad range of tasks. Woodman and Hardy (2001) refer to anxiety being generally accepted to be an unpleasant emotion. Lang (2000) describes human emotions to have developed around two key motivational systems that play key roles in evolution and survival, namely the appetitive and defensive systems.

Emotions in general and anxiety in particular influence the selective attention and information processing capability of an individual. The general propensity of an individual to experience high anxiety and the short-term anxiety he or she experiences in a particular situation are distinguished as trait anxiety and state anxiety respectively. However, trait anxiety does have a bearing on the level of state anxiety experienced by an individual; typically, individuals with trait anxiety have a tendency to experience higher levels of trait anxiety in stressful situations when compared to low trait anxious individuals (Williams, Watts, MacLeod & Matthews, 1988).

2.1.1 Measuring Anxiety

Anxiety is mainly measured through paper-and-pencil self-report questionnaires. One of the most frequently used scales to measure levels of state and trait anxiety is the Spielberger’s State and Trait Anxiety Inventory (STAI) (Spielberger, Gorsuch., & Lushene, 1970). The STAI has both a state version (STAI-S) and a trait version (STAI-T); the state version is the most commonly used inventory to measure state anxiety. Both versions consist of 20 questions; the STAI-T consists requires individuals to rate how they generally feel on a 4-point frequency scale (from 1 = almost never to 4 = almost always) while its state counterpart asks them to rate their feelings at that moment on a 4-
point intensity scale (from 1 = not at all to 4 = very much so). Maximum possible score on both versions is 40 while the minimum possible is 10. The test has been reported to have high internal consistency (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983).

Other popular inventories include the Autonomic Perception Questionnaire (APQ) (Mandler, Mandler & Urviller, 1958), the Affective Adjective Checklist (AAACL) (Zuckerman, 1960), and the Activation-De-activation Checklist (AD-CL) (Thayer, 1967). One drawback common to all self-report methods is the inability of the participants to accurately and reliably report on their cognitive processes (Nisbett & Wilson, 1977).

2.1.2 Cognitive Models of Anxiety

A number of cognitive models of attention and attentional biases have been put forward to explain the relationship between anxiety and attention. One key feature emerging from these models is that attentional bias is critical in the origin and maintenance of anxiety and emotional disorders like GAD. Up until the latter part of the 1980’s, there were two main theories explaining how anxiety affects attention, Beck’s cognitive theories of emotional disorders (Beck, 1976; Beck, Rush, Shaw, & Emery, 1979; Beck, Emery & Greenberg, 1986) and Bower’s theories based on his network model (1981). Beck’s theories, in particular, have been influential in devising new treatment protocols for depression and anxiety, specifically cognitive-behavioral therapy (Butler, Fennell, Robson, & Gelder, 1991; Simons, Murphy, Levine & Wetzel, 1986).

Beck (1976) proposed that humans have set schema that they use to process the information. All incoming information passes through the schema, which attaches semantic meaning to it. In people suffering from anxiety disorders, the schemata pertaining to processing threat or danger are dysfunctional, resulting in selective
processing of schema-congruent information when activated. Similarly, depression is
associated with dysfunctional schemata related to loss or failure.

Bower (1981) explained the same phenomenon using an associative network of
emotions and memories of events. In such a network, emotions are connected to
memories of relevant events (happy and sad) to form nodes of the network. Activation
can travel in either direction; activating a particular emotion node can trigger specific
memories and vice versa. When activated, a node also activates, to some extent, the
nodes connected to it. For instance, normally the node representing “sadness” is linked to
nodes representing memories of sad events. Feeling “sad” will trigger memories of sad
events and thinking of these events will activate sorrow. This means that events are
tagged with their emotional value before being stored in the network. In depression,
events are tagged as negative more frequently and as having higher intensity of
negativity. This increase in strength of connection between the memory nodes and the
incoming information nodes implies that even events with low values of sadness have a
higher negative impact on depressed individuals, as compared to normal individuals and
also strengthen the connections more. Similarly, attentional bias towards anxiety causes a
tendency towards selective processing of negative or threatening information, causing the
individual to experience higher levels of anxiety.

Both models explain the role of anxiety in causing and maintaining attentional
biases. Overwhelming evidence exists in support of most of the predictions of both the
above models (e.g., Clark & Teasdale, 1982; Bradley & Matthews, 1983; MacLeod et al.,
1986). However, the models fail to explain some of the findings emerging from the
studies on attentional biases. In particular, both models predict that both anxiety and
depression should be associated with attentional biases on all aspects on information processing, namely selective attention, reasoning, and memory (Mogg & Bradley, 1998a). However, studies have failed to uncover any evidence either of an attentional bias towards threatening stimuli in depression (MacLeod et al., 1986) or of a recall (memory) bias in anxiety (Mogg, Matthews & Weinman, 1987). On the contrary, research suggests anxiety is linked to an attentional bias towards threat while depression is associated with a memory bias towards negative information (Mogg & Bradley, 1998a).

Figure 2.1 Cognitive mechanisms underlying biases in initial orienting to threat in anxiety

To explain these findings, Williams et al. (1988) proposed a new model relating attentional biases to anxiety and depression. To begin with, they associated anxiety with a tendency for preattentive vigilance for threat and depression with a bias towards postattentive elaborative processes, thereby explaining the lack of a recall bias is anxiety and a similar lack of bias in preattentive processes in depression. The model proposes two mechanisms for directing preattentive and attentional bias towards threat stimuli in
high anxiety individuals; the Affective Decision Mechanism (ADM) evaluates the threat value of incoming stimuli and outputs the result into the Resource Allocation Mechanism (RAM) (Figure 2-1). The RAM allocates attentional resources towards or away from threat based on trait anxiety of the individual with high trait anxious individuals having a tendency to orient towards threat and low trait anxious people orienting away from the same; this is the interaction hypothesis. As indicated in Figure 2.1, the difference between high and low anxious individuals becomes more pronounced with an increase in the threat-value attached to the stimulus by the ADM. Consequently, the ADM can be thought of as a mechanism to assign priorities to incoming stimuli. Three main themes of the model are:

1. Anxiety is associated with different patterns of selective attention rather than with detailed information processing. Cognitive bias for anxiety acts on the preattentive stage, looking for threatening stimuli in the environment.

2. Individuals who have a tendency to display an attentional bias towards threatening stimuli are more prone to anxiety and anxiety disorders under stress.

3. Trait anxiety influences the direction of the attentional bias, implying individuals with high trait anxiety are more likely to experience a higher level of state anxiety in more situations.

Williams, Watts, MacLeod and Matthews (1997) revised their 1988 model within a connectionist framework Parallel Distributed Processing model of Cohen et al. (1990). The revised model is explained following the review of PDP models of the Stroop task below.

Recent theories of anxiety appear to complement the Williams et al. (1988) model. For example, Matthews (1990) proposed that emotions serve to assign processing priorities to incoming stimuli based on the view of evolutionary functions of emotion (Oatley & Johnson-Laird, 1987). Eysenck (1992) devised the hypervigilance theory on
the basis of the interaction hypothesis. He proposed that not only are high trait anxious people more biased towards attending to threat information (i.e., high specific hypervigilance) but they may attend to any task irrelevant stimuli when anxious (i.e., high general hypervigilance or distractibility). The theory further suggests that high trait anxiety results in a higher rate of environmental scanning, with the general focus of attention being very broad but becoming narrow when focusing on cue relevant stimuli.

Eysenck and Calvo (1992) explained the effect of anxiety on selective attention and subsequently on performance with the Processing Efficiency Theory (PET). According to the theory, processing of anxiety-relevant stimuli increase demands on working memory, thereby reducing the amount of resources available to process task relevant information. Another possibility emerging from the PET is that performance decrement exhibited by high trait anxious individuals on experimental tasks in individuals occurs because they selectively attend to stimuli that are relevant to their anxiety and not to the task at hand.

The PET is a formalization of the line of cognitive research being followed in the study of anxiety. Research emphasis has been to focus on the patterns of allocation of selective attention to tasks associated with high anxiety (Matthews & McLeod, 1994). The susceptibility of high anxious individuals to attentional bias towards processing information relevant to their anxiety is a robust finding and has been repeated in various situations with a host of different populations.

The role of attentional bias in selective attention has been explained using a searchlight analogy (Williams et al., 1997): Selective attention is likened to a searchlight beam, with the area illuminated by the beam as the center of attention. Some peripheral
attention is devoted to information in the area not illuminated directly by the searchlight. This information can cause an involuntary shift in attention causing attentional draining from the main task and thereby leading to deterioration in performance. Attentional bias is explained as an attentional vigilance to threat and has been projected as the main factor in causing and maintaining anxiety (Matthews, 1990; Eysenck, 1992). Bradley, Mogg, Falla, & Hamilton (1998, p. 737) explain this cycle: “Individuals with a tendency to adopt such a vigilant attentional style would be more likely to detect potential sources of danger in their environment, which in turn would exacerbate their anxious mood.”

Öhman (1993; Öhman & Soares, 1993, 1994) reached similar conclusions from a different research perspective. They suggested that preattentive processes also regulate vulnerability to phobias with fear evoking stimuli working in much the same way as threat stimuli in anxiety. Specifically, fear responses to stimuli are initiated by automatic analysis mechanisms. These mechanisms are guided by biologically prepared threat stimuli and direct attention to the stimulus once it is analyzed.

### 2.2 Attentional Bias

The above discussion reveals the importance of attentional bias in evaluating incoming stimuli and directing attention. Williams et al. (1997) assume an attentional bias to have occurred when there is a discrete shift in attention to some change in the environment of the individual. They specify three assumptions regarding the shift in attention resulting from the bias essential in studying attentional biases:

4. The shift encompasses all sense modalities (vision, touch, taste, smell, etc)

5. Although usually passive and involuntary, the shift can be voluntary (i.e., attention can be deliberately focused on the area).

6. Onset of the shift is brought about by a discrete change in the environment (i.e. by the onset or offset of some event).
Researchers have used two basic paradigms to study attentional biases, interference methodologies (specifically, the dichotic listening paradigm and its visual analog, the Stroop task), and methods to test directly for attentional bias (namely, the visual search and dot probe paradigms).

2.2.1 Dichotic Listening Paradigm

Mainly used in studying selective attention, the dichotic listening paradigm and its visual analog are now the least preferred paradigms for studying attentional biases. The basic version of the task consists of simultaneously playing a different audio message in the left and right ear of the subject using headphones. Participants are then asked to “shadow” (say out aloud) one of the messages as it is played. Early studies found participants could effectively follow only one of the messages, though some information (e.g. a high pitch tone, or change from a male to female voice) from the unshadowed message still got through. The explanation offered was that the messages were distinguished on the basis of some physical characteristics (e.g., pitch, amplitude, etc.). For a full review of theories of selective attention see (Abernethy, 2001). The paradigm was instrumental in establishing that attentional bias acts early in the information processing system.

In the study of attentional bias using dichotic listening paradigm, the task was based on the premise that anxious individuals were more likely to attend to threat and other stimuli that directly related to their life events. Parkinson and Rachman (1981) used the task to study lowered auditory thresholds in concerned mothers. They presented audio messages consisting of words representing pain and other unpleasant stimuli embedded at various volumes to two groups of mothers; those whom had children admitted to the hospital for some surgical procedure and a control group with no children admitted.
Results indicated the experimental group identified more embedded words than the control group.

The paradigm lost stature following a study on PTSD sufferers (Trandel & McNally, 1987), in which it failed to find any attentional bias in war veterans towards Vietnam-related words. Participants with and without PTSD experienced similar disruptions to all threatening stimuli; this was a major shortcoming since it does not allow the method to be made reliably sensitive to the specific fears of the population being studied.

2.2.2 Stroop Task

The classic Stroop task (Stroop, 1935) consists of displaying names of colors written in different colored inks (Figure 2.2(a)) in which participants are required to either read the word or name the color of the ink as quickly as possible. The dependent measure in this case is the response time to name the color or read the word. Typically, participants are able to ignore the effects of ink color while reading the word aloud but experience significant interference with the word when trying to name the color of the ink. Interference is greatest if the word is an antagonistic color-name (e.g., word “GREEN” printed in red ink (Figure 2.2(a)) or represents an antagonistic color (e.g. word “GRASS” printed in red ink) (Jensen and Rohwer, 1966, MacLeod, 1991). Meaningless stimuli (e.g., a row of X’s) do not interfere with the ink naming at all while congruent colors slightly facilitate naming the color (“RED” printed in red ink [Figure 2.2(b)]).

Various explanations have been presented for the observed discrepancy in response times for color-naming and word-reading. The simplest one explained the observed interference on the basis of discrepancy in processing times required for the word reading and color naming. Researchers proposed that color naming is more
automatic than word reading and so takes longer to process than word reading, causing interference at the output level. So, although both inputs were detected simultaneously, response from word reading arrived at the output level before its counterpart from the color-naming task. However, Glaser and Glaser (1982) put the explanation to test and proved it inadequate; they provided participants with advanced knowledge of the ink color by displaying a color patch of the same color as the color of the ink. Participants displayed interference effects even when the color patch was displayed 400 ms before the word. The effects came to be known as Stimulus Onset Asynchrony (SOA) effects.

A more robust explanation was offered by MacLeod and Dunbar (1988) on the basis of degree of automaticity of different tasks. Numerous studies have established that automaticity on a task increases by practice according to the power law (Kolers, 1976; Newell & Rosenbloom, 1981; Anderson, 1982; Logan, 1988). MacLeod and Dunbar (1988) reasoned that if amount of practice could make a task more automatic, it would show up in appropriate changes in Stroop interference with the more automatic task interfering with the performance on the less automatic one. To test the hypothesis, they trained individuals to associate four different shapes with four different colors.

Figure 2.2 Versions of the Stroop task.

<table>
<thead>
<tr>
<th>GREEN</th>
<th>RED</th>
<th>SPIDER</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Conflict condition, GREEN in red ink</td>
<td>b. Congruent condition, GREEN in red ink</td>
<td>c. Emotional Stroop for spider phobies, in red ink</td>
</tr>
</tbody>
</table>
Participants were trained by presenting the shape in a neutral color (white) with each shape being presented 72 times. The same pattern of training was carried out daily for 20 days. At the end of the first day of training, participants were administered the Stroop task with the shape. On an average, participants were 100 ms slower at shape naming than at naming colors. The end of the fifth day saw significant increases in speed of shape naming, with the shapes interfering with color naming. By the time the study was completed (20 days of 72 trials per stimulus for a total of 2,520 trials per stimulus), participants displayed significant interference with color naming and a small amount of facilitation in naming colors (evident from increased RT when the shape and color were conflicting and reduced RT when the shape and the color were congruent). On the other hand, colors showed very little effect on shape naming (shapes had taken the place of words and become the more automatic task).

Word reading is considered an automatic task because individuals have arguably practiced it more than the more controlled and less practiced task of color naming. The theory has since been modified based on the findings of MacLeod and Dunbar (1988), who proposed that tasks have varying degrees of automaticity and control. Specifically, the degree of automaticity of tasks is a continuum rather than a dichotomy (i.e., tasks are not simply controlled and automatic but vary in their degree of automaticity or control, with some being more automatic than others). For instance, if a study consists of two tasks with one being more automatic than the other, performance on the less automatic task will suffer due to interference from the more automatic process.

The emotional version of the Stroop task used a negative affective word instead of a color name. Participants had to name the color of the negative word (words
representing specific phobias for phobics). Figure 2.2(c) depicts a slide presented to a spider phobic. The emotional Stroop is an adaptation of the classic Stroop task that compares the response times of participants on color naming a series of emotional words as opposed to color-names (Matthews & Harley, 1996; Williams, Matthews & MacLeod, 1996). For anxiety, the affective word is either threatening (e.g. “death”, “injury”, “sickness”) or non-threatening word (e.g. “chair”, “picture”). Individuals with high trait anxiety were hypothesized to display higher levels of interference in naming the ink color of a threatening word rather than a non-threatening word. Studies have found results congruent with the hypothesis.

The first study using this paradigm was conducted on patients suffering from GAD (Matthews & MacLeod, 1985). They found the experimental group to be significantly slower at naming the color of threat words. Similar results have since been observed in patients suffering from a host of different emotional and anxiety disorders, including Post traumatic stress disorder (PTSD) (Threasher, Dalgleish & Yule, 1994), obsessive compulsive disorder (OCD) (Lavy, van Oppen & van den Hout, 1994), specific phobics (IAS, like social and spider phobics) (Lavy, van den Hout & Arntz, 1993) and panic disorders (McNally, Amir, Louro, Lukach, Reimann & Calamari, 1994). One key finding apparent from studies with population groups suffering from different anxiety and emotional disorders was that individuals with these disorders exhibit the greatest difference in interference with the corresponding control groups when the valence words used represent threats relevant to their specific condition. For example, in a study with GAD patients worried about physical injury, Mogg, Matthews & Weinman (1989) found that participants displayed the most interference when the word was related to physical
danger (e.g. injury, fracture, etc.). Similarly, the greatest interference effects for social phobics are induced by words representing socially threatening situations (Hope, Rapee, Heimberg, & Dombeck, 1990), and words of physical threat cause the greatest influence for panic disorder patients (McNally, Amir, Louro, Lukach, Reimann, & Calamari, 1994).

Researchers put forward different explanations to account for the findings associated with the emotional Stroop task. One view was that anxious individuals allocate more attentional resources to threat words and process them in greater detail due to an attentional bias toward threat. The increase in resources consumed by processing threat information lead to the interference. A second explanation posits that threat words cause a spike in the level of state anxiety, disrupting performance on color-naming task. Some researchers (MacLeod, 1990; de Ruiter & Brosschot, 1994) questioned the validity of both the above explanations, stating that a tendency to divert attention from emotional cues can also lead to observed interference. The latter explanation was only the first of several criticisms levied against the emotional Stroop task.

One major drawback of the emotional Stroop variations was in interpreting the results of the task; the task offered no evidence as to whether the interference occurred at the information processing stage or at the response selection stage. Further, it failed to shed any light on the role of state and trait anxiety in the observed effects. Interpretative difficulties apart, the paradigm also lead to some unexpected results. Specifically, studies with phobics did not find any threat-related interference when participants were in close proximity (physical or chronological) with their threat situation or object. For example, snake phobics in the presence of snakes (Matthews & Sebastian, 1993) and social phobics
getting ready to give a speech right after testing (Amir, McNally, Reimann, & Clements, 1996) did not reveal expected interference effects for snake and threatening social situation related words. Finally, the task measured only deterioration in performance due to attentional bias (Williams, et al., 1997). However, interpretative difficulties remain, by far, the more serious shortcoming of the task.

2.2.3 Dot Probe Task

Interference paradigms like the dichotic listening paradigm and the Stroop task failed to offer a direct indices of the mechanisms underlying attentional bias, as explained by the interpretative difficulties encountered in the Stroop task. Alternatives to these tasks are the visual search paradigm and the dot probe paradigm. The dot probe is a direct measure of attentional bias experienced by individuals (Williams, et al., 1997). Developed by MacLeod, Matthews and Tata (1986), it was modified from paradigms in cognitive psychology that used response time to visual probes to assess attention (Posner, Snyder & Davidson, 1980; Navon & Margalit, 1983). These paradigms suggested that participants would respond faster to a probe stimulus when it appeared in an attended rather than unattended region of visual attention.

The paradigm measures attentional drain due to existing biases in attention without confounds of response selection by measuring the reaction time of a neutral response (button click) to a neutral stimulus (dot-probe). The basic steps (Figure 2.3) consist of simultaneously displaying an emotional cue (word or picture) paired with a neutral cue for a short duration of time (traditionally 500 ms though other times have been used). Following cue offset, a dot appears in the spatial location of one of the two cues. Participants are instructed either to indicate the position of the probe (probe position task, i.e., indicate whether the probe appears on the left or right, or top or bottom by pressing
the appropriate button) or indicate its type (probe classification task; two different probes are used, say the letters ‘E’ and ‘F’, and participants are required to indicate the letter) as quickly as possible. The driving hypothesis for the MacLeod task (another name for the dot probe used in literature) was that high anxiety individuals systematically attend to threat-related stimuli and this would be reflected by faster response times to probes replacing emotional cues as opposed to non-threatening cues and also response times of non-anxious individuals for the same cues.

2.2.3.1 Initial studies (basic dot probe task)

The first dot probe study (MacLeod, et al., 1986) used the paradigm to measure attentional bias in GAD patients. Sixteen individuals diagnosed with GAD and referred for anxiety management by their practitioner were tested against a group of sixteen low anxiety (LA) controls. The GAD group obtained mean scores of 44.7 and 52.5 on the state and trait versions of the STAI while the LA group scored 36.3 and 39.5, respectively. On the Beck Depression Index (BDI), GAD sufferers and controls groups
scored 13.9 and 7.6, respectively, with GAD patients being significantly more depressed than their LA counterparts (the significant difference in the depression levels of the two groups added a confound that was later removed by testing a low-anxious depressed sample against controls). Patients were matched with controls for age, gender and verbal intelligence (measured by the Mill Hill Synonyms Test). Each group was shown a total of 288 word-pairs on a computer monitor; 48 consisted of a threat word paired with a neutral word while the remaining 240 were a pair of neutral words. Of the threat words, half represented physical threat while the other half represented social-threat.

Words were displayed centered on the vertical axis of a VDU (Visual Display Unit), separated by a distance of 3 cm from each other (constituting a visual angle of less than 2 degrees), for 500 ms. Participants were instructed to read out loud the word appearing on the top in every trial and to press a button as quickly as possible when a probe appeared to indicate its presence. The probe (a white dot that appeared with equal probability in the spatial location of one of the two words) appeared in 96 trials and remained on the screen until participants pressed a button to indicate its presence. All threat-neutral pairs (48 in all) were followed by the probe while the other 48 probed trials consisted of filler items chosen at random from the neutral pairs. Trials could thus be classified into three types; probed-threat, probed-neutral, and unprobed-neutrals. On trials without the probes, the next picture was displayed following a delay of one second.

Results confirmed the hypothesized preferential attention to threat information by the GAD group and an avoidance of the same displayed by controls. When the probe appeared at the top, the GAD group was significantly faster at responding when it followed a threat word (593 ms) than a neutral word (652 ms). The same trend was
observed for probes appearing in the lower section of the screen, with the high anxiety group responding faster when the probe was preceded by a threat word (663 ms) than when it followed a neutral word (695 ms). Reaction times for the control group followed the reverse trend, with controls reacting faster to probes replacing neutral words than threat words, implying an avoidance of threat cues. Specifically, controls recorded reaction times of 540 ms when the probe replaced a threat word in the upper area vs. 524 ms when it replaced a neutral word in the same location; for probes appearing in the lower area of the display, controls were 32 ms faster in responding to probes following neutral words (584 ms) as compared to threat words (616 ms).

The study was among the first to offer an explanation for attentional bias towards threat stimuli without any confound from response bias (as would have been the case if Stroop task had been used) of the results. Results from this and other dot probe studies were critical in formulation of important assumptions regarding the nature of attentional bias. Williams et al. (1997) acknowledged the contribution of the paradigm as,

It [the MacLeod et al. (1986) study] showed that we needed to assume the existence of a decision mechanism which (a) was at a preattentive level, (b) was sensitive to general differences in threat, (c) allocated attention to different parts or aspects of the environment, and (d) was independent of response bias (Williams, et al. 1997, p. 83)

In a subsequent study using the paradigm, Broadbent and Broadbent (1988) attempted to answer some of the questions emerging from MacLeod et al. (1986). They investigated whether preferential allocation of attention to threat stimuli was a characteristic of only clinically anxious people or if people with sub clinical levels of anxiety also display a similar bias. Further, they questioned whether the effects were a function of the personality of the individual (and therefore permanent) or more a function
of the state of the individual regardless of personality characteristics (and so more fleeting).

Making a few minor changes to the setup of MacLeod et al. (1986), they tested a total of 104 women in four different experimental setups to answer the above questions. In each experiment, they divided the women into a HA group and a LA group based on their STAI scores. Individuals scoring greater than 35 on the trait form of the STAI were classified as HA while those scoring less than that were classified as LA.

Results confirmed the existence of a similar bias in the sub-anxious sample and an avoidance of threat information by the LA group. On the whole, HA participants responded faster when probes appeared in place of threat stimuli as opposed to when the probe replaced the neutral stimulus in the threat-neutral pair while the reverse was true for LA participants. When threat words appeared in the upper area, HA individuals responded faster to probes replacing the threat word (587 ms) than to probes that replaced the neutral word in the threat-neutral pair (637 ms). Similarly, when the probe and the threat word, both appeared on the bottom, the HA group took 650 ms to press the button while taking 667 ms when the word appeared on the bottom and the probe appeared on top. Opposite readings were observed for the LA group; individuals were slower to respond to probe appearing in the location of the threat word (RT 656 ms and 678 ms for probes and threat words on the top and bottom, respectively), while reacting faster to probes that replaced the neutral word in the threat-neutral pair (649 ms for threat word in upper position and probe in lower and 657 ms for the opposite). They also found trait anxiety a more reliable indicator of attentional biases as compared to state anxiety; high trait anxious participants in their study consistently displayed similar patterns of
attentional bias while the effect of state anxiety differed from one experiment to the other.

Around the same time, MacLeod and Matthews (1988) ran a follow-up study to investigate the effects of state and trait anxiety on attentional biases. They tested 36 high and low trait anxious (non-clinical) college students for attentional biases towards exam related cues under conditions of low stress (12 weeks before the exam) and under high stress (one week before the exam). Here again the STAI (trait) score (dividing median score 39.5) was used to stratify the students into high and low anxious categories. Participants were presented 288 word pairs, 96 of which were probed. The probed pairs consisted of an equal number of threat and neutral pairs. Half the threat words used in this case were related to examinations while the other half were general threat words chosen from earlier studies. Words were rated for threat value and pertinence to examinations on a scale of 1-5 (1 being the least and 5 the maximum on both scales) by eight independent judges and both groups of threat words had the same threat rating (4.1).

Authors computed the attentional bias score to analyze the results by subtracting the mean RT when the probe occurs in the same place as the threat word from the mean RT when the probe and the threat cue occur in different locations.

\[
\text{Attentional Bias} = \frac{(UP/\ LT - UP/\ UT) + (LP/\ UT - LP/\ LT)}{2}
\]

\[UP = \text{Upper Probe, } UT = \text{Upper Threat}
\]

\[LP = \text{Lower Probe, } LT = \text{Lower Threat}
\]

Positive values of bias score signified vigilance of threat and negative values indicated avoidance of threat. Attentional bias scores were used to obtain a single index of probe and threat positions so as to simplify computing a four-way interaction of trait
anxiety, test time, threat position and probe position. Results revealed a tendency to attend to threat stimuli in general by the high trait anxious group and avoidance by the low trait anxious group. Exam words did not attract much attention from either group in the first test but did so in the second. Also both groups recorded equivalent increases in state anxiety but with opposite effects. The HA group responded even faster when the probe and threat appeared in the same location while the low anxious group recorded shorter latencies to probes replacing the neutral word in critical pairs. This pattern of change could not be explained on the basis of trait anxiety alone and lead the authors to infer that the patterns were in fact due to an interaction of state and trait anxiety. Researchers later referred to this pattern of attention allocation and the effect of state and trait anxieties on it as the interaction hypothesis (Williams et al., 1988).

The dot probe has proved to be an effective measure of attention allocation and preattentive bias to different kinds of stimuli. Studies have replicated the task and confirmed the existence of similar patterns of attention allocation in several different samples. Different studies effected changes in the task methodology, making the task more sensitive to the sample being studied. One shortcoming with early versions of the task was due to probing in only a portion of the trials. This limited the amount of data that could be collected. Also, because each threat word was probed, the appearance of such a pair could act to prime participants for the probe and hence result in a faster RT.

Mogg, Bradley and Williams (1995) disposed of this requirement by probing participants at the end of every trial and instructing them to indicate the position of the probe (top or bottom for cues displayed centered on the vertical axis) by pressing the appropriate button. Trials on which the threat-word, neutral word pairs were displayed
came to be known as critical trials. Using this methodology also rendered the requirement of reading aloud the top word infeasible, eliminating another confound. One potential flaw of probing in every trial was the possibility of participants adopting a strategy to attend to the spatial location of one word only (and press the button indicating presence of the probe depending on whether the probe appeared on the side they were attending to or not). MacLeod and Chong (1999) overcame this potential pitfall with the “forced reaction time” version of the task. Essentially, they used two different probes (two dots in vertical and horizontal orientation ‘:’ and ‘..’) and participants were instructed to perform probe classification, pressing a different key depending on the type of probe used.

One criticism the basic dot probe shared with the Stroop was that it too provided only a snapshot of attention at the instant of probe onset. Specifically, both tasks only provide a definite answer of the direction of attention when the dot probe appeared, but no information to the direction of attention allocation prior or subsequent to the probe. Three explanations have been offered to account for the observations of the dot probe. First is the vigilance-avoidance pattern of processing (Mogg, Matthews and Weinman, 1987; Williams et al., 1988), which says anxious individuals follow a pattern of first attending to and then avoiding the threat cue in an effort to mitigate their anxious state. Such a pattern may also act to maintain their anxiety-state by preventing anxious individuals from habituating to threatening events. A second possibility, consistent with the models of Beck (1976) and Bower (1981), is that anxious individuals orient themselves toward threat and subsequently have trouble disengaging attention. To answer these questions required modifying the task of MacLeod et al. to measure the time course of attention.
2.2.3.2 Manipulation of stimulus duration

Mogg, Bradley, de Bono and Painter (1997) modified the task to measure attention at three different times by manipulating stimulus duration. They presented 192 word pairs (96 threat-neutral and 96 neutral-neutral pairs) to 35 volunteers in top-and-bottom orientation, and displayed each word pair randomly for 100 ms, 500 ms or 1500 ms. The first condition was designed to be shorter than the inter-saccadic interval during active visual search (varies between 200-300 ms: Kowler, 1995) and therefore did not allow any shifts in attention. On the other extreme, 1500 ms allowed for detailed processing of the stimuli and multiple overt shifts in attention. The 500 ms condition represented the most frequently employed time period for the dot probe. They also varied the inter-stimulus period randomly among 750, 1000 and 1250 ms. Threat words consisted of an equal number of words referring to social threats (e.g., stupid, despised, criticism) and physical threat (e.g., illness, injury, fracture). Word-pairs in critical trials were ordered such that there was an equal probability of the type of threat-word displayed (social, physical), its location (top, bottom) and the probe position (top, bottom). Participants’ emotional states were assessed after they completed the task by having them fill out the STAI, BDI, and Social Desirability Scale, among others, and were divided into two groups (high and low state anxiety) based on their STAI-state scores (dividing median score 30) for analysis.

After removing outliers from the data (trials with high error rates and RT more than three standard deviations from the mean), Mogg et al. (1997) they found a significant main effect for exposure in that individuals tended to respond faster to probes in the 100 ms duration (485 ms) as opposed to the other two conditions (latencies of 498 ms and 503 ms for 500 ms and 1500 ms conditions) regardless of trait anxiety. High state anxious individuals showed a significant trend to respond to threat, recording response times10
ms faster to probes replacing threat words (mean 493 ms) than those replacing neutral words (503 ms). On the other hand, low anxiety scorers displayed a non-significant effect of threat-avoidance, with RT for threat words 8 ms slower than for neutral words (502 ms and 494 ms, respectively). Bias scores for the high state anxiety group were 10, 9, and 11 and –11, -10 and –1 for the low state anxiety group for the 100, 500 and 1500 ms conditions respectively. Post hoc analysis revealed significant difference in bias scores averages over the three conditions for the two groups, with significant differences in the 100 ms condition and non-significant trends between the other two.

Findings did not support the vigilance-avoidance hypothesis, that is, there was no significant difference of attentional bias with display duration. According to the vigilance-avoidance hypothesis, dysphoric individuals have an initial attentional bias towards threat cues, which puts them in an aggravated state of fear. In order to escape this state, they direct attention away from the stimuli. The authors refrained from making any generalizations on the basis of this study, as it was the first in the field, and offered two explanations for the observations; first, they suggested that approach-avoidance could be more likely a characteristic of individuals with anxiety disorders (GAD, panic disorder, etc.) rather than those with sub clinical anxiety, and secondly, they suggested that attentional avoidance could be influenced by the relative threat value of the stimulus.

### 2.2.3.3 Backward masking

Backward masking of stimuli was a technique used in the Stroop task to restrict awareness and measure strictly preattentive bias towards threat cues. It involved displaying the word for a very short time (e.g., 14 ms) and then replacing it with a length-matched mask (random characters). The mask worked to prevent detailed processing of the word and thus strictly measured pre-attentional bias. The same was adapted for the
dot probe. All word pairs were displayed for a very short duration (typically 14 ms), and then covered by a mask (consisting of random letters or symbols, one for each letter of the word, or contortions of the letters of the words themselves) for a similar duration. The probe followed stimulus-offset and participants were required to perform probe position or probe classification task. The hypothesis for the masked dot probe remained virtually unchanged from the basic version; HA individuals were predicted to attend to the spatial location formerly occupied by the mask covering the threat word. Results of studies using this version of the dot probe confirmed the hypothesis with HA individuals and those with GAD (Bradley, Mogg & Lee, 1997b; Mogg et al., 1997).

A barrier in generalizing findings from the above studies and other dot probe studies using single words as threat stimuli is the amount of threat information that single words can convey. As noted by Bradley et al. (1997a) and Mogg et al. (1997), single words convey a limited amount threat-information and, once that information is extracted, the word loses part or all of its threat value. Additionally, a potential confound exists due to the threat value and relative frequency of use of threat words as HA individuals use threat words more often more than LA individuals. Finally, research suggests that attentional biases are guided by innate, evolution-driven mechanisms; words do not fulfill the criteria fit of ecologically valid threat stimuli (LeDoux, 1995). An alternative to single words is using photographs of threat stimuli (mutilated bodies, attacking animals, angry faces). Pictorial cues are a much more ecologically valid threat cue than single threat words.

### 2.2.3.4 Pictorial dot probe task

Bradley and his colleagues (Bradley, Mogg, Millar, Bonham-Carter, Fergusson, Jenkins, & Parr, 1997) designed a pictorial dot probe task, using pictures rather than
words as the emotional stimuli. They displayed pictures of faces with happy, threatening or neutral expressions for 500 ms to a sample consisting of sub-clinical HA students (grouped according to scores obtained in the upper and lower tertiles of the Fear of Negative Evaluation scale (FNE; Watson & Friend, 1969)). On critical trials, an emotional face was paired with a neutral face and participants were required to indicate the position of the dot. Results did not indicate a relationship between social anxiety and attentional bias but post-hoc tests revealed a tendency of dysphoric individuals to avoid the threatening faces.

Some later studies used general threat pictures rather than emotional faces. In one such study, Bradley, Mogg, Falla & Hamilton (1998) used pictures that were either severe (e.g., assault victims, mutilated bodies) or moderate (e.g., man behind bars, soldiers) in threat value based on evaluation by judges. Critical trials were the same as the preceding study, displaying a threat picture paired with a neutral photograph, displayed side by side for 500 ms. Results showed that HA participants were quicker to react to probes replacing higher rather than moderate threat pictures, implying a greater vigilance for higher threat cues. In a subsequent study, Mogg et al. (1998) employed pictures from the International Affective Picture System (IAPS; Lang Bradley & Cuthbert, 1995) and reached the same conclusions.

The same authors (Mogg & Bradley, 1999) repeated the study using probe classification rather than probe position. Probe classification produced three times as many errors as probe position, and participants were slower by approximately 200 ms in responding the probes. Mean response times for probe position are in the order of 300-400 ms and 500-600 ms for probe classification (Mogg et. al, 1998, Mogg & Bradley,
1998b). However, results still provided more evidence in favor of an attentional bias towards threat stimuli.

Using the pictorial dot probe offered other options to measure the time course of attentional allocation and bias. An additional means to gain insight regarding direction of attention was the addition of tracking eye-movement to the basic task. One such study was undertaken by Bradley, Mogg and Millar (2000); they added eye tracking to the basic dot probe task using pictures of happy, threatening and neutral facial expressions displayed for 500 ms. Gaze tracking measured “overt” shifts in attention, that is, voluntary shifts in attention, while reaction time to probes provided a measure of covert orienting of attention (involuntary shifts). Dysphoric individuals were faster in responding to probes replacing threatening stimuli and eye tracking patterns revealed that they also tended to initially orient to the threat stimuli as opposed to non-dysphoric individuals.

A masked version of the pictorial dot probe has also been developed (Mogg & Bradley, 1999), as have studies to investigate the time course of attention by manipulating stimulus duration. Bradley et al. (1998 a) investigated the time course of attentional processes and found HA individuals displayed higher vigilance towards threat faces (but not towards emotional faces in general) when the stimuli were displayed for 500 ms and 1250 ms.

A substantial amount of research using dot probe has been conducted on removing the uncertainty surrounding information processing biases in social anxiety (review by Heinrichs & Hoffmann, 2001). Some studies on this topic have suggested vigilance for social-threat cues while others have found avoidance, partly due to the difference in
methodologies of the studies. Early studies using social-threat words did not reveal a clear relationship between social anxiety and attentional biases. Asmundsen and Stein (1994) were the first to investigate this relationship using social phobics. They used a modified version of the dot probe; displaying word pairs in top-and-bottom orientation for 500 ms and instructing participants to read aloud the top word in every trial. Following stimulus offset, participants were to respond as quickly as possible to probe onset by pressing the appropriate button indicating probe position. Results indicated that social phobics responded quicker to the probe regardless of probe location when the social threat word appeared on the top. Thus, although the study proved that social phobics selectively attend to socially evaluative words, it suffered from interpretative problems due to the aforementioned decrease in RT regardless of probe position. As such the results could also be interpreted as individuals displaying enhanced vigilance after reading a threat word. Two other studies with similar populations also resulted in no significant effects towards social threat cues by socially anxious people (Horenstein & Segui, 1997; Sanz, 1997).

2.2.3.5 Social anxiety

Mansell, Clark, Ehlers and Chen (1999) tested socially anxious individuals under conditions of socially evaluative threat and no-threat. Participants were divided into groups on the basis of their social anxiety scores (lower (<8) and upper quartile (>17) scores on the FNE, respectively). Conditions of social threat were induced by informing participants that they were to give a speech to a live audience after the test. An equal number of participants were tested under threat and no-threat conditions. Stimuli in this case consisted of pictures of faces (happy, threat and neutral) paired with a picture of a household object. The threat condition induced attention away from emotional faces
(both positive and negative) while the no-threat condition did not find any differences in attentional bias between the test and control groups. The authors also covaried out the differences in trait anxiety and depression indices and found an attentional avoidance effect. The fact that social phobics avoid emotional faces stimuli while other phobics (see below) direct attention to emotional stimuli lead the authors to suggest that phobics exhibit attentional biases in directions which reduce the uncertainty around the threat stimuli. For instance, findings for individuals with high social anxiety have not been consistent with those obtained from various other groups afflicted by anxiety disorders and phobias. Lavy and van den Hout (1993) found a similar attentional bias for spider related words and pictures with spider-phobics, which according to Mansell et al. (1999) was the best way for individuals to reduce uncertainty about the spider. A social phobic, on the other hand, breaks eye contact by looking away from the face cue to achieve the same end result. However, a later study by the same authors (Mansell, Ehlers, Clark, & Chen, 2002) using threat words as the salient stimuli with high and low socially anxious college students under conditions of social threat and no-threat did not find any bias towards or away from the threatening stimuli.

Using pictorial stimuli, Mogg & Bradley (2004) examined social phobics for bias to threat cues and the time course of their attentional processes. Pictures of emotional facial expressions served as the emotional cue and were displayed for either 500 ms or 1250 ms. A significant trend to attend to preferentially to negative faces as opposed to positive or neutral faces was observed for the clinical population in the 500 ms condition. In the 1250 ms condition however, no bias was found for either the clinical or the control group.
2.2.3.6 Drug abuse

Lubman, Mogg and Bradley (2000) compared methadone-sustained drug users to age-matched controls for attentional bias towards drug-related cues. Participants were shown drug-related pictures (needles, spoons, heroin wraps, etc.) paired with neutral pictures for 500 ms. The hypothesis predicted the existence of a drug related bias as posited by some cognitive theories (Robinson & Berridge, 1993); consequently relapse has been linked to an attentional bias towards drug related stimuli (Wikler, 1965; Siegel, 1979; Stewart et al., 1984; Childress et al., 1986; Baker et al., 1987; Tiffany, 1990). Findings supported the prediction in that opiate drug users displayed an attentional bias towards drug related information.

2.2.3.7 Smoking and alcoholism

The same theories incited research for attentional bias in smokers and alcoholics towards their respective drugs. Townsend and Duka (2001) extended the research of Lubman et al. (2000) and adapted it to investigate for a bias in non-dependent heavy social drinkers towards alcohol-related cues as opposed to occasional social drinkers. Critical trials consisted of an alcohol-related cue (word or picture) paired with a neutral non-alcoholic cue. All cues were displayed for 500 ms. Results confirmed a bias towards alcohol related cues in the heavy drinker group. Ehrman et al. (2002) based their research on the two studies mentioned above and examined whether current cigarette smokers displayed an attentional bias towards smoking cues as opposed to non-smokers and former smokers, respectively. They found smokers to have a significantly higher bias towards smoking cues than non-smokers while former smokers had an intermediate level of bias.
2.2.3.8 Eating disorders

The cognitive model of Vitousek and her colleagues (Vitousek & Holon, 1990; Vitousek & Orimoto, 1993) identifies two basic cognitive factors to blame for causing and maintaining eating disorders; individuals’ body image (shape and weight) and the schema biased processing of the body image. Clearly, the second factor is reminiscent of the models of Beck (1976) and Bower (1981). Earlier studies on body image and eating disorders relied heavily on data mainly collected through self-report questionnaires. Using self-report measures is potentially limiting because it may be confounded by distortions of self-image and denial (Fairburn et al. 1991; Vitousek & Orimoto, 1993). The dot probe, on the other hand, can provide an objective measure of attention and response to food cues. Following this line of reasoning, Reiger, Schotte, Touyz, Beumont, Griffiths and Russell (1998) examined the existence of a bias toward body and shape related stimulus words in patients of anorexia nervosa, bulimia nervosa, and controls. They used words reflecting large or thin physiques paired with neutral words on critical trials. Individuals with eating disorders exhibited a bias towards words describing a large physique and away from neutral words and words representing a thin physique. Taken together, these results may indicate that individuals with eating disorders process information related to fatness while ignoring information related to thinness. If so, it could indicate a fear of attaining a large physique despite evidence to the contrary, explaining why patients of these eating disorders show an aversion to food.

In yet another study, Placanica and her associates (Placanica, Faunce, Soames Job, 2002) tested high and low scorers on the Eating Disorder Inventory-2 (EDI-2) under fasting and non-fasting conditions for bias towards food stimuli. They found a bias towards high-calorie foods under the fasting condition across all participants while high
EDI-2 scorers showed a bias to low-calorie (non-fat) foods only when not fasting. This hunger-driven bias towards high-calorie foods may shed some light on the binge-purge and cycle found in bulimic-nervosa.

2.2.3.9 Pain and miscellaneous areas

Some other non-traditional areas where the dot probe has been applied of late include attentional bias towards pain stimuli (Keogh et al, 2001; Dehgani, Sharpe & Nicholas, 2003) in chronic pain sufferers, towards words related to Irritable Bowel Syndrome (IBS) for IBS sufferers (IBS; Gutierrez, 2001), to sexual and violent words in victims of sexual trauma (Bush, 2000).

2.2.3.10 Limitations of the dot probe

Although it has been used with considerable success in various studies with a host of anxiety and other disorders, the dot probe task has several limitations. The most glaring insufficiency of the task is that it does not provide a complete picture of the timecourse of attention but only of attention at the instant of probing.

One of the more serious criticisms of the task, and one it shares with the Stroop task is that the salient stimuli are presented in the foveal region. Although foveal vision and attention are not the same, it is believed that it is impossible not to attend to information presented within a 1-degree radius of fixation. Thus, results of the two tasks cannot conclusively say whether threatening stimuli attract attention or hold it once they are detected. However, the task is easy to administer and provides a direct reading of attentional bias at the instant of probing.

2.3 Connectionist Models of Attention

Williams et al., (1997) identified two main issues concerning the study of attentional biases: the cause and the mechanism. Causes refer to the reasons why
attentional biases manifest themselves only in some people under certain conditions. Mechanisms refer to the point in the information processing system at which they act. The first issue has been addressed by different paradigms used to study attentional biases. For example, robust relationships have been established between trait anxiety, state anxiety, stress and their affect on attentional biases using the dichotic listening paradigm, the Stroop and dot probe tasks. Studying mechanisms, on the other hand, has not been as straightforward. Developing a clear understanding of the mechanisms is important to understand attentional biases more thoroughly and devise more effective treatments for the various disorders.

An attractive method to study mechanisms of various constructs is through computer simulations. Developing such a simulation allows researchers to intervene, change variables and measure their effect. Neural Networks (NN) models (also known as Parallel Distributed Processing (PDP) models) are computational modeling paradigms based brain operations and are the preferred modeling paradigm when simulating attentional biases on computers. In fact Williams et al. revised their 1988 model in 1997 to explain in PDP models (Williams et al., 1997).

The second analysis tool employed in the current study is a Bayesian belief network, or Bayesian network (BN) for short. BN are probabilistic graphical models. The current study represents the first known attempt to use BN to arrive at probabilistic relationships between variables involved in the dot probe task. Current popular applications of these models are in the fields of mainstream computer science (e.g., datamining- discovering relationships between relationships from data, expert diagnostic systems, etc.) and business and finance (e.g., risk analysis for insurance and other
projects, stock market prediction). Bayesian networks’ application to cognitive reasoning has been rather limited, mainly said to be more suited to higher order reasoning tasks than simulating lower, automatic processing tasks. One of their most widely known applications is in the sometimes annoying Microsoft Office Helper and troubleshooter. BN are most applicable in areas where relationships between variables are known (this is explained below). BN also offer an intuitive method of modeling the relationships graphically. The current study will use BN to develop a causal model of attentional bias as per the dot probe task and then fine-tune the probabilistic relationships between the variables.

In this section, the basics of the main NN models developed for studying attentional bias, the models to simulate the Stroop task by Cohen et al. (1990) and the extension of the same to the emotional Stroop by Matthews and Harley (1996) are summarized. The section first explains the theory and working of NN. The two models are then discussed in detail. The theory and working of BN are explained next along with an example of how they will be applied in the current study.

2.3.1 Neural Networks

A neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

- Knowledge is acquired by the network from its environment through a learning process.

- Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge. (Haykin, 1998, p.2)

2.3.1.1 Overview

A NN consists of a number of small processing elements that are also referred to as neurons. Each neuron is capable of performing only very simple calculations;
computing the input and using a *transfer* or *activation* function to compute an output. A single neuron in the human brain is much slower than a microprocessor (by about an order of $10^6$, with a neuron taking $10^{-3}$ s per operation compared to the $10^{-9}$ s of a microprocessor). The brain overcomes this disadvantage of speed by using parallel processing. Each neuron is connected to numerous other neurons, with the connections between them known as *synapses* allowing simultaneous parallel activation of varying neural circuits. Shepherd and Koch (1990) estimated the number of neurons in the brain at 10 billion with 60 trillion synapses. As such, NN try to emulate this natural parallelism. Each synapse has a strength associated with it; referred to as the “*interneuron connection strengths* or *synaptic weights*” (Haykin, 1998, p.2). Each units’ input is a summation of the weighted output of all the other active units that project to it. Synaptic weights can be positive (excitatory) or negative (inhibitory). Clearly, these weights are the most basic variable in a NN; they are adjusted according to the application area of the NN using a *learning algorithm*.

2.3.1.2 Learning

Learning in a NN is the process of adjusting the synaptic weights according to the variations in the learning data (i.e., the problem). Indeed, learning paradigms are classified into learning with a teacher (*supervised learning*) and learning without a teacher (*reinforcement* and *unsupervised learning*) (Haykin, 1998). In supervised learning, for example, the NN is provided with an input and the known result for that input (called *target output*). The network computes the output based on the input (called *computed output*) and compares it against the target output. The weights are adjusted using a function to minimize the error between the target and computed outputs. A detailed explanation of supervised learning is provided in the next section. Alternatively,
when learning without a teacher, the network uses one of two techniques: (1) reinforcement learning, where the NN has an in-built “critic” (some scalar index of performance) and learns by minimizing the scalar index or (2) unsupervised learning which, instead of a critic has a task independent measure that optimizes the free parameters of the network. Specifically, in unsupervised learning, the network uses the task independent measure to find statistical regularities in the input data enabling it to form internal representation of data to automatically derive new classes of data (Becker, 1991). For a full discussion on supervised and unsupervised learning, see (Haykin, 1998).

2.3.1.3 Supervised learning.

Supervised learning consists of two phases, a training phase followed by a test phase. The training phase consists four steps that culminate in the network being able to compute the correct output for each of the input cases. The steps are:

1. Input data with known outputs (target outputs) to the network.
2. Allow the network to compute its output (computed output) based on the given input.
3. Compute the difference between the computed output and the target output.
4. Use the learning algorithm to adjust the synaptic weights based on the magnitude of the error difference.

The test phase supplies new inputs not previously seen by the network.

2.3.1.4 An example of supervised learning.

An example of supervised learning is a model to predict whether it will rain on a particular day or not. The first step is to identify the independent and dependent variables in the model and choose the NN architecture\(^2\) for the problem. In this example, the inputs

\(^2\) Architecture of a NN refers to how the neurons in the network are connected to each other. Three basic architectures are single-layer feedforward networks, multi-layer feedforward networks and recurrent networks. Multi-layer feedforward network architecture is the one most commonly used in supervised learning including in the current study and in the simulation of the Stroop (Cohen, et al. 1990) and the
are the four of the most basic variables known to affect weather on a given day, pressure, humidity, temperature, and wind-direction. The output consists of two nodes, rain or no-rain; signaled by activation of the respective nodes. The next step is to initialize the network by assigning weights to the inter-neuron connections. To start with, these weights may be generated randomly within some range. Once the network is initialized, it must be trained, constituting the third step. Input for the training phase consists of the four input variables and the value of the observed (target) output for a given day (i.e., whether it rained or not). The training phase involves the network computing the output for a large number of training input cases, and adjusting the synaptic weights accordingly until the computed output matches the target output for each case. The successful completion of all four steps culminates in the network being ready to predict whether or not it will rain on a given day given the temperature, pressure, humidity and wind-direction for that day.

2.3.2 Details and Theory

2.3.2.1 Overview

Figure 2.4 shows a basic multi-layer feedforward NN (multi-layer refers to multiple layers of weights) that uses the back-propagation algorithm for learning (also known as the backpropagation network (BPN) or the backpropagation architecture. The BPN consists of an input layer, one or more hidden layers and one output layer. Figure 2.4 provides an example of a network that contains an input layer of two nodes ($I_1$ and $I_2$), one hidden layer of four nodes ($H_1$ through $H_4$) and one output layer of two nodes ($O_1$ and $O_2$). Input and output neurons are always in one of two states, firing emotional Stroop (Matthews & Harley, 1994). For detailed information on network architectures see Haykin (1994).
(active) or not-firing (inactive), whereas hidden neurons, can have a discrete output or use the activation level itself as the output for the node (explained in detail below). Input to the network is provided by applying an activation pattern over the input nodes. Output of the network is given by activation of one of the output nodes. The network has one input unit for each input of the problem being modeled, and one output unit for each possible output (e.g., the NN in the weather example in the previous section contains four input variables and two output variables). The number of hidden layers and the number of units in each layer is also problem specific, but typically the number is between the number of input nodes and the number of output nodes; lesser than the number of input units and greater than the number of output units (Blum, 1992). Hidden nodes provide non-linearity to the network.

A BPN is considered fully connected if every neuron in one layer is connected to every neuron in the next layer. Figure 2.4 is an example of a fully connected BPN. When the neurons in one layer are only connected to certain neurons in the next layer, the BPN is considered partially connected. In either case, there are no connections between neurons in the same layer.
2.3.2.2 Notation

This section describes a notation based that used by Haykin (1998) to clearly refer to weights on any connection in any of the weight layers and to refer to any unit in any of the three layers of units. As mentioned earlier, a NN consists of layers of units (input, hidden and output). Weights are denoted by $w_{ij}^n$, which is the weight $w$ on the $n$th layer of weights between nodes $i$ and $j$ (note that nodes $i$ and $j$ are in different layers of units, for instance, the input and hidden or hidden and output layers). For example, the NN of Figure 2.4 consists of two layers of weights; the first (denoted by $w^1$) between the layer of input units and the hidden layer and the second (denoted by $w^3$) between the hidden and output layers. Therefore the weight of the connection between the first unit in the input layer and the third unit in the hidden layer is denoted by $w^1_{13}$. Similarly, the weight
on the connection between the second unit in the hidden layer unit and the first output unit is denoted by $w_{21}^2$. Input to a unit (not shown) is denoted by the letter $x$; so $x_j$ represents the input to the $j^{th}$ unit in a layer. Activation level of the unit is denoted by $a$, so $a_j$ denotes the activation level of the $j^{th}$ neuron.

### 2.3.2.3 Initialization

The network can be initialized by randomly assigning randomly generated weights to the connections between units. Alternatively, the designer can assign weights to the connections according to the problem. In either case, the weights are adjusted according to the problem domain using the learning algorithm during the training phase.

### 2.3.2.4 Node Details

Figure 2.5 is a detailed representation of a node of the network of Figure 2.4. The figure refers to the third node in the hidden layer of the network; receiving input from the two input units ($I_1$ and $I_2$) and transmitting the result to the two output units (not shown). As illustrated, the processing unit performs two basic tasks:

1. Computes the net input (denoted by $x_3$ with the computations performed by the summation unit).
2. Computes the activation level from the net input (denoted by $a_3$, using the transfer function to perform the computations).

Output of the node may or may not be the same as the activation level.

### 2.3.2.5 Net input

The net input to a hidden unit is typically the sum of the product of the output of each input unit and the synaptic weight of the connection between the two.

Sometimes a bias value ($b$) may be added to the net input of each node to lend
variability to the network. The bias can be a constant set by the experimenter so it is the same for every node, or a random number generated from within a specified range. Therefore in this case, the net input is:

\[ x_3 = I_1 w_{13}^1 + I_2 w_{23}^1 + b \] \hspace{1cm} --2.1

where \( I_1 \) and \( I_2 \) are the outputs of the input units. The above equation can be generalized to give the net input for any arbitrary node (\( j^{th} \) node of the \( n^{th} \) layer) as:

\[ x_j = \sum a_i w_{ij}^n + b \] \hspace{1cm} --2.2

where \( a_i \) is the activation of the units in the previous layer.

### 2.3.2.6 Activation level

Computing the activation level consists of applying the transfer (activation) function to the net input. A BPN requires the transfer function to be non-linear,
continuous and monotonous (i.e., differentiable at any point of the curve) (Haykin, 1998; see also Cohen, et al. 1988). A class of functions known as sigmoid functions fulfills these requirements and the function that is most commonly chosen as the activation function among the functions of this class is the logistic function given by:

Figure 2.6 Graph of the logistic sigmoid function

\[ Activation, f(x_j) = a_j = \frac{1}{1 + e^{-x_j}} \]  

---2.3

Figure 2.6 depicts the graph for the logistic function of equation 2.3. Note that the slope of the function is the highest when the net input is zero and increases more slowly the higher the value of activation.

2.3.2.7 Output

As mentioned above, the output of the node depends on whether the node is a hidden or output node. For a hidden node, the output of the node is usually the same as the activation level. However, a threshold value is set for output nodes (by the experimenter) and the node fires if and only if the activation level of the node exceeds the
set threshold. Alternatively, the output may be selected as some function based on the activation level of an individual unit or all the units in the output layer. The number of output units of the network depends on the problem being modeled and the number of units that can be active simultaneously depends on the problem parameters. In short, The network loops until an output is obtained (i.e., one of the nodes fires). Cohen et al. (1990) used one such function in their model; the process is explained in the next section.

2.3.2.8 Training

During training, the NN is supplied with the target output for each input pattern. The network computes the output (computed output) and the learning algorithm compares the computed output with the target output to calculate the error. The learning algorithm used in the two simulations of the Stroop task is known as the delta rule. Error-information is then propagated back through the network and each unit of the network adjusts the weights of its connections according to some error-minimizing function (e.g., gradient-descent, so called because the function estimates the direction in which to move down the slope of the function to as to minimize the function). Error correction may be carried out after each of the

Figure 2.7 shows the flow of computation and weight-correction for a BPN. Backpropagation algorithm is by far the most popular error driven learning algorithm.

Summarizing, the training phase consists of the following steps:

- Activate the appropriate input units.
- Compute the output. The network is allowed to run until one of the output units fires.
- Compute the error (difference between the computed and target outputs).
- Propagate the error information backwards through the network to all units
Figure 2.7 Flow of activation (solid lines) and error (dotted lines) in a multi-layer backpropagation neural network.

and adjust the synaptic-weights using the backpropagation algorithm. Error correction may be performed after a single input activation is presented to the network; or after all the input patterns have been presented to the NN. In the latter case, the network performs the corrections based on the total error (the mean-squared-error [see Haykin, 1998]) for all input patterns, which is the square root of the mean of sum of the squares of error for each input pattern. Mathematically,

\[
\text{mse} = \sqrt{\frac{\text{error}_1^2 + \text{error}_2^2 + \ldots + \text{error}_n^2}{n}}
\]

The network performs error-correction until the error falls within the specified limit set according the problem.
2.3.2.9 Testing

Once the training is complete, the network is ready to accept new inputs (as opposed to the training data) and produce an output.

2.4 Connectionist Models of the Stroop Task

Two PDP models of the Stroop task have been developed to date, the first by Cohen et al. (1990) that simulated the findings of the basic Stroop task and the second, a simulation of the emotional Stroop, reported by Matthews and Harley (1996). This section discusses the design of both networks, along with issues concerning training and testing of the two. In order to be considered a successful simulation of the Stroop task, a model must replicate the main empirical findings of the task.

- **Word reading is faster than color naming.** Mean time to read a color word is 350-450 ms while naming a color patch of a row of X’s takes about 200 ms more (550-650 ms) (Dyer, 1973; Glaser & Glaser, 1982).
- **Word reading is not affected by color ink.** Color of the word to be read has virtually no affect on the time to read the word.
- **Words can influence color naming.** Content of the word interferes with color naming, conflicting words cause a substantial increase (variable but commonly 100 ms) in the RT to name colors. Conversely, congruent words facilitate performance in color naming, reducing RT by 20 ms (Regan, 1978) to 50 ms (Kahneman & Chajczyk, 1983).
- **Facilitation is less than interference.** Although congruent stimuli have been used only sparsely, general findings are consistent with the pattern mentioned above in that the amount of facilitation (20 ms) is much less than interference (100 ms).

2.4.1 The Cohen Model

Cohen et al. (1990) used a partially connected PDP model to simulate the Stroop task (Figure 2.8). The model produced the main empirical findings of the Stroop task. Input to the network consisted of specifying the task and task parameters (i.e., whether to
perform color naming or word reading and the ink color and the word). Output consisted of generating the response for the input through activation of the correct ink color for the color-naming task and the correct word for the word-reading task. The Cohen model computed the time course of a psychological process by presenting results in terms of reaction time (RT), computed by deriving a linear relationship between the number of iterations taken by the network to compute the output and typical RT for the task. As such, the authors required the network to mimic the variability in reaction times of human participants performing the Stroop task. This mandated some changes to be made in the way each unit computes input and the selection of the final output by the network.

2.4.1.1 Structure

Cohen et al. (1990) used a partially connected BPN, consisting of two processing pathways, one for color naming and one for word reading. Each of the pathways can be thought of as a distinct neural network with both competing for the final output, achieved by connecting the hidden layer to both output units (see Figure 2.8). The network contained six input units to specify ink color, task and the words to be read, and two output units representing the possible outputs (i.e., red or green). The model could handle only two words, “Red” and “Green”, printed in two possible ink colors, RED and GREEN. Inputs were presented as a pattern of activation over the input nodes.

2.4.1.2 Initialization

Cohen et al. (1990) assigned small, random weights to the connections between the hidden and input layers and intermediate values (either +2 or –2) to connections between the input and hidden layers. In the latter case, values were chosen to obtain a straightforward mapping from the input layer to the hidden layer.
Figure 2.8. Neural network model for simulation of the Stroop task

2.4.1.3 Net input

Each unit computed net input in a similar manner to the one described in equation 2.2. Two changes were made to the manner in which net input was computed for each unit; the first was to allow the network to simulate the time course of a psychological process. The initial change was based on the cascade models (McClelland, 1979), which also simulated the time course of psychological processes; net input at any instant of time \( t \) was defined as the running average of its net input over time. Mathematically:

\[
\bar{x}_j(t) = \tau x_j(t) + (1 - \tau) \bar{x}_j(t-1)
\]

where, \( \tau \) is a rate constant and,
\[ x_j(t) = \sum_i a_i(t)w_{ij} + b \] --2.6

Note that equation 2.6 is the same as equation 2.2 except that \( a_i(t) \) is the activation at time \( t \). Additionally, using this function guaranteed that the network would always reach a stable asymptotic state and result in an output.

The second change was made to add variability of performance to the network; a normally distributed random bias was associated with each hidden and output unit and added to the net input (denoted by \( b \) in equation 2.6).

2.4.1.4 Activation

The logistic function (equation 2.3), applied to the net input computed according to equation 2.6, was used to compute the activation. Mathematically,

\[ \text{Activation, } f(x_j(t)) = a_j(t) = \frac{1}{1 + e^{-x_j(t)}} \] --2.7

Again, note the only difference between equation 2.7 and equation 2.3 is in the value of the net input.

2.4.1.5 Output

Output of the network was indicated in the same way as a typical BPN, by the firing of an output node (i.e., “red” if the node representing “red” fired and “green” if the other node fired). Cohen et al. (1990) introduced variability in this step by making the firing of the output unit dependent on the result of a random walk (Link, 1975) or diffusion process (Ratcliff, 1978). Adaptation of these processes to computing the output of the network consisted of associating evidence accumulators with each of the output units. The accumulators were set at 0 at the beginning of each trial and a small amount of evidence was added at the end of each time step; evidence added was random and
normally distributed with mean $\mu$ based on the difference between the activation of the unit and activation of the most active alternative and fixed standard deviation ($\sigma = 0.1$). So for the $i^{th}$ output unit, the mean $\mu_i$ is given by

$$\mu_i = \alpha(a_i - \max(a_{i,j}))$$

$\alpha$ is the rate of accumulation of evidence and was set at 0.1 through all their trials.

The threshold was set at 1.0 for the evidence; so the output unit fired when the evidence associated with a unit exceeded 1.0.

2.4.1.6 Training

One difference between training the Stroop model and the weather example cited earlier was that the training in this case was completed separately on the word reading and color naming tasks, rather than both tasks at the same time. An input pattern consisted of the ink color or the word and the task to be performed. So an input for the color-naming task with color “red” was represented by “RED-COLOR-NULL”, activating the input unit for color “red” and the task demand unit (TDU) for “COLOR NAMING” only. The target output in this case was “red”. The network was then allowed to reach an asymptotic level of activation and generate an output and correct the weight in accordance with the error between the computed and target outputs using the backpropagation algorithm.

The Cohen model differed from a typical BPN in initializing and updating the weights. Firstly, connections in the first layer were randomly assigned values of either +2 or −2. Secondly, weights on the connections between the TDUs and the intermediate units were kept constant (and not allowed to be changed during training) so that the
activation of the TDU did not provide any extra information to the intermediate units.

However, the authors suggested that these could be learned.

One training objective was to make word reading more automatic than color naming, achieved by training the network on ten times as many word-reading stimuli than color naming stimuli, allowing it to strengthen the word-reading connections much more than color-naming ones. Table 2.1 lists all the training patterns.

Table 2.1: Input patterns and corresponding outputs used for training the network by Cohen et al. (1990)

<table>
<thead>
<tr>
<th>Ink color (Input)</th>
<th>Word (Input)</th>
<th>Task</th>
<th>Condition</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>RED</td>
<td>WORD READING</td>
<td>Congruent</td>
<td>RED</td>
</tr>
<tr>
<td>Green</td>
<td>GREEN</td>
<td>WORD READING</td>
<td>Congruent</td>
<td>GREEN</td>
</tr>
<tr>
<td>Red</td>
<td>GREEN</td>
<td>WORD READING</td>
<td>Incongruent</td>
<td>GREEN</td>
</tr>
<tr>
<td>Green</td>
<td>RED</td>
<td>WORD READING</td>
<td>Incongruent</td>
<td>RED</td>
</tr>
<tr>
<td>Red</td>
<td>GREEN</td>
<td>COLOR NAMING</td>
<td>Congruent</td>
<td>RED</td>
</tr>
<tr>
<td>Green</td>
<td>RED</td>
<td>COLOR NAMING</td>
<td>Congruent</td>
<td>GREEN</td>
</tr>
<tr>
<td>Red</td>
<td>RED</td>
<td>COLOR NAMING</td>
<td>Incongruent</td>
<td>GREEN</td>
</tr>
<tr>
<td>Green</td>
<td>GREEN</td>
<td>COLOR NAMING</td>
<td>Incongruent</td>
<td>GREEN</td>
</tr>
</tbody>
</table>

2.4.1.7 Testing

Testing involved providing all inputs for the task (color, word and task) to the network and allowing it to cycle through until the output reached an asymptotic value. In all, the network was tested on 12 different input patterns (listed in Table 2.1) representing every possible condition, for both color-naming and word-reading. The conditions tested were congruent (word same as color), conflict (different word and color) and control (only the color or the word depending on whether testing is for color naming or word reading). A conflict condition of naming ink color for the word “Green” printed in “Red” ink was presented to the network as activation pattern “RED-COLOR-GREEN” (i.e., activating nodes representing ink color “red”, word “green” and color naming TDU). Cohen et al. (1990) recorded the number of iterations it took the network to reach an
asymptotic activation of the output units and used it to derive a relationship between the
number of iterations and the reaction time by comparing the number of iterations against
established reaction times for each condition. Naturally, this meant the relationship
between the number of iterations and RT was dependent on the task being simulated.

2.4.1.8 Simulations and results

Cohen et al. (1990) performed six different simulations to test findings of the
Stroop task in four different categories and explained those in terms of the PDP model.
The four categories were:

1. **Strength of processing** which primarily explained the main empirical
findings of the Stroop task on the basis of connection strengths.

2. **Stimulus onset asynchrony (SOA)** effects, which investigated observing
interference even when the ink color was displayed before the actual word. This
simulation was the only one not in agreement with actual data. Specifically, the
model displayed some influence of color on word reading when the color is
presented early.

3. **Practice effects**: The simulation was based on the study conducted by
MacLeod and Dunbar (1988) who trained individuals to associate shapes with
colors, creating a novel task which the individuals had not practiced before. The
Stroop task was then constructed in which individuals were presented with shapes
in different colors and were required to name color the shape was originally
associated with. Results indicated that practice on associating colors with shapes
increased performance on color naming the shapes in the Stroop task consistent
with the Power law. Simulation of this task included two different simulations
which investigated the Power law and practice in the Stroop task and developing
automaticity with practice. The first of the two simulations plotted RT as a function
of number of training trials ($N$) and found performance of the network increased on
color naming the shapes according to the Power law.

4. **Allocation of attention**: The final two simulations tested the affect of
attention allocation to performance on the two tasks. Researchers have proposed
somewhat opposing views on attention allocation and its affects on performance.
Some researchers define automatic tasks as requiring absolutely no attention
(Posner & Snyder, 1975; Shiffrin & Schneider, 1977), going so far as saying lack of
attention should not influence performance on such tasks (Posner & Snyder, 1975),
while others have challenged this claim, saying few, if any, processes can function
without attention (Kahneman & Treisman, 1984; Logan, 1980).
Cohen et al. addressed the issue with their model. In the first of two simulations, they performed the two tasks on their model with varying degrees of attention allocation (activation of the TDUs) and found that color naming required more attention than word reading to maintain a given level of performance. More importantly, however, performance on both tasks degraded by reduced attention. In the second of the two simulations, the authors investigated response-set effects. More specifically, they evaluated whether words and objects that are not part of the response set cause significantly less interference than those that are (e.g., the word BLUE is never a correct response in the current simulation and so is not a part of the response set). The authors successfully simulated the picture-naming task (Dunbar, 1985) and observed significantly less interference for words that were not part of the response set.

Williams et al. (1997) updated their earlier model (Williams et al., 1988) within a PDP framework following the simulations by Cohen et al. (1990). They re-conceptualized the ADM as the input units, assigned the task of tagging input with a threat value and the RAM as the TDU. Despite these revisions, their core assumptions of the interaction between state and trait anxiety remained virtually unchanged.

Matthews and Harley (1996) built upon the model of Cohen and colleagues (1990) by adapting it to simulate the emotional Stroop. However, their focus was on trying to explain the cause of attentional bias, and as such they were not concerned with adding variability to their model.

2.4.2 The Matthews and Harley Model

Matthews and Harley (1996) extended the Cohen model by applying it to the emotional Stroop task. The authors set realistic objectives for the simulations, keeping in mind the lack of simulations and other investigations into the mechanisms of the task.
Their main objective was to investigate three different mechanisms for generating attentional bias (exposure, intensity and attentional), based on different explanations of attentional biases. Simulating the emotional Stroop involved changing the network to present emotional words as inputs. Also, the network was not designed to simulate the time course of attention, that is, in contrast to the work of Cohen et al. (1990), Matthews Harley (1996) did not present results by comparing RT but compared relative activation of the output units. Details are explained in the sections below.

2.4.2.1 Hypotheses

As stated, the model tested three qualitatively different explanations for attentional biases. The first, known as the exposure hypothesis, proposed that repeated exposure to emotional stimuli lead to an attentional bias towards emotional stimuli causing an anxious person to be more practiced, and therefore more automatic, in reacting to emotional stimuli.

The intensity hypothesis was second viable alternative, and explained the basis of attentional biases as distressed individuals perceiving the same emotional stimuli as more potent (higher intensity) than normal individuals, and therefore placing a higher priority on processing that information. The hypothesis was further branched into state and trait portions, with high state anxiety individuals perceiving higher intensity only during test conditions while high trait anxiety individuals felt the same (higher) intensity all the time.

The third hypothesis stemmed from Matthews and Wells (1995) explanation of observed interference in the Stroop task. They suggested that attentional bias occurs due to a coping strategy adopted by distressed individuals to monitor potential sources of threat. Arguably, the coping strategy leads such individuals to pay more attention to
emotional stimuli, which in turn causes larger interference effects. This called for activation of the task demand unit during testing.

2.4.2.2 Structure

Matthews and Harley (1996) followed an evolutionary approach, extending the Cohen model for the emotional Stroop. After two unsuccessful architectures (Figure 2.9 (a)) that did not yield satisfactory results for standard Stroop interference, they arrived at the final model (Figure 2.9 (b)). The first model was a straightforward extension of the Cohen model, while the final architecture featured extra connections as well as an additional TDU to monitor threat. The final model had nine input units, six hidden units, and five output units.

As can be seen from Figure 2.9, input units represent semantic features of the word rather than the word itself. So each word was presented as activation of a unique combination of input units, and the network was trained to output the word corresponding to the semantic units activated. Color indicated a color word and color type indicated the degree of redness, thus by activating both the units simultaneously word “red” was represented. “Monster” is associated with a large object (being) and negative emotion and therefore is presented to the network by the activation of (large) SIZE and (negative) EMOTION. The semantic codes words for all the words used by the network as seen by the output produced are listed in Table 2.2. Presenting input in this manner had two distinct advantages; firstly, it was consistent with psycholinguistic theories stating speech processing is a two level process. Secondly, it allowed the network to learn semantic similarities between words.
2.4.2.3 Initialization

The network was initialized using the same as the initialization method used by Cohen et al. (1990).

2.4.2.4 Input

Since the network was not intended to model the RT of the psychological process, Matthews and Harley used the basic BPN equations to compute net input (equation 2.2) and activation (2.3) for each unit.

2.4.2.5 Output

Obtaining output was rather straightforward in this case, given that the output produced in the first iteration of the network became the final output and was used for comparisons against baseline conditions.

2.4.2.6 Training and testing

Differences in architecture apart, both networks implemented the BPN architecture and were trained using the same basic algorithm. Presenting all training patterns to the network is called an epoch. Matthews and Harley trained the network for 400 epochs (i.e., cycling 400 times through all the training patterns shown in Table 2.2). The number of training patterns of each type used for training and the activation values depended on the hypothesis. Training patterns for the baseline condition are displayed in Table 2.2. The table highlights an interesting distinction in the training approaches for the two networks; the Matthews and Harley network was trained to ignore input in the absence of a task specification, that is, when a TDU was not activated. Otherwise, Matthews and Harley still used Cohen et al.’s (1990) approach to make word reading more automatic than color-naming, training the network more on word reading than color
Figure 2.9 Matthews and Harley Model (a) the first two models. The dotted lines were connected in Model 2 while non-existent in model 1, (b) model 3 shared the same connections for the 2nd weight layer with model 1. Connections that differ in layer 1 are shown as solid lines while those carrying over from 1 and 2 are shown in dotted lines.
naming. Results were obtained by comparing the activations of the output unit for each hypothesis condition against baseline activations. The authors manipulated training sets for each of the hypotheses. Exposure condition involved doubling (chosen arbitrarily) the number of training patterns of emotional word reading. To train for the intensity condition, Matthews and Harley changed the input from 1 to 8 for appropriate units, simulating hypersensitivity to the stimuli. Finally, attentional manipulations did not entail any changes in training but the unit was set to a low positive (0.3) value during testing.

The input patterns for which no output is listed were used to train the network to produce an output only if a TDU was activated.

Table 2.2: Training patterns and number of times each condition was presented to the network to train for the emotional Stroop task.

<table>
<thead>
<tr>
<th>Stimulus Input</th>
<th>TDU Activated</th>
<th>Output</th>
<th>Repetitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color + Color Type</td>
<td>Word Reading</td>
<td>Red</td>
<td>16</td>
</tr>
<tr>
<td>Color</td>
<td>Word Reading</td>
<td>Green</td>
<td>16</td>
</tr>
<tr>
<td>Emotion</td>
<td>Word Reading</td>
<td>Spider</td>
<td>16</td>
</tr>
<tr>
<td>Size</td>
<td>Word Reading</td>
<td>House</td>
<td>16</td>
</tr>
<tr>
<td>Emotion + Size</td>
<td>Word Reading</td>
<td>Monster</td>
<td>16</td>
</tr>
<tr>
<td>Color + Color Type</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Color</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Emotion</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Size</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Emotion + Size</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>“Red”</td>
<td>Color Naming</td>
<td>RED</td>
<td>1</td>
</tr>
<tr>
<td>“Green”</td>
<td>Color Naming</td>
<td>GREEN</td>
<td>1</td>
</tr>
<tr>
<td>“Red” + Emotion</td>
<td>Color Naming</td>
<td>RED</td>
<td>1</td>
</tr>
<tr>
<td>“Green” + Emotion</td>
<td>Color Naming</td>
<td>GREEN</td>
<td>1</td>
</tr>
<tr>
<td>“Red” + Size</td>
<td>Color Naming</td>
<td>RED</td>
<td>1</td>
</tr>
<tr>
<td>“Green” + Size</td>
<td>Color Naming</td>
<td>GREEN</td>
<td>1</td>
</tr>
<tr>
<td>“Red” + Emotion + Size</td>
<td>Color Naming</td>
<td>RED</td>
<td>1</td>
</tr>
<tr>
<td>“Green” + Emotion + Size</td>
<td>Color Naming</td>
<td>GREEN</td>
<td>1</td>
</tr>
<tr>
<td>“Red”</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>“Green”</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Color + Color Type</td>
<td>Threat Monitoring Unit</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Color</td>
<td>Threat Monitoring Unit</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Emotion</td>
<td>Threat Monitoring Unit</td>
<td>Spider</td>
<td>2</td>
</tr>
<tr>
<td>Size</td>
<td>Threat Monitoring Unit</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Emotion + Size</td>
<td>Threat Monitoring Unit</td>
<td>Monster</td>
<td>2</td>
</tr>
</tbody>
</table>
2.4.2.7 Results

The three test mechanisms tested in this simulation are stated above, namely exposure, intensity and attention mechanisms. The first condition was further divided into two parts, emotion word exposure and emotion task exposure. The former simulated performance on the two tasks after increased exposure (presenting the appropriate patterns 32 times as opposed to 16 during training) to reading emotional words. Results found an increase in reading emotional words while almost no effect was observed for color naming. The latter condition trained the network to respond to emotional words only when the TMU was activated, indicating an acknowledgement of the threat value of the word. This manipulation resulted in a marginal impairment of color naming and a similar improvement in word reading. However, the TMU was not activated during testing.

In intensity manipulations, the researchers trained the network to simulate chronic hypervigilance to threat words by increasing the activation of the threat words from 1 to 8, resulting in stronger emotional Stroop interference but impaired performance on reading emotional words.

The final condition tested performance while attending to the emotional content of the words for the tasks (word reading and color naming). The TMU was set to a low positive value (0.3) while testing to simulate concurrent attention to the threat value of the word, resulting in a significant increase in interference in color naming emotional words coupled with a small impairment the same for neutral words. The same manipulation resulted in a minor impairment of reading neutral words while having almost no effect on reading color words. All results were consistent with the findings of
emotional Stroop interference having a smaller magnitude than standard Stroop interference.

2.4.3 Pros and cons of using PDP models

Although computational models in general, and PDP models in particular are attractive methods to simulate human behavior, to understand underlying mechanisms, they are not without advantages and potential pitfalls. O’Reilly and Munakata (2000) summarized the major advantages and disadvantages of using PDP models.

2.4.3.1 Advantages

1. Models can aid in understanding phenomena and their mechanisms. For example, the Cohen, et al. NN explained some findings of the Stroop task in terms of the weights on the connections between units.

2. Models deal with complexity explaining phenomena that would be impossible to explain verbally.

3. Models are explicit. Models force researchers to think clearly about the assumptions made in the models. Further, the results obtained from these models are clear and cannot be written off as some other processes interfering in the main task. An example is the lack of conformity between the results obtained for SOA effects in the Cohen model.

4. Models allow control. Different variables can be assigned different values, using different activation functions and so on.

2.4.3.2 Criticisms

1. Models are too simple. Models have to simplify a lot of the variables in the task. Further, usually models involve only the variables used in the tasks and no extraneous variables are modeled. They do not model the biological and physical variables in any detail.

2. Models are too complex. Some researchers believe that models are too complex to be useful in explaining their behavior. This is especially true for NN, although it is clear at an abstract level that the connections between the units in different layers are strengthened, they do not really provide an explanation of what the intermediate units represent. In essence, NN follow a black box approach to modeling.
3. Models can do anything. Given enough data, PDP models can be trained to simulate just about any condition, which is analogous to a theory that explains everything.

2.5 A Belief Network Model of Attentional Bias in Dot Probe Paradigm

Neural networks are founded in theories of neuroscience and mimic the architecture of the brain; so they offer an intuitive way to simulate attention and analyze its underlying processes. One limitation of such models however, is their “black box” approach, meaning that although one has access to the synaptic weights, interpreting their values is not straightforward. Questions still abound regarding the role of the hidden units and meaning of the weights on synapses connected them. For example, both models of the Stroop task discussed above (Cohen et al., 1990; Matthews & Harley, 1996) list the weights on the connections between input and hidden units and between hidden and input units and how they are strengthened or weakened depending on the training data. However, hidden units do not represent discrete variables; therefore knowledge of the connection strength cannot be interpreted in terms of a relationship between input and output variables.³

Further, as mentioned earlier, research into the causes and mechanisms of attentional biases have yielded very consistent and robust relationships between different variables of attention and the paradigm used to test it. However, no studies attempting to quantify the said relationships have been found. The current study attempts to do just that; determine probabilistic (or belief) values of variables involved in attentional biases from the dot probe perspective.

³ As an illustration, consider the fact that the model of the Stroop task could have been implemented using a fully connected BPN as opposed to the partially connected ones that the authors used. In light of this evidence, what is the significance of the connections and the weights on those connections.
A Bayesian network (or belief network (BN)) is a tool that allows simple and elegant modeling of a system with known variables and established relationships. A BN is a graphical model that encodes probabilistic relationships among a set of variables (Heckerman, 1996).

Graphical models are a marriage between probability theory and graph theory. They provide a natural tool for dealing with two problems that occur throughout applied mathematics and engineering -- uncertainty and complexity -- and in particular they are playing an increasingly important role in the design and analysis of machine learning algorithms. Fundamental to the idea of a graphical model is the notion of modularity -- a complex system is built by combining simpler parts. Probability theory provides the glue whereby the parts are combined, ensuring that the system as a whole is consistent, and providing ways to interface models to data. The graph theoretic side of graphical models provides both an intuitively appealing interface by which humans can model highly-interacting sets of variables as well as a data structure that lends itself naturally to the design of efficient general-purpose algorithms (Jordan, 1988)

Clearly, understanding BN requires an understanding of the basic laws of probability and graph theory, both of which are explained in the sections that follow.

2.5.1 Nothing is Certain

Logic provides the tools for reasoning with absolutely certain values, like “if it rains, the grass will be wet”. Such a statement deals with absolute certainty, that is, if it is known that it is raining, the grass will be wet. Logical reasoning, however does not work well with uncertain events. Consider the problem of trying to predict whether or not it will rain given that it is cloudy. Two uncertain variables complicate prediction in this case: the state of cloudiness (e.g., the number and type of clouds) and the probability of rain given the state of cloudiness. Is the statement, “if it is cloudy, it will rain” an absolute certainty? If the number of mistakes made by the National Weather Service considered, the clear answer is “no.” Weather prediction has to deal with uncertainties; statements like, “if it is cloudy, it will probably rain.” The same is the case with the vast
majority of contexts as few real world problems have absolute certainties associated with them. Statements like “you will fail the course because of your laziness”, or “reckless driving causes accidents” (from Pearl, 2000, p.1) reflect some amount of uncertainty. Surely not all lazy people fail the course, and not all instances of reckless driving result in accidents. What these statements imply is that the particular actions mentioned increase the likelihood (or probability) of the consequence. The goal is to compute the probability in each case.

Probability implies doubt, lack of regularity, exceptionality. In other words, it is a measure of uncertainty. Logical reasoning offers four different logical connectives, namely conjunction (“both the grass and the pavement are wet”), disjunction (“either the grass is wet or it is not”), implication (“if it rains, then the grass will get wet”), and negation (“the grass is not wet”). Combining two statements can lead to an inference about an event not explicitly specified; for example, combining “if it rains, then the grass will get wet” and “the grass is not wet” lead to the conclusion that it did not rain (Jensen, 2000). Probabilistic reasoning warrants development of a similar set of rules on the lines of logical operators to combine probabilistic values. For instance, to compute the probability of rain when the probability of “rain when cloudy” is 0.8 and the probability of “cloudy” is 0.7 requires developing a method to combine the two probabilities to arrive at the required one.

Another question that begs to be answered is “how does one know the probabilities in the first place?”. There are two ways of computing the probabilities. The traditional approach, called the frequentist approach, bases the probability of an event on the frequency of prior occurrences of the same event. Perhaps the simplest example is
computing the probability of getting “tails” in a fair-coin toss; the probability is based on the number of times the coin landed on tails in an arbitrary number (say 100) of trials. Clearly, the approach lacks applicability to many real world problems. The alternative is to assign belief values to the event, called the subjectivist or Bayesian\textsuperscript{4} approach, “according to which probabilities encode degrees of belief about events in the world and data are used to strengthen, update or weaken those degrees of belief. In this formalism, degrees of belief are assigned to propositions (sentences that take on true or false values) in some language, and those degrees of belief are combined and manipulated according to the rules of probability calculus.” (Pearl, 2000 (p.2)).

Two characteristics of belief values are worth noting. First, such values are assigned based on some degree of belief the experimenter has in the occurrence of the particular event and not on the frequency of the same. Second, they are governed by the laws of probability. Using the second characteristic, experimenters change the degree of belief assigned to different variables so that the assigned belief values are consistent with the laws of probability. The next section explains the basic axioms of probability theory. Subsequent sections build on the same laws and explain their application to BN.

2.5.2 Axioms of Probability

Probability calculus defines three basic axioms:

i. Probability of variable \( A \) being in state \( a_i \) (denoted by \( P(A=a_i) \))\textsuperscript{5} is a number between 0 and 1. Thus,

\textsuperscript{4} After Reverend Thomas Bayes (more detail here)

\textsuperscript{5} As an example, consider the variable \( A \) that represents the probability of the car being of a certain color, then the possible states of \( A \) are the possible colors of the car. If the possible colors (states) are red, blue, green and white, then in order for the states to be mutually exclusive and exhaustive, the car has to have one color and can never have more than one color.
$0 \leq P(A = a_i) \leq 1 \quad \text{--2.9}$

and $\sum_i P(A = a_i) = 1$

ii. $P(A = a_i) = 1$ if and only if $a_i$ is certain.

iii. If $A$ and $B$ are two *mutually exclusive* events, then the probability that either one or the other will occur is the sum of their individual probabilities,

$$P(A=a_i \text{ or } B=b_i) = P(A=a_i) + P(B=b_i) \quad \text{--2.10}$$

This is known as the *additive rule* or the *theorem on the addition of probabilities*.

### 2.5.3 Law of Total Probability

In contrast with the additive rule, *joint probability* of two events $A$ and $B$ is the probability that both $A$ and $B$ will be in a given state at the same time. Joint probability of *independent events* is the *product* of their individual probabilities. Therefore,

$$P(A = a_j, B = b_j) = P(A = a_j)P(B = b_j) \quad \text{--2.11}$$

Generalizing, probability of $n$ independent events, $E_1$ through $E_n$ is given by

$$P(E_1, \ldots, E_n) = \prod_{i=1}^{n} P(E_i) \quad \text{-- 2.12}$$

For example, $(A=a_j, B=b_i)$ gives the *joint probability* of $A=a_j$ and $B=b_i$ (i.e., the probability of $A$ being in state $a_j$ and $B$ in state $b_i$ simultaneously). $P(A=a_j)$ and $P(B=b_i)$ are known as the *marginal probabilities* of $a_j$ and $b_i$ respectively.

One implication of joint probability is the *law of total probability*. The law provides a method to compute the *marginal probability* of an event (say $P(A=a_j)$) given the *joint probability* by summing over all the states of the other variable. Mathematically,
\[ P(A = a_j) = \sum_{i=1}^{m} P(A = a_j, B = b_i) \]  

where \( m \) is the number of possible states of \( B \).

The operation of summing over all values of \( b_i \) is also called “marginalizing over \( B \)”. As an example, assume two fair coins, \( A \) and \( B \) and consider the question, “What is the probability of getting “heads” on \( A \)?” The question can be answered by marginalizing over \( B \) (i.e., summing over all the states of \( B \) in the joint probability of \( A \) and \( B \)). The joint probability gives the probability of \( A \) and \( B \) being in some given state concurrently.

Let \( h \) represent getting “heads” and \( t \) represent “tails”, so in this case, \( P(A = h, B = h) \) represents the joint probability of getting “heads” in \( A \) and \( B \) simultaneously and \( P(A = h, B = t) \) gives the joint probability of “heads” in \( A \) and “tails” in \( B \). Adding the two joint probabilities yields \( P(A = h) \),

\[ P(A = h) = P(A = h, B = h) + P(A = h, B = t) \]

Or \( P(A = h) = \sum_{i} P(A = h, B = b_i) \)

where \( b_i \) are the possible states of \( B \), in this case \( h \) and \( t \).

From the above example it is clear that the marginal probability of any variable can be computed by marginalizing (summing over all its states) over the variable in the joint probability. This law forms the basis of probabilistic inference.

A corollary of this law is that the sum of the joint probabilities over all states of all variables is 1,

\[ \sum_{j=1}^{n} \sum_{i=1}^{m} P(A = a_j, B = b_i) = 1 \]  

where \( n \) and \( m \) are the number of states of \( A \) and \( B \) respectively.
2.5.4 Conditional Probability

Two other concepts of probability calculus that play an important role in BN are \textit{conditional probability} and \textit{conditional independence} of variables. This section deals with the first of the two: Conditional probability gives the probability of an event given the probability of another event.

\[ P(A=a | B=b) = x \] \hspace{1cm} -- 2.15

is read as the probability of \( A=a \) given the event \( B=b \) is \( x \). Traditionally, conditional probability is defined in terms of joint probability as

\[ P(A | B) = \frac{P(A, B)}{P(B)} \] \hspace{1cm} -- 2.16

The following example will be used throughout this section to explain application of relationships to a real problem. The problem statement is to compute the probability of the grass being wet \((W)\) given the states of rain \((R)\) and cloudiness \((C)\). So the probability of rain given cloudiness is

\[ P(R | C) = \frac{P(R, C)}{P(C)} \] \hspace{1cm} -- 2.17

This example will be referred to as the “wet grass” example in subsequent sections.

2.5.5 Chain Rule

According to Pearl (2000), the Bayesian approach to probability deems conditional probability as a more basic relationship between variables than joint probability as it is closer to the organization of human knowledge. Accordingly, equation (2.13) can be written so as to compute the \textit{joint probability} in terms of the \textit{conditional probabilities}.

\[ P(A, B) = P(A | B)P(B) \] \hspace{1cm} -- 2.18
The above equation denotes the product over all the possible states of $A$ and $B$. That is, if $A$ and $B$ are dichotomous variables with states 0 and 1, the joint probability $P(A,B)$ is given by

$$P(A,B) = P(A = 0 | B = 0)P(A = 0 | B = 1)P(A = 1 | B = 0)P(A = 1 | B = 1)P(B = 0)P(B = 1)$$

-- 2.19

This is known as the chain rule of probability and plays a critical role in BN. Generalizing the chain rule to a set of $n$ events $E_1,...,E_n$ their joint probability is given by the product of the conditional probabilities of each of the variables.

$$P(E_1,E_2,...,E_n) = P(E_n | E_{n-1},...,E_2,E_1)...P(E_2 | E_1)P(E_1)$$

or $$P(E_1,E_2,...,E_n) = \prod_{j=1}^{n} P(E_j | E_{j-1},...,E_2,E_1)$$

--2.20

The chain rule provides a method to compute the joint probability of all the variables represented in a graphical model (BN). As an illustration, consider a variable representing the state of a sprinkler ($S$) added to the wet grass example of the previous section. The joint probability of all the variables is denoted as

$$P(W,S,R,C) = P(W | S,R,C) P(S | R,C) P(R | C) P(C)$$

### 2.5.6 Bayes’ Theorem

Bayes’ theorem provides method of updating belief in hypothesis $B$ in light of evidence $A$. Elaborating, equation (2.16) yields the following

$$P(A,B) = P(A | B)P(B) = P(B | A)P(A)$$

Rearranging these terms leads to the inversion formula, which is the heart of Bayesian inference.
\[ P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \]  \[ \text{-- 2.21} \]

\( P(A \mid B) \) is known as the posterior probability or posterior belief (Bayesian theorists prefer using the term belief values rather than actual probability values to highlight the notion that the values are not traditional probabilistic values) and is the product of the prior belief \( P(A) \) and the likelihood of \( B \) given \( A \) (\( P(B \mid A) \)) is the likelihood that \( B \) will occur given that \( A \) is true) divided by a normalizing constant \( P(B) \) can be obtained by marginalizing as in equation (2.11)). The inversion formula (equation 2.19) and the chain rule (equation 2.16) make it possible to compute the conditional probability of any variable in a set of variables if the joint probability over the variables is known.

### 2.5.7 Conditional Independence

One problem with computing the joint probability distribution is the number of computations required. As equation (2.17) illustrates, the number of terms required for computing the joint probability increases exponentially with the number of variables; conditional independence offers to reduce the number terms in many cases.

For two variables, \( A \) and \( B \), if knowing the state of \( B \) does not affect the probability of \( A \), then \( A \) and \( B \) are said to be independent of each other; mathematically

\[ P(A \mid B) = P(A) \]  \[ \text{-- 2.22} \]

Similarly, \( P(A \mid B, C) = P(A \mid C) \) implies that \( A \) and \( B \) are conditionally independent of each other given \( C \); that is, once \( C \) is known, knowing \( B \) does not change the belief value of \( A \). For example, in the wet grass example, the state of wetness of the grass is conditionally independent of cloudiness given rain (i.e., given that it is raining, the probability of cloudiness will not have any effect on the state of wetness of the grass).
Generalizing the relationship for a set \( E \) of \( n \) events \( \{E_1, \ldots, E_n\} \), say there exists a set \( \text{pa}_j(E) \) consisting of variables \( E_j \) is conditionally dependent upon, such that

\[
P(E) = P(E_1, E_2, \ldots, E_n) = \prod_{j=1}^{n} P(E_j \mid E_{j-1}, \ldots, E_1) \quad \text{(from equation (2.20))}
\]

and,

\[
P(E_j \mid E_{j-1}, \ldots, E_1) = P(E_j \mid \text{pa}_j(E)) \quad \text{-- 2.23}
\]

Clearly, if the number of elements in \( \text{pa}_j(E) \) is less than \( n \), it results in considerable reduction in terms of computation required.

### 2.5.8 Graphical Notation

A brief review of graphical notation and concepts is necessary before further discussion into BN. A graph \( G(V,E) \) is a function of the set of vertices \( V \) and a set of edges \( E \). Each edge is denoted by a pair of vertices \((A, B)\) is an edge connecting the vertices \( A \) and \( B \) and may be directional (denoted by an arrowhead on one end) or unidirectional (also called bi-directional, denoted by arrows on both ends or no arrows at all). For bi-directional edges, \((A, B) = (B, A)\) while the same is not true for directional edges, which are denoted by an ordered pair of vertices, the first one denoting the originating vertex for the edge and the second one denoting the terminating vertex. So \((A, B)\) denotes an edge starting at \( A \) and ending at \( B \) and \((B, A)\) denotes one from \( B \) to \( A \) and \((A, B) \neq (B, A)\). The vertex of origin is known as the parent of the ending vertex, which is known as the child of the parent. A vertex may have multiple parents or multiple children. Parents of parent nodes are known as ancestors; likewise all vertices descending from any given node are known as its descendants. Two vertices with an edge between them are called adjacent vertices. A path in a graph is a sequence of “connected” edges.
(the terminating vertex of one edge is the originating vertex for the next, e.g.,

\((A, B), (B, W), (W, E), (E, Z)\)). A directed path has all the edges pointing in the same
direction. Two edges in a graph are said to be connected if there exists a path between the
two edges (in the path listed in the previous example, \(A\) and \(Z\) are connected), else they
are disconnected. A path is called a cycle if it begins and ends on the same node (e.g.,
\(X \rightarrow Y \rightarrow X\), distinct from self-loops, e.g., \(X \rightarrow X\)). A graph is cyclic if it contains at least one
cycle. The same definition of cycles applies to both directed and undirected graphs. A
directed graph that does not contain any cycles is called a directed acyclic graph (DAG).

### 2.5.9 Causal Networks and d-separation

A causal network provides a method for reasoning under uncertainty by
constructing a graphical model that represents causal relationships between events
(Jensen, 2001). A directed graph represents the causal relationships, each elementary
variable in the problem forms a node (vertex) of the network and an edge from \(A\) to \(B\) can
be thought of as “\(A\) causes \(B\)”. This property makes deciding the structure of graphical
models more intuitive as compared to other modeling techniques.

The d-separation criterion decides how evidence (probability of belief information)
is blocked from being transmitted from one node to another (Jensen, 2001) in a causal
network. Essentially, there are three types of connections in any directed graph; serial,
diverging and converging as illustrated in Figure 2.10 (a), (b) and (c) respectively.
Intuitively, it is easy to see that information can flow in serial (e.g., \(A = \text{Cloudy}, V = \text{Rain},
B = \text{Wet}\)) and diverging (\(A = \text{hair-length}, V = \text{Sex}, B = \text{Stature}\) (Jensen, 2001)) connections
unless vertex \(V\) is instantiated (has received some evidence). In situations depicted in
Figure 2.10 information can flow from \(A\) to \(B\) if and only if \(V\) is not instantiated.
Specifically, in Figure 2.10 (a), knowing it is raining or not will negate any effect of cloudiness on the state of wetness of grass and in Figure 2.10 (b) knowing the sex of a person nullifies any effect knowledge of stature may have on hair-length and vice versa.

For converging connections however (e.g., \( A = \text{Rain} \), \( B = \text{Sprinkler} \), \( V = \text{Pavement Wet} \) and \( D = \text{Pavement slippery} \) from Figure 2.10 (c)), information can flow only if the diverging node or \textit{any of its descendants} is instantiated. For the situation in Figure 2.10 (c), information can travel from \( A \) to \( B \) if and only if \( V \) or \( D \) is instantiated. (Jensen, 2001).

2.5.10 Bayesian Networks

A \textit{probabilistic model} encodes information that allows computing probability of any number of variables (\textit{propositions}) represented in the model connected using Boolean operators. As an example, the probabilistic model of the variables \( A, B \) and \( C \) should allow computing the probability of all statements like \( A \) \textit{and} \( B \), \( (A \text{ and } B) \) \textit{or} \( \neg C \), and so on. Each such term, formed by connecting the propositions using Boolean operators, is called a \textit{well-formed sentence} (Pearl, 2000) and is denoted by \( S \).

Any \textit{joint distribution function} (JDF) represents a \textit{complete probabilistic model} over the variables, since it allows computing the probability of \textit{any} well-formed sentence. The JDF is computed using the \textit{additive rule} (equation 2.10) and using the following two properties; (a) that every Boolean formula can be represented as a \textit{disjunction} of elementary events\(^6\) (Pearl, 2000), and (b) that elementary events are, by definition, mutually exclusive. \textit{Conditional probabilities} can be computed in a similar manner using equation (2.16).

\(^6\) For example, \( A \) and \( B \) is the same as not (not \( A \) or not \( B \)). This is obtained from DeMorgan’s theorem, which states \([\neg (A \text{ or } B) = A \text{ and } B]\) and \([\neg (A \text{ and } B) = (\neg A) \text{ or } (\neg B)]\).
By checking if sufficient information is available to compute the probability of every elementary event in the function domain and if the probabilities sum up to 1, the JDF can determine if sufficient information is available to specify a complete probabilistic model and verify whether or not it is consistent with the data. If the model is inconsistent, it specifies the additional information required and the points at which it is required (Pearl, 2000). From the above discussion, it should be clear that deriving the correct JDF for a given set of variables allows computing the probability of any well-formed sentence (i.e., any combination of the variables). For most practical problems, specifying joint probability functions depends on the problem domain; for continuous variables, they are specified as algebraic expressions (like the normal distribution, exponential distribution, etc.) while various indirect representation methods have been developed for problems involving discrete variables. Graphical models are the most promising of such indirect representations (Pearl, 2000).
Figure 2.10 d-separation in (a) serial, (b) diverging and (c) converging connections.

Pearl (2000, p. 13) lists three advantages of using graphs in probabilistic and statistical modeling:

1. they provide convenient means of expressing substantive assumptions;
2. they facilitate economical representation of joint probability functions; and
3. they facilitate efficient inferences from observations.

The second advantage is the most critical; as equation (2.17) illustrates, computing the joint probability of two dichotomous variables requires four terms. Extending this, computing the same for $n$ variables requires $2^n$ terms, making the entire calculation very expensive computationally. Clearly, computational complexity can be reduced if each term depends only on a small subset of the total terms. Graphical models help achieve just such an economy.

Graphical models that use undirected graphs are referred to as *Markov models* (Pearl, 1988) and are used mainly to represent symmetrical spatial relationships (Isham, 1981; Cox & Wermuth, 1996; Lauritzen, 1996). *Directed graphical models*, on the other hand, employ directed graphs and are used to represent *causal or temporal relationships* (Lauritzen, 1982; Wermuth & Lauritzen, 1983; Kiiveri et al, 1984) and are known as *Bayesian networks*. “[BN are] so named to emphasize three aspects: (1) the subjective nature of the input information; (2) the reliance on Bayes’s conditioning as the basis for updating information and; (3) a distinction between causal and evidential modes of reasoning.” (Pearl, 2000, p. 14). More formally, BN are *graphical models* with the following properties (Jensen, 2001):

1. The graph of the models consists of a set of *variables* and a set of *directed edges* between variables. The variables form the nodes of the graph and are connected by directed edges. Each variable has a finite set of mutually exclusive states.
2. The variables together with the directed edges form a *directed acyclic graph* (DAG).

3. A conditional probability table (CPT) is attached with each variable. The CPT gives the conditional probability of the variable given its parents. In a BN, conditional independence is implied by the structure of the network. Conditional independence relationships are encoded in the structure of a BN such that a node is independent of its ancestors given its parents. So for a variable $A$ with parents $B_1, \ldots, B_n$, the CPT is $P(A | B_1, \ldots, B_n)$ and the node is conditionally independent of any parents of $B_i$.

Marginal probability of any variable in a BN can be computed by marginalizing over the joint probability distribution of all the variables in the network. The conditional independence relationships encoded in the structure of the network allow some savings in the number of computations required to compute the JDF (according to equation 2.23).

Two unknowns must be resolved in order to successfully compute the JDF and thereby the marginal probabilities, namely the *structure* of the network and the values of the *parameters* (or *conditional probabilities*). Structure refers to the edges of the network and how the various nodes (variables) are connected to each other. Structure encodes conditional independence relationships between the variables; lack of an edge between $A$ and $B$ denotes $A$ is independent of $B$\(^7\). Specifically, BN encode conditional independence relationships such that a node is independent of its ancestors given its parents. Recall conditional independence from equation (2.23). In this case, $pa_j(E)$ is consists of the parents of node $E_j$. As an example, consider the network shown below (Figure 2.11), the node $E_6$ is independent of $E_1$ and $E_2$ given its parents $E_3$, $E_4$ and $E_5$. The d-separation criterion translates to relationships of conditional independence in BN\(^8\).

---

\(^7\) Causality is used in devising the structure of the network and makes it very intuitive. An arrow from $A$ to $B$ can be thought of as $A$ causes $B$ and also represents conditional dependence between the two.

\(^8\) The definition of BN does not make any reference to causality and the edges do not necessarily represent causal impact, instead the definition requires for the *d-separation* properties implied by the network.
One point to bear in mind while designing the structure of a BN is that the aim of the BN is to give certainties of events, called *hypothesis events* that are not directly observable or are observable at a very high cost. These events are grouped into mutually exclusive groups and are called *hypothesis variables*. Variables that are directly observable and provide information regarding the hypothesis variables are called *information variables*.

Parameters of a BN are conditional probabilities of each of the variables in the network. Conditional probabilities are represented using a structure known as the

![Figure 2.11. Illustration of conditional independence relationships.](image)

*Conditional Probability Table* (CPT), which is essentially a table listing the conditional probabilities for each variable given the rest of the variables in the network. A BN can “learn” the structure from data when the number of variables used is very large structure to hold (Jensen, 2002). However, using causality has proved to be a good method of arriving at the structure of the BN (Murphy, 2000).
and a substantial amount of data is available. When the structure is known, as in the current study, CPTs are initially specified by an expert (representing the belief of the expert) and are subsequently corrected to make the network conform to real world data. This essentially means that the initial values in the CPTs should be reasonably close or they should at least conform to the same pattern as the actual values (i.e., one should not specify the probability of rain as 0.0 when the probability of cloudiness is 0.96). Determining the correct structure for the network and values of the parameters is referred to as learning in BN context. The conditional probability relationships between two variables can be thought of as the strength of connection in BN. Once the structure and parameters have been learned, the network can be used for probabilistic inference, that is to infer the state of one variable given a change in the state of others (or to find the states of variables for given uncertain state values for other variables), or to explain away, that is, to find the probability of the alternative “cause” given the probability of one cause. Using BN for inference is more common of the two.

Inference in BN can be exact or approximate, which means that the values of variables are computed to conform exactly to the laws of probability or to obtain an approximate value of the same. The first is suited to conditions with a small number of variables with a limited number of states so that computing the values of the variables is not computationally intractable. On the other hand, when the number of variables is large, approximate inference is more practical. The current study involves only five dichotomous variables, and therefore implements exact inference.

2.5.11 An Example of a Bayesian Network

This section describes the steps involved to build a BN model of the wet grass example. The aim is to develop a model to compute the probability of the grass being wet
given the other variables. The first step in building the structure is to identify hypothesis and information variables and determine the relationships between them. This problem has four variables with two possible states for each, true (T or 1) and false (F or 0). The variables are cloudy (C), sprinkler (S), rain (R) and wet grass (W).

The network for the problem is depicted in Figure 2.12. Relationships among variables were determined using causality: Clouds lead to (cause) rain and not the other way around and they also determine whether the sprinkler is to be turned on or off. Clearly, it would be imprudent to suggest the sprinkler affects the state of cloudiness. The figure also shows the CPT for each variable in the network. Conditional probabilities can be set by analysis of existing data (e.g., to determine the probability of rain given the probability of clouds from meteorological data) or they can be subjective belief values set by an expert as explained above. The example and the current study use the latter. Subjective values can be changed so as to make them consistent (according to the rules of probability) with observed data.

The JDF must be computed before the model can be used for any predictions. The JDF for the current problem is given by,

\[
P(C, S, R, W) = P(C) \cdot P(S \mid C) \cdot P(R \mid C) \cdot P(W \mid C, S, R)
\]

The above equation can be simplified using conditional independence between cloudiness and wet grass given rain and sprinkler, that is

\[
P(C, S, R, W) = P(C) \cdot P(S \mid C) \cdot P(R \mid C) \cdot P(W \mid S, R)
\]

For this example, the observed variable is that the grass is wet \((W=1)\) and the problem is to compute which of the two causes, rain or sprinkler, is the more likely in this case (i.e., to compute and compare the conditional probabilities of rain given wet grass
According to Bayes’ theorem
\[
P(S = 1 | W = 1) = \frac{P(W = 1 | S = 1)P(S = 1)}{P(W = 1)}
\]

Note the numerator term in the above equation is the joint probability of \(W=1\) and \(S=1\), that is

Figure 2.12 A Sample Bayesian network to determine model the probability of the grass being wet given states of cloudiness (C), Rain (R) and Sprinkler (S).
\[ P(W = 1 \mid S = 1)P(S = 1) = P(S = 1, W = 1) \]

Therefore equation (2.26) can be rewritten as

\[ P(S = 1 \mid W = 1) = \frac{P(S = 1, W = 1)}{P(W = 1)} \quad \text{--2.27} \]

Both the numerator and denominator terms are obtained by marginalizing over appropriate variables in the joint probability distribution of the entire model. So

\[ P(S = 1, W = 1) = \sum_{c, r} P(C = c, S = 1, R = r, W = 1) \quad \text{--2.28} \]

and,

\[ P(W = 1) = \sum_{c, s, r} P(C = c, S = s, R = r, W = 1) \quad \text{--2.29} \]

In case of the current network, equation 2.28 is expanded as

\[ P(S = 1, W = 1) = P(C = 0, S = 1, R = 0, W = 1) + \\
\quad P(C = 0, S = 1, R = 1, W = 1) + \\
\quad P(C = 1, S = 1, R = 0, W = 1) + \\
\quad P(C = 1, S = 1, R = 1, W = 1) \quad \text{--2.30} \]

The first term in equation 2.30 can be solved according to the JDF equation for the network (equation 2.25), therefore

\[ P(C = 0, S = 1, R = 0, W = 1) = P(C = 0)P(S = 1 \mid C = 0)P(R = 0 \mid C = 0)P(W = 1 \mid S = 1, R = 0) \quad \text{--2.31} \]

All the values required to compute the above equation are listed as part of the CPT in Figure 2.12. Plugging the values in the equation,

\[ P(C = 0, S = 1, R = 0, W = 1) = (0.5)(0.1)(0.8)(0.9) = 0.036 \quad \text{--2.32} \]

Plugging in the values, equation 2.27 yields,

\[ P(S = 1 \mid W = 1) = \frac{P(S = 1, W = 1)}{P(W = 1)} = \frac{0.2781}{0.6741} = 0.4126 \quad \text{--2.33} \]
Similarly the equation for rain given wet grass results

\[
P(R = 1 | W = 1) = \frac{P(R = 1, W = 1)}{P(W = 1)} = \frac{0.4581}{0.6741} = 0.6796
\]

where the numerator is,

\[
P(R = 1, W = 1) = 0.4581
\]

while the denominator term is the same from equation 2.29,

\[
P(W = 1) = 0.6741
\]

Clearly, as seen from the results of equations 2.33 and 2.34, the sprinkler is the more likely cause of the grass being wet under the given conditions. For an illustration of explaining away, consider the probability of the sprinkler being on when it is known that the grass is wet and that it is raining. Knowing grass is wet instantiates the converging connection and allows information to pass through from one variable to another. As a result, knowing grass is wet and that is raining, should result in a decrease in the probability of the sprinkler being on. This is exactly what happens, with

\[
P(S = 1 | W = 1, R = 1) = 0.1945
\]

as opposed to 0.5 when only the state of cloudiness is known.

### 2.6 Summary

Research on attentional bias has been focused at understanding the causes and the underlying mechanisms that drive these biases. An overwhelming majority of research has focused on understanding the causes of attentional bias, whereas investigating mechanisms has been largely neglected. Various cognitive theories emphasize that underlying mechanisms hold at least the same import in understanding attentional biases as their causes (e.g., Williams et al., 1997). Study of the mechanisms of attentional bias using paradigms such as the Stroop task and the dot probe task is proceeding relative
slowly (compared to advancing of knowledge about the causes), primarily because these paradigms offer only an indirect method to study these mechanisms of attentional biases.

A direct method of studying these mechanisms is by developing computer models. Computer simulation of attentional paradigm provides an attractive method of investigating underlying mechanisms due to the increase in computational power. Computational power has increased roughly by a factor of 10 over the past eight years, which incidentally is the time since the last study involving simulation of an attentional paradigm (i.e., Matthews and Harley (1996)). Ironically, although it provides a direct measure of attentional bias, the dot probe paradigm has not been simulated. At the same time, research with the paradigm has generated a very large database and robust relationships between the variables of the task. A computer model can potentially utilize the data and information available about the relationships between variables to possibly provide new insights into the mechanisms of attentional bias from a dot probe perspective.

The current study aims to address these issues by constructing a NN simulation of the dot probe task. One important objective of the study is also to present BN as a viable and easy to use method of modeling and analyzing data. The study will use established relationships to devise the models and verify (or challenge) the models using existing data; leading to questions about relationships and opening avenues of new research in the area.
This chapter presents details regarding the design, training, and testing of two dot probe models with the Neural and Bayesian network models, respectively. Development and testing specifics will be described for each version of the dot probe that will be simulated, followed by explanations of the two models.

3.1 The Task

The standard version of the dot probe task will be simulated, using pictures as the emotional stimuli (Mogg & Bradley, 1999). The task consists of a pair neutral images, or a neutral image paired with a negative image, displayed side-by-side for 500 ms. Stimulus offset is followed by appearance of a dot probe stimulus in the spatial location of one of the pictures. Participants are instructed to press a button to indicate the side on which the dot appears. The probe remains on the screen for 1100 ms or until the user presses a button to indicate its position.

3.2 Neural Network Model of the Dot Probe task

The purpose of the neural network model of the dot probe task will be to compare three different potential causal mechanisms of attentional bias in the dot probe task. The three mechanisms are:

7. **Interaction**, based on the model by Williams et al. (1988), which states that attentional bias is a product of the state and trait anxiety levels of an individual and that high anxiety individuals orient towards threatening stimuli and that low anxiety individuals orient away from them.

8. **Exposure**, based on the premise that all individuals (both high and low anxiety) initially respond in the same manner to both threat and neutral stimuli but that high anxiety individuals are exposed to threat stimuli more than low anxiety individuals which strengthens the connections for negative stimuli.

9. **Intensity**. The third one essentially states that attentional bias occurs because high anxiety individuals perceive a stimulus as having a higher threat value at that moment than individuals with low levels of trait anxiety.
Two simulations were performed; results for the first are presented in terms of reaction time computed from the number of iterations required to obtain the output. Results for Simulation 2 are presented as activation levels of the output units after one iteration.

Each mechanism trained the network to be more aroused (activated) for negative stimuli as opposed to neutral ones. The mechanisms used the same network, but differed in the training and test patterns presented, as well as the parameters used for training and testing. Results of all three training patterns are compared against the results of the baseline condition. The baseline condition assumed that (1) high and low anxiety individuals respond the same way to all kinds of stimuli, and (2) that both these groups experience a higher level of arousal with for negative stimuli than neutral ones. The details of the mechanisms and the training procedure for each are explained below.

According to the initial proposal, the network was to be trained for 400 epochs. This requirement was revised to training the network until its mean squared error for that mechanism was less than 0.0001 or the patterns had been presented to the network 500 times (i.e., 500 iterations). Weights were adjusted online, or at the end of each cycle (that is, after all patterns listed in the table have been presented to the network once).

### 3.2.1 Mechanisms

The network was trained for the negative and neutral conditions in isolation. In other words, training input to the network specified the response to a given picture displayed and anxiety level. Training sets for a given mechanism consisted of a fixed number of patterns of each type. The patterns and the frequency with which they were presented to the network are listed in Table 3.1. The mechanism for attention selection,
called the *attention mechanism*, was encoded into the network by training the network more at least twice as extensively on patterns representing attention to the side of the dot. These patterns are listed in Table 3.2.

Table 3.1. Training patterns used to train the NN for baseline, exposure 3x and exposure 5x conditions

<table>
<thead>
<tr>
<th>Probe Side (Left/Right)</th>
<th>Output (Left/Right)</th>
<th>Number of patterns (baseline)</th>
<th>Number of patterns (exposure 3x)</th>
<th>Number of patterns (exposure 5x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>Left</td>
<td>10</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>Right</td>
<td>Right</td>
<td>10</td>
<td>30</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 3.2. Training patterns for the attention mechanism for baseline, exposure 3x and exposure 5x conditions.

<table>
<thead>
<tr>
<th>Negative picture side (Left/Right)</th>
<th>Neutral Picture Side (Left/Right)</th>
<th>Trait Anxiety (High/Low)</th>
<th>Output (Left/Right)</th>
<th>Number of patterns (baseline)</th>
<th>Number of patterns (exposure 3x)</th>
<th>Number of patterns (exposure 5x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>-</td>
<td>High</td>
<td>Left</td>
<td>5</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>Right</td>
<td>-</td>
<td>High</td>
<td>Right</td>
<td>5</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>Left</td>
<td>-</td>
<td>Low</td>
<td>Left</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Right</td>
<td>-</td>
<td>Low</td>
<td>Right</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Left</td>
<td>Left</td>
<td>High</td>
<td>Left</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Right</td>
<td>Right</td>
<td>High</td>
<td>Right</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Left</td>
<td>Left</td>
<td>Low</td>
<td>Left</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Right</td>
<td>Right</td>
<td>Low</td>
<td>Right</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

The number of patterns training the network to attend to the side of the dot was twice the number of patterns training the network to attend to the negative image in each condition.

---

1 This is the mechanism that simulates deliberate human attention towards to attend and respond to the side on which the dot appears.
3.2.1.1 Baseline condition

The baseline condition was trained using the patterns in Table 3.1. The training assumed that threatening stimuli cause a slightly higher activation than neutral stimuli and so each epoch consists of five patterns presenting a negative stimulus against four patterns for neutral stimuli.

3.2.1.2 Exposure mechanism.

The exposure mechanism investigated the possibility of etiology of attentional bias due to repeated exposure of a high anxiety individual to threat stimuli. Training the network more extensively on certain patterns results in strengthening the connections to produce output consistent with those patterns. Cohen et al. (1990) and Matthews and Harley (1996) used this property of NN to train their networks to exhibit a higher degree of automaticity for the word-reading task than for color naming, presenting 10 times and 8 as many word reading patterns as color naming patterns, respectively. As listed in Table 3.1, the current model used the same technique to make the NN more sensitive to negative stimuli for high levels of trait anxiety and neutral stimuli for low levels of trait anxiety.

Two exposure mechanisms were simulated, Exposure 3x that trained the network on 3 times as many negative stimuli under high anxiety as the baseline condition, and Exposure 5x that trained the network on 5 times as many negative stimuli as the baseline mechanism. The training patterns for the two mechanisms are listed in Table 3.1. The number of times each pattern was presented to the network is listed in the last column of Table 3.1.
The number of training patterns for the attentional mechanism were also increased to offset the increased activation of negative stimuli under high anxiety. These training patterns are listed in Table 3.2.

### 3.2.1.3 Interaction mechanism.

The interaction mechanism trained the network to produce results according to the interaction hypothesis, requiring the network to exhibit the following behavior:

1. Attend toward the negative picture for high levels of trait anxiety and respond faster when the dot replaces the negative picture.
2. Attend toward from the neutral picture for low levels of trait anxiety and respond faster when the dot replaces the neutral picture.
3. Produce an output if and only if one of the probe units is activated (to mimic responding to the probe), and the output of the network should be the same as side on which the dot appears.

Alternatively, the first two requirements can be viewed from the reverse perspective as:

1. Attend away the neutral picture for high levels of trait anxiety and respond slower when the dot replaces the neutral picture.
2. Attend away from the negative picture for low levels of trait anxiety and respond slower when the dot replaces the negative picture.

Training patterns for the interaction mechanism trained the network to specifically attend towards the negative stimulus (and away from the neutral stimulus) under high anxiety and towards the neutral stimulus (and so away from the negative stimulus) under low anxiety. Additionally, the condition assumed that both negative and neutral images resulted in similar activation of the network. This was programmed into the network by presenting the network with an equal number of patterns for negative and neutral stimuli. The training patterns are listed in Table 3.3.
3.2.1.4 Intensity condition

The intensity condition was intended simulate performance under both trait and state anxiety in the NN. The mechanism is based on the assumption that levels of trait and state anxiety regulate the perceived intensity of the stimulus for the individual.

Table 3.3. Training patterns used to train the network for the interaction mechanism.

<table>
<thead>
<tr>
<th>Negative picture side (Left/Right)</th>
<th>Neutral Picture Side (Left/Right)</th>
<th>Trait Anxiety (High/Low)</th>
<th>Output (Left/Right)</th>
<th>Number of patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>-</td>
<td>High</td>
<td>Left</td>
<td>4</td>
</tr>
<tr>
<td>Right</td>
<td>-</td>
<td>High</td>
<td>Right</td>
<td>4</td>
</tr>
<tr>
<td>Left</td>
<td>-</td>
<td>Low</td>
<td>Right</td>
<td>4</td>
</tr>
<tr>
<td>Right</td>
<td>-</td>
<td>Low</td>
<td>Left</td>
<td>4</td>
</tr>
<tr>
<td>-</td>
<td>Left</td>
<td>High</td>
<td>Right</td>
<td>4</td>
</tr>
<tr>
<td>-</td>
<td>Right</td>
<td>High</td>
<td>Left</td>
<td>4</td>
</tr>
<tr>
<td>-</td>
<td>Left</td>
<td>Low</td>
<td>Left</td>
<td>4</td>
</tr>
<tr>
<td>-</td>
<td>Right</td>
<td>Low</td>
<td>Right</td>
<td>4</td>
</tr>
</tbody>
</table>

The mechanism essentially was proposed to be simulated by applying a supernormal input (activation of 8.0 opposed 1.0) to the network. Trait anxiety in this condition was to be simulated by applying the supernormal inputs during training while testing with normal inputs. On the other hand, state anxiety was to be simulated by testing the network with supernormal inputs for the negative image applied to the network trained by the baseline mechanism. The network could not be successfully trained to simulate trait anxiety but results were obtained for the state anxiety condition while two different state anxiety conditions were tested, using an activation value of 4.0 and 8.0 for the negative image under high anxiety respectively.
3.2.2 Simulations

Two different simulations were performed for each of the mechanisms; one presented the results in terms of the number of iterations required to compute the output while the other presented them as activations of the output units obtained in one iteration.

As one of the simulations listed above simulated the temporal characteristics of the dot probe and presented results in terms of RT, it required changes in computing the net input and output of the network. The following sections explain details of the network and further differences in the two simulations.

3.2.3 Structure

A two-layer fully connected backpropagation neural network depicted in Figure 3.1 will be used for all three simulations, in contrast to a partially connected BPN used in the two models of the Stroop task discussed earlier (Cohen et al., 1990; Matthews & Harley, 1994).

Input to the network will specify the type of image displayed on the left and the right, the anxiety level of the participant and the side on which the dot appears. The model will be built to handle only negative and neutral images. Two units each will be used to specify the characteristics (negative or neutral valence) of the picture on the left \( (I_1=\text{negative}, I_2=\text{neutral}) \) and right \( (I_7=\text{negative}, I_8=\text{neutral}) \) sides as depicted in Figure 3.1. Two units will specify the anxiety level \( (I_4 \text{ and } I_5 \text{ for high and low anxiety respectively}) \) of the individual and two units will be used to indicate the side on which the dot probe appears \( (I_3 \text{ and } I_6 \text{ for left and right sides, respectively}) \).

The present model will feature one hidden layer of five units, while the output layer will be characterized by two units, indicating the possible responses of “left” and “right”.
Activation of output unit $O_1$ will correspond with the participant pressing the button to indicate the dot appearing on the left side and $O_2$ will correspond to the right side.

### 3.2.4 Net Input

Simulations 1 and 3 will model the time-course of attention and present results in terms of RT. Simulation 2 will compare activation levels obtained in a single iteration and compare activation level and mean square error of the network for the input pattern. Therefore, net input at any step for simulations 1 and 3 will be a function of the running average of the net input of the preceding time steps, computed using equations 2.5 and 2.6. Net input for simulation 2 will be computed using the typical formula presented in equation 2.3. The three equations are listed below.$^2$

$$\tilde{x}_j(t) = \tau x_j(t) + (1 - \tau) \tilde{x}_j(t-1)$$  

---2.5

where

$$x_j(t) = \sum_i a_i(t) w_{ij} + b$$  

---2.6

$$x_j = \sum_i a_i w_{ij}^o + b$$  

---2.3

A normally distributed random bias will be added to the net input in all three simulations to mimic variability of performance, reminiscent of the Cohen model (Cohen et al., 1990).

### 3.2.5 Activation

Activation for the hidden and output units will be computed by applying the logistic sigmoid function of equation 2.3 to the net input computed in the previous step.

---

$^2$ The numbers of the equations are the same as used in Chapter 2.
Figure 3.1 Neural network model for simulating dot probe task.

3.2.6 Output

For all simulations, activation level of the unit will form the output of each hidden unit. Output of the output units (and therefore the network) will be computed on the basis of a random walk or diffusion process by adding a small amount of evidence according to equation 2.8 to each unit on each iteration. A threshold value of 1.0 will be set for these units such that they will be considered to be active only when the evidence of the unit exceeds 0.1. No threshold will be set for output of simulation 2 and the activation level of each unit will be considered as its output.
3.2.7 Initialization

Initialization of the network for all three simulations will involve assigning random weights between 0.0 and 1.0 to the connections in the network. These weights will be adjusted to yield the correct output for given input during the training phase.

3.2.8 Training

Training patterns for each of the simulations have already been explained in the previous section (Simulations).

3.2.9 Testing

The network will simulate presenting images to 10 high anxiety and 10 low anxiety individuals. Each participant will be presented with five trials of each possible combination of stimuli (Table 3.1). Results for Simulation 1 are presented in terms of RT computed by deriving relationships between the number of iterations required to compute the output and the typical output for the condition while the activation of the output units are compared for Simulation 2.

3.3 Bayesian Network

The first task in constructing the BN model of the dot probe task is to identify the hypothesis and information variables. Recall that hypothesis variables are the ones that either are not observable or observable only at a very high cost and information variables provide information about the hypothesis variables. The hypothesis and information variables to be modeled in the BN are chosen from among the independent and dependent variables of the dot probe task\(^3\) to be modeled. The primary independent variables are:

4. The valence and arousal level of the picture.
5. Participant’s level of trait anxiety.

\(^3\) The same version of the dot probe will be modeled by the BN as will be simulated by the NN.
6. The side on which the probe appears. 
   And the dependent variables are

1. The direction of attention.
7. The $RT$ in responding to the probe. $RT$ constitutes the fifth variable in the BN. 
   Independent variables are by definition, observable and therefore classified as 
   information variables. Of the two dependent variables, $RT$ is also an observed variable 
   and so is classified as an information variable. That leaves only the direction of attention 
   as the hypothesis variable. As direction of attention is precisely what the dot probe (and 
   other attentional paradigms) are used to infer, the BN provides an ideal modeling 
   framework.

   To model a variable in a BN, it must be discrete and have a finite number of states. 
   The BN representing a model of the five variables of the dot probe task is shown 
   in Figure 3.1. The number of states, their names, their possible values and the meaning of 
   each variable in a Bayesian and real world context are explained below.

8. Arousal Rating (AR): As variables should be discrete and have a finite number of 
   states, this node of the BN will represent the arousal rating of the image. The 
   “probability” of the variable was the normalized arousal rating of the displayed 
   image, thereby making the model independent of the rating tool used. Thus, it is 
   possible to represent either the IAPS rating of the picture or its SAM rating. The 
   current study used the former. Indeed the network could also prove to be a tool to 
   check how closely one rating matches the other. Valence of the picture will not 
   represented in the model, so it is assumed the arousal rating represents negative 
   image. Further, the neutral image in a critical trial is assumed to have an arousal 
   rating of 0.0 and so is ignored in the model.
   The variable had two possible states, negative = true (1.0) and negative = false 
   (0.0) (or neutral= true). A value of 1.0 denoted an image with the highest possible 
   negative rating (the specific highest value will depend on the arousal scale used). A 
   value of 0.75 means that the image is negative with probability 0.75 for the model 
   and the image has an arousal rating of 0.75 on the arousal scale in the dot probe 
   context.

9. Anxiety (Anx): This node denoted the level of trait anxiety of an individual. The 
   variable also had two possible states, high = true (1.0) and high=false (0.0) (or 
   low= true). The variable represented the probability of the individual having high 
   level of trait anxiety computed by normalizing the anxiety score of the individual
on an anxiety questionnaire. Different anxiety questionnaires could have been used to obtain the trait anxiety score but the current study will relied on the STAI-Trait.

10. **Attention Direction (AD):** This was the sole hypothesis variable (variable not directly observable) in the current Bayesian model; it is directly influenced by $AR$ and $Anx$. According to the cognitive model proposed by Williams et al. (1988), perception of the valence and arousal level of the stimulus coupled with level of trait anxiety decide the direction in which an individual will orient attention. Following the same line of reasoning, attention direction can also be interpreted as a measure of the strength of attentional bias. States of the variable reflect whether attention direction is towards or away from the negative image, as $towards = true$ (1.0) and $towards = false$ (0.0) (or $away = true$) from the negative image. Note that the connection between $AR$, $Anx$ and $AD$ is a converging connection. By the rules of d-separation, this means that information cannot travel from $AR$ to $Anx$ and vice-versa if either $AD$ or at least one of its descendants is instantiated (i.e., their state is known). For instance, if it is known that an individual is attending towards ($towards = true$) the negative image, information about anxiety ($Anx$) will not add to knowledge about the arousal level of the picture ($AR$).

11. **Probe Side (PS):** This variable simply denoted the side on which the dot appeared. Like $AD$, its possible states are either on the side of the negative picture ($PS = Neg [1.0]$) or on the opposite side as the negative picture ($PS = Ntrl [0.0]$). Unlike the other variables, values between 0.0 and 1.0 were not allowed to represent its states.

12. **Reaction Time (RT):** For this model, RT was assumed to have only two states, high and low. The two are denoted as $RT= low$ (1.0) and $RT= high$ (0.0). One of the issues the model was intended to was the existence of a linear relationship between the probability of obtaining a high RT and the actual RT. CPT for each variable are not listed in the figure. The CPT table for a variable indicates the probability of the variable being in one of the state given its parents. The crucial CPT required was that of $P(AD|AR,Anx)$, that is, the conditional probability of $AD$ given the probability of $AR$ and $Anx$. These values were computed from a study conducted at the Motor Behavior Laboratory at the University of Florida as follows:

13. **Outliers were eliminated (RT of < 200 ms and >1000 ms).**

14. **If the RT was greater one standard deviation (positive) of the mean, attention direction was assumed to be on the side of opposite of probe-side, else it was assumed to be on the side of the probe. Average of individuals attending towards and away from the negative picture was computed to form the CPT.**
Attention was assumed directed towards the negative image when RT was low for a probe replacing the negative image and high for an image replacing the neutral image. RT was said to be low if it was one SD deviation less than the mean and was said to be fast if it was more than one SD more than the mean. Average of these numbers were com

Figure 3.2 Bayesian network model of the dot probe task.

After establishing the CPTs for each of the variables, the model will was used to attempt to answer the following questions:

10. What are the probabilistic relationships between the variables? Specifically, the model was used to obtain the probability of an individual attending to the negative image given the anxiety level of the individual and the arousal rating of the picture displayed. In order to do this, $AD$ was set to 1 (Negative = 1, i.e., participant is attending towards the negative cue with probability 1.0), and the $AR$ was obtained for different levels of anxiety. This effectively yielded the arousal level for the given probability, required so that the individual attends to the negative cue with probability 1. Similarly, the relationship was investigated by recording the change in the value of anxiety due to changes in the arousal rating.
11. Is there a specific anxiety score that can be used as a threshold to high anxiety? This was ascertained from the pattern of change in the arousal level given the anxiety score.

12. What is the accuracy of the model? Determined by checking the results obtained from the model against results of the actual study.
CHAPTER 4
RESULTS

4.1 Neural Network Model

The neural network model was trained on three different conditions, namely the basic, exposure, and interaction, until either the network achieved a mean squared error below a preset value (0.0001) or exceeded a predetermined arbitrary number of iterations (500). The number of iterations that the network took to train for each of the conditions, and the final error after training, is listed in Table 4.1.

Two different sets of test patterns were used for the neural network (NN) model; a basic model, that only specified whether a given unit was activated or not, and a set of patterns specifically to test for the intensity condition. Test patterns for the intensity mechanism had elevated levels of activation for negative stimuli to simulate a higher perceived arousal level of the negative stimulus. As a result, anxiety for the intensity condition refers to state anxiety, while for all the other test cases it refers to trait anxiety. Basic test patterns are presented in Table 4.2 and the intensity test patterns are listed in Table 4.3.

Table 4.1. Number of training iterations and MSE for each training mechanism.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Number of training iterations</th>
<th>Mean Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>18</td>
<td>0.000049</td>
</tr>
<tr>
<td>Exposure 3x</td>
<td>24</td>
<td>0.000096</td>
</tr>
<tr>
<td>Exposure 5x</td>
<td>18</td>
<td>0.000060</td>
</tr>
<tr>
<td>Interaction</td>
<td>500</td>
<td>0.000956</td>
</tr>
</tbody>
</table>

Each column of Table 4.2 and Table 4.3 represents the activation presented to an input unit of the network. The first two columns represent information about the picture
appearing on the left side and the last two columns represent information about the picture on the right side. So, a negative picture appearing on the left is represented by setting the first input unit to 1.0. Similarly, the side on which the dot appears is indicated by the appropriate probe unit. Denoted by a “1” if activated, anxiety of the individual is encoded by activating the high or low anxiety units also designated by a “1” if active. The numbers in parentheses denote the index-number of each input unit. Patterns 1 through 4 represent the possible combinations for high anxiety; patterns 1 and 2 represent the match condition, that is, when the dot follows the negative picture. Specifically, pattern 1 represents match-left and pattern 2 represents match-right (left and right refer to the side of the dot). Patterns 3 and 4 denote the mismatch conditions, that is, conditions in which the dot replaces the neutral image. Specifically pattern 3 is called mismatch-right (because the dot probe appears on the right) and pattern 4 is called mismatch-left (because the dot probe appears on the left).

Patterns 5 through 8 encode the same information for a low anxiety individual and so differ from the first four patterns only in the activation of the anxiety unit. Consequently, the numbers 1 through 4 are used to refer to patterns 5 through 8 along with stating the anxiety level specifically. All results refer to these patterns.

The intensity mechanism, presented in Table 4.3, differs from the other conditions in an elevated level of activation of the negative stimulus is designated for high levels of state anxiety. Therefore, patterns 5 through 8 remain unchanged from the basic patterns and are not listed, while the activation level for the negative picture in patterns 1 through 4 is increased to 4.0.
The weights and biases for the network after training for the basic, exposure and interaction conditions are listed in the Appendix. Although it was initially proposed that training for the exposure condition would ‘expose’ the network to three times as many

Table 4.2. Basic test patterns for the neural network model

<table>
<thead>
<tr>
<th>Pattern Number</th>
<th>Left Pic. Negative (1)</th>
<th>Left Pic Neutral (2)</th>
<th>Left Probe (3)</th>
<th>High Anxiety (4)</th>
<th>Low Anxiety (5)</th>
<th>Right Probe (6)</th>
<th>Right Pic. Negative (7)</th>
<th>Right Pic. Neutral (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 match left</td>
<td>1 0 1 1 0</td>
<td>0</td>
<td>0</td>
<td>1 1</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2 match right</td>
<td>0 1 0 1 0</td>
<td>1</td>
<td>1</td>
<td></td>
<td>0 1</td>
<td>1</td>
<td>1 0</td>
<td>0</td>
</tr>
<tr>
<td>3 mismatch right</td>
<td>1 0 0 1 0</td>
<td>1</td>
<td>1</td>
<td></td>
<td>0 1</td>
<td>1</td>
<td>0 1</td>
<td>1</td>
</tr>
<tr>
<td>4 mismatch left</td>
<td>0 1 1 1 0</td>
<td>0</td>
<td>0</td>
<td>1 1</td>
<td></td>
<td>0</td>
<td>1 0</td>
<td>0</td>
</tr>
<tr>
<td>5 match left</td>
<td>1 0 1 0 1</td>
<td>0</td>
<td>0</td>
<td>1 1</td>
<td></td>
<td>0</td>
<td>0 1</td>
<td>1</td>
</tr>
<tr>
<td>6 match right</td>
<td>0 1 0 0 1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1 1</td>
<td>1</td>
<td>1 0</td>
<td>0</td>
</tr>
<tr>
<td>7 mismatch right</td>
<td>1 0 0 0 1</td>
<td>0</td>
<td>1</td>
<td></td>
<td>1 1</td>
<td>1</td>
<td>0 1</td>
<td>1</td>
</tr>
<tr>
<td>8 mismatch left</td>
<td>0 1 1 0 1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>0 1</td>
<td>0</td>
<td>1 0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.3. Test patterns for the intensity condition.

<table>
<thead>
<tr>
<th>Pattern Number</th>
<th>Left Pic. Negative (1)</th>
<th>Left Pic Neutral (2)</th>
<th>Left Probe (3)</th>
<th>High Anxiety (4)</th>
<th>Low Anxiety (5)</th>
<th>Right Probe (6)</th>
<th>Right Pic. Negative (7)</th>
<th>Right Pic. Neutral (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4 0 1 1 0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0 1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0 1 0 1 0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4 0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>4 0 0 1 0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4 1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0 1 1 1 0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0 1</td>
<td>0</td>
</tr>
</tbody>
</table>
presenting a high anxiety individual with five times as many negative stimuli. Final weights and biases for this condition are also presented in the Appendix.

4.1.1 Results for Simulation 1: RT.

The intent of this simulation was to present results for all three conditions in terms of the response time (RT), computed by deriving equations to convert the number of iterations required to compute the output (RT) for that condition. A robust set of such equations (to convert number of response iterations to RT) could not be established and so results for this simulation are presented in terms of the number of iterations required to compute the output.

Table 4.4 lists the results for all conditions with high levels of trait anxiety while Table 4.5 lists the same for low levels of trait anxiety. The interpretation of the anxiety condition (i.e., whether it is trait or state anxiety), depends on the condition being simulated. Specifically, the intensity condition simulates performance on the dot probe task under high levels of state anxiety while the other two conditions, exposure and interaction, simulate performance under high levels of trait anxiety. The first column in each table lists the name of the condition.

The dot probe task requires participants to indicate the side on which the dot appeared. The same was also required of the model; that is, to have the model output the side on which the dot appeared. The interaction and both intensity mechanisms could not produce the correct output for mismatch patterns.

Baseline mechanism. In this training mechanism, the network was exposed to five negative stimuli and four neutral ones in each epoch for both high and low anxiety. The condition assumed similar reaction to both negative and neutral stimuli under both high and low anxiety conditions, thus simulating the response of a participant with normal
level of anxiety. As expected, the network responded with the correct output for all input patterns, that is, it correctly followed the side on which the dot appeared. An unexpected result was that under high anxiety, the network responded faster when the negative pattern appeared on the left than when it appeared on the right. The network took 8 iterations to compute the output under the match condition when the negative picture was on the left as opposed to 10 when the picture appeared on the right. Similarly, it took 9

Table 4.4. Results of Simulation 1 under conditions of high anxiety.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Number of Iterations</th>
<th>Evidence-Left Output Unit</th>
<th>Evidence-Right Output Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pattern 1: Anxiety- High, Negative Pic. - Left, Probe- Left. Correct Output- Left (Match Left)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>8</td>
<td>1.00412</td>
<td>-0.83204</td>
</tr>
<tr>
<td>Exposure 3x</td>
<td>8</td>
<td>1.00486</td>
<td>-0.83278</td>
</tr>
<tr>
<td>Exposure 5x</td>
<td>8</td>
<td>1.00487</td>
<td>-0.83279</td>
</tr>
<tr>
<td>Interaction</td>
<td>8</td>
<td>1.00476</td>
<td>-0.83268</td>
</tr>
<tr>
<td>Intensity</td>
<td>8</td>
<td>1.0047</td>
<td>-0.83262</td>
</tr>
<tr>
<td>Intensity 8</td>
<td>8</td>
<td>1.00472</td>
<td>-0.83264</td>
</tr>
<tr>
<td><strong>Pattern 2: Anxiety- High, Negative Pic. - Right, Probe- Right. Correct Output: Right (Match Right)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>10</td>
<td>-0.69662</td>
<td>1.09567</td>
</tr>
<tr>
<td>Exposure 3x</td>
<td>9</td>
<td>-0.84056</td>
<td>1.07813</td>
</tr>
<tr>
<td>Exposure 5x</td>
<td>9</td>
<td>-0.8562</td>
<td>1.09377</td>
</tr>
<tr>
<td>Interaction</td>
<td>10</td>
<td>-0.67263</td>
<td>1.07168</td>
</tr>
<tr>
<td>Intensity</td>
<td>9</td>
<td>-0.84577</td>
<td>1.08334</td>
</tr>
<tr>
<td>Intensity 8</td>
<td>9</td>
<td>-0.85023</td>
<td>1.0878</td>
</tr>
<tr>
<td><strong>Pattern 3: Anxiety- High, Negative Pic. - Left, Probe- Right. Correct Output: Right (Mismatch Right)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>9</td>
<td>-0.8549</td>
<td>1.09246</td>
</tr>
<tr>
<td>Exposure 3x</td>
<td>9</td>
<td>-0.85619</td>
<td>1.09375</td>
</tr>
<tr>
<td>Exposure 5x</td>
<td>9</td>
<td>-0.8562</td>
<td>1.09376</td>
</tr>
<tr>
<td>Interaction</td>
<td>11</td>
<td>1.09849</td>
<td>-0.70685</td>
</tr>
<tr>
<td>Intensity</td>
<td>10</td>
<td>1.14387</td>
<td>-0.74483</td>
</tr>
<tr>
<td>Intensity 8</td>
<td>8</td>
<td>1.00243</td>
<td>-0.83035</td>
</tr>
<tr>
<td><strong>Pattern 4: Anxiety- High, Negative Pic. - Right, Probe- Left. Correct Output: Left (Mismatch Left)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>10</td>
<td>1.13939</td>
<td>-0.74035</td>
</tr>
<tr>
<td>Exposure 3x</td>
<td>11</td>
<td>1.11164</td>
<td>-0.72</td>
</tr>
<tr>
<td>Exposure 5x</td>
<td>8</td>
<td>1.00476</td>
<td>-0.83268</td>
</tr>
<tr>
<td>Interaction</td>
<td>10</td>
<td>-0.60555</td>
<td>1.00459</td>
</tr>
<tr>
<td>Intensity</td>
<td>9</td>
<td>-0.79285</td>
<td>1.03041</td>
</tr>
<tr>
<td>Intensity 8</td>
<td>10</td>
<td>-0.63408</td>
<td>1.03313</td>
</tr>
</tbody>
</table>
iterations for the mismatch right (probe on right but negative image left) condition versus 10 on the mismatch right. A similar pattern of faster response to negative picture was not

Table 4.5. Results of Simulation 1 under condition of low anxiety.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Number of Iterations</th>
<th>Evidence Left Output Unit</th>
<th>Evidence Right Output Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern 1: Anxiety- Low, Negative Pic. - Left, Probe- Left. Correct Output- Left (Match Left)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>10</td>
<td>1.1538</td>
<td>-0.75475</td>
</tr>
<tr>
<td>Exposure 3x</td>
<td>10</td>
<td>1.07328</td>
<td>-0.67424</td>
</tr>
<tr>
<td>Exposure 5x</td>
<td>8</td>
<td>1.00486</td>
<td>-0.83279</td>
</tr>
<tr>
<td>Interaction</td>
<td>10</td>
<td>1.14758</td>
<td>-0.74853</td>
</tr>
<tr>
<td>Intensity</td>
<td>10</td>
<td>1.14923</td>
<td>-0.75018</td>
</tr>
<tr>
<td>Intensity 8</td>
<td>10</td>
<td>1.14979</td>
<td>-0.75075</td>
</tr>
<tr>
<td>Pattern 2: Anxiety- Low, Negative Pic. - Right, Probe- Right. Correct Output: Right (Match Right)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>10</td>
<td>-0.663</td>
<td>1.06205</td>
</tr>
<tr>
<td>Exposure 3x</td>
<td>9</td>
<td>-0.83967</td>
<td>1.07724</td>
</tr>
<tr>
<td>Exposure 5x</td>
<td>9</td>
<td>-0.8562</td>
<td>1.09377</td>
</tr>
<tr>
<td>Interaction</td>
<td>9</td>
<td>-0.77923</td>
<td>1.01679</td>
</tr>
<tr>
<td>Intensity</td>
<td>10</td>
<td>-0.67575</td>
<td>1.0748</td>
</tr>
<tr>
<td>Intensity 8</td>
<td>10</td>
<td>-0.66966</td>
<td>1.0687</td>
</tr>
<tr>
<td>Pattern 3: Anxiety- Low, Negative Pic. - Left, Probe- Right. Correct Output: Right (Mismatch Right)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>9</td>
<td>-0.85449</td>
<td>1.09205</td>
</tr>
<tr>
<td>Exposure 3x</td>
<td>9</td>
<td>-0.85618</td>
<td>1.09375</td>
</tr>
<tr>
<td>Exposure 5x</td>
<td>9</td>
<td>-0.85617</td>
<td>1.09374</td>
</tr>
<tr>
<td>Interaction</td>
<td>9</td>
<td>-0.83441</td>
<td>1.07198</td>
</tr>
<tr>
<td>Intensity</td>
<td>9</td>
<td>-0.85485</td>
<td>1.09242</td>
</tr>
<tr>
<td>Intensity 8</td>
<td>9</td>
<td>-0.85467</td>
<td>1.09224</td>
</tr>
<tr>
<td>Pattern 4: Anxiety- Low, Negative Pic. - Right, Probe- Left. Correct Output: Left (Mismatch Left)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>10</td>
<td>1.14272</td>
<td>-0.74367</td>
</tr>
<tr>
<td>Exposure 3x</td>
<td>10</td>
<td>1.04377</td>
<td>-0.64473</td>
</tr>
<tr>
<td>Exposure 5x</td>
<td>8</td>
<td>1.00457</td>
<td>-0.83249</td>
</tr>
<tr>
<td>Interaction</td>
<td>11</td>
<td>1.10457</td>
<td>-0.71292</td>
</tr>
<tr>
<td>Intensity</td>
<td>10</td>
<td>1.14661</td>
<td>-0.74757</td>
</tr>
<tr>
<td>Intensity 8</td>
<td>10</td>
<td>1.14914</td>
<td>-0.75009</td>
</tr>
</tbody>
</table>

that apparent for low anxiety. Results of the baseline mechanism provided a yardstick by which the results of the other conditions could be measured and thereby
allow measurement of the efficacy of the training mechanism in producing the desired results.

**Exposure.** Training the exposure condition took longer than the baseline condition (24 cycles versus 18). The mechanism consisted of exposing the network to three times as many patterns of negative stimuli as the baseline mechanism under high anxiety for each epoch. The resulting network produced the output faster than the baseline mechanism in the match right (9 iterations against 10) and slower in the mismatch right (11 iterations against 10) conditions. The trend of reacting to negative pictures appearing on the left persisted in this condition as well, although the difference was reduced in the match condition and increased in the mismatch conditions.

To determine if repeated exposure to the same stimuli had any further effect on the performance of the network, the network was trained on five times as many negative stimuli under high anxiety as in the baseline mechanism. Results of this are listed as “Exposure 5x”. To avoid confusion, the basic exposure mechanism will be referred to as “Exposure 3x” throughout the remainder of this text. As can be seen, performance mostly remained the same as the Exposure 3x mechanism except for an observed speed up in case of mismatch left. Both the exposure conditions displayed the existence of an attentional bias by responding faster for the match conditions as opposed to mismatch conditions when the negative picture appeared on the same side.

Performance of the network for the Exposure 3x mechanism for low anxiety (Table 4.5) remained virtually unchanged from the baseline condition except for a slight decrease observed in the number of response iterations for match-right condition (9 iterations as opposed to 10). On the other hand, continued exposure, as represented by the
Exposure 5x mechanism, resulted in quicker response times for low anxiety in three of the four trials (namely, match-left, match-right and mismatch-left) as compared to the baseline mechanism, and in half the trials (match-left and mismatch-left), when compared with the Exposure 3x mechanism. Implications of this result are discussed in Chapter 5. Further, the trend of responding to salient stimuli appearing on the left was largely absent in this case.

**Interaction.** Although the interaction mechanism correctly simulated response to all test patterns for low anxiety and the match conditions for high anxiety, the mechanism completely failed to output the correct response in both mismatch conditions for high anxiety. The trend of favoring negative pictures appearing on the left side continued with the current mechanism also. The mechanism gives some insight into the inappropriateness of the interaction hypothesis as represented in this network.

**Intensity.** Two intensity mechanisms were simulated, each one presenting the input unit for the negative unit with a higher activation under high anxiety as opposed to low anxiety. Both the mechanisms used the network trained in the baseline condition. The basic intensity mechanism presented an activation value of 4.0 while the second intensity mechanism (*Intensity 8*) doubled the activation value to 8.0. The mechanism reduced the response times in the match conditions from the corresponding times under the baseline mechanism but, like the interaction mechanism, could not produce the correct output for both the mismatch conditions under high anxiety.

4.1.2 Results for Simulation 2: Activation.

Simulation 2 involved computing the final results obtained by running the network for one iteration only. Table 4.6 lists the results for this simulation for high and low
Table 4.6. Simulation 2: Output activations for high and low anxiety.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Activation Left Output Unit</th>
<th>Activation Right Output Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pattern 1: Negative Pic. - Left, Probe- Left. Correct Output- Left</strong>&lt;br&gt; (Match Left)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Anxiety</td>
<td>Low Anxiety</td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>Right</td>
<td>Left</td>
</tr>
<tr>
<td>Basic</td>
<td>0.999906</td>
<td>0.000149</td>
</tr>
<tr>
<td>Exposure</td>
<td>0.999996</td>
<td>0.000000</td>
</tr>
<tr>
<td>Exposure 5x</td>
<td>0.999911</td>
<td>0.000105</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.999968</td>
<td>0.000024</td>
</tr>
<tr>
<td>Intensity</td>
<td>0.999980</td>
<td>0.000022</td>
</tr>
<tr>
<td>Intensity 8</td>
<td>0.999978</td>
<td>0.000044</td>
</tr>
<tr>
<td><strong>Pattern 2: Negative Pic. - Right, Probe- Right. Correct Output- Right</strong>&lt;br&gt; (Match Right)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Anxiety</td>
<td>Low Anxiety</td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>Right</td>
<td>Left</td>
</tr>
<tr>
<td>Basic</td>
<td>0.000686</td>
<td>0.999647</td>
</tr>
<tr>
<td>Exposure</td>
<td>0.000000</td>
<td>0.999983</td>
</tr>
<tr>
<td>Exposure 5x</td>
<td>0.000500</td>
<td>0.999654</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.000062</td>
<td>0.999965</td>
</tr>
<tr>
<td>Intensity</td>
<td>0.000056</td>
<td>0.999962</td>
</tr>
<tr>
<td>Intensity 8</td>
<td>0.015610</td>
<td>0.988949</td>
</tr>
<tr>
<td><strong>Pattern 3: Negative Pic. - Left, Probe- Right. Correct Output: Right</strong>&lt;br&gt; (Mismatch Right)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Anxiety</td>
<td>Low Anxiety</td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>Right</td>
<td>Left</td>
</tr>
<tr>
<td>Basic</td>
<td>0.002218</td>
<td>0.999012</td>
</tr>
<tr>
<td>Exposure</td>
<td>2.76E-10</td>
<td>0.999991</td>
</tr>
<tr>
<td>Exposure 5x</td>
<td>0.001175</td>
<td>0.998576</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.999982</td>
<td>0.000022</td>
</tr>
<tr>
<td>Intensity</td>
<td>0.999979</td>
<td>0.000032</td>
</tr>
<tr>
<td>Intensity 8</td>
<td>0.999951</td>
<td>0.000079</td>
</tr>
<tr>
<td><strong>Pattern 4: Negative Pic. - Right, Probe- Left. Correct Output: Left</strong>&lt;br&gt; (Mismatch Left)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Anxiety</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>Right</td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>0.999713</td>
<td>0.000483</td>
</tr>
<tr>
<td>Exposure</td>
<td>0.999919</td>
<td>8.80E-08</td>
</tr>
<tr>
<td>Exposure 5x</td>
<td>0.999884</td>
<td>0.000181</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.000186</td>
<td>0.999953</td>
</tr>
<tr>
<td>Intensity</td>
<td>0.000152</td>
<td>0.999942</td>
</tr>
<tr>
<td>Intensity 8</td>
<td>0.094360</td>
<td>0.955398</td>
</tr>
</tbody>
</table>

anxiety individuals. Although activations of output units do not apply to the dot probe as clearly and directly as the number of iterations to compute the response times, such values are more robust than the number of response iterations as fewer variables are
required to be in agreement to replicate the results. More importantly, activation data is indispensable in verifying the accuracy of a NN model. The activation values are consistent with the results of Simulation 1.

One finding emerging from the activation values is that when the activation value is increased (as in the intensity-8 mechanism), the activation of the correct output unit decreases while the activation of the incorrect unit increases. This points towards an increase in error in the dot probe performance under high state anxiety.

4.2 Bayesian Network Model

One of the most crucial aspects of probabilistic modeling using Bayesian networks (BN), at par with establishing the structure of the network, is determining the conditional probability tables (CPTs). The CPTs for each variable were computed according to the method outlined in Chapter 3, and are shown in Table 4.7. Arousal rate (AR), anxiety (Anx) and probe side (PS) are independent variables. Values in the CPTs list the prior probabilities of each of these variables. Table 4.7 (a) lists the prior probability for AR; the values of 0.5 for negative arousal rate indicates a mildly arousing negative picture. Similarly, in Table 4.7 (b), prior probability of 0.5 specifies that a given individual is equally likely to have high and low levels of trait anxiety. The fact that for any trial, the dot probe is equally likely to appear replacing the negative image as it is for the neutral image is highlighted with prior probability values of 0.5.

The critical CPT is that of the variable attention direction (AD). The variable
depends upon the anxiety and arousal level of the negative image. CPTs for this variable were established using the procedure outlined in Chapter 3. Three different CPTs were set up, (a) using the data of all the conditions, (b) using data when the negative image appeared on the left only, and (c) by using data from trials in which the negative picture appeared on the right. Values for these conditions are shown in Table 4.8.

After establishing the CPTs, the network was used to compute the posterior probabilities of $AD=\text{negative}$, $AR=\text{negative}$, and $Anx=\text{high}$. Following the initial computation of prior probabilities, the effect of knowing $Anx=\text{high}$ in addition to

Table 4.7. Conditional Probability Tables for variables $Arousal \text{ Rating (AR)}$, $Anxiety (Anx)$, $Probe \text{ Side (PS)}$, and $Reaction \text{ Time (RT)}$

<table>
<thead>
<tr>
<th>$AR= \text{Neg}$</th>
<th>$AR= \text{Ntrl}$</th>
<th>$Anx= \text{High}$</th>
<th>$Anx= \text{Low}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

(a)  

<table>
<thead>
<tr>
<th>$PS = \text{Neg}$</th>
<th>$PS = \text{Ntrl}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

(c)

<table>
<thead>
<tr>
<th>$AD= \text{Neg}$</th>
<th>$PS= \text{Neg}$</th>
<th>$RT= \text{Low}$</th>
<th>$RT= \text{High}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0</td>
<td>0.95</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>0 1</td>
<td>0.05</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>1 0</td>
<td>0.05</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>1 1</td>
<td>0.95</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

(d)

<table>
<thead>
<tr>
<th>$AR= \text{Neg}$</th>
<th>$Anx = \text{High}$</th>
<th>Case 1: All data</th>
<th>Case 2: Data from left only</th>
<th>Case 3: Data from right only</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>0 1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1 0</td>
<td>0.5035</td>
<td>0.4965</td>
<td>0.5416</td>
<td>0.4583</td>
</tr>
<tr>
<td>1 1</td>
<td>0.5068</td>
<td>0.4932</td>
<td>0.5211</td>
<td>0.4789</td>
</tr>
</tbody>
</table>

Table 4.8. Conditional Probability Tables for variable $AR$ for the Bayesian network.
PS=Neg and RT=Low on the posterior probabilities of AD and AR was computed. The same effect for knowing the state of AR on the probabilities of AD and Anx was also computed. These posterior probabilities are listed in Table 4.9, Table 4.10, and Table 4.11, respectively.

The first row of each table yields the posterior probabilities of AR=negative (i.e., the image is negative), AD=negative (i.e., the individual attends towards the negative image) and Anx=high (i.e., the probability that anxiety level of the individual is high) given that the probe replaces the negative image (PS=Neg) and the recorded RT is low (RT=low). A comparison of the three tables highlights the import of accuracy of the CPTs. This is discussed further in Chapter 5.

Table 4.9 Posterior probabilities of various variables computed given the evidence (in bold) using the CPT derived from all data.

<table>
<thead>
<tr>
<th>AR= Neg</th>
<th>AD=Neg</th>
<th>Anx=high</th>
<th>PS=Neg</th>
<th>RT=low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>Post</td>
<td>Prior</td>
<td>Post</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5023</td>
<td>0.5026</td>
<td>0.9505</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4977</td>
<td>0.5026</td>
<td>0.0505</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.503</td>
<td>0.5026</td>
<td>0.9506</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4969</td>
<td>0.5026</td>
<td>0.0506</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5016</td>
<td>0.5026</td>
<td>0.9503</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4984</td>
<td>0.5026</td>
<td>0.0503</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>NA</td>
<td>0.5026</td>
<td>0.951</td>
<td>0.5015</td>
</tr>
<tr>
<td>1</td>
<td>NA</td>
<td>0.5026</td>
<td>0.051</td>
<td>0.4985</td>
</tr>
</tbody>
</table>

Table 4.10 Posterior probabilities of various variables computed given the evidence (in bold) using the CPT derived from data using pictures appearing on the left only.

<table>
<thead>
<tr>
<th>AR= Neg</th>
<th>AD=Neg</th>
<th>Anx=high</th>
<th>PS=Neg</th>
<th>RT=low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>Post</td>
<td>Prior</td>
<td>Post</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5137</td>
<td>0.5157</td>
<td>0.9529</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4855</td>
<td>0.5157</td>
<td>0.0531</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5093</td>
<td>0.5157</td>
<td>0.952</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4903</td>
<td>0.5157</td>
<td>0.052</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4806</td>
<td>0.5157</td>
<td>0.541</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.518</td>
<td>0.5157</td>
<td>0.9538</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>NA</td>
<td>0.5157</td>
<td>0.956</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>NA</td>
<td>0.5157</td>
<td>0.0563</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Table 4.11 Posterior probabilities of various variables computed given the evidence (in bold) using the CPT derived from data using pictures appearing on the left only.

<table>
<thead>
<tr>
<th>AR= Neg</th>
<th>AD=Neg</th>
<th>Anx=high</th>
<th>PS=Neg</th>
<th>RT=low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>Post</td>
<td>Prior</td>
<td>Post</td>
<td>Evidence</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4902</td>
<td>0.4893</td>
<td>0.9479</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5095</td>
<td>0.4893</td>
<td>0.048</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.497</td>
<td>0.4893</td>
<td>0.9494</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>0.503</td>
<td>0.4893</td>
<td>0.0494</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4832</td>
<td>0.4893</td>
<td>0.9464</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5158</td>
<td>0.4893</td>
<td>0.0467</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>NA</td>
<td>0.4893</td>
<td>0.9458</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>NA</td>
<td>0.4893</td>
<td>0.0461</td>
<td>0.5</td>
</tr>
</tbody>
</table>

The BN model was also used to test the effect of a change in trait anxiety levels on the posterior probability of both $AR$ and $Anx$ given that the individual is known to be attending towards the negative image. These computations were carried out only with the CPT obtained by considering all trials from the available data and are presented in Table 4.12 through Table 4.15. Although the model performed the computations, interpretation of the results obtained is not straightforward and is riddled with ambiguities. These are highlighted in Chapter 5, along with an explanation and discussion of the results.

Table 4.12 Prior and posterior probabilities for various prior probability values of $AR= Neg$.

<table>
<thead>
<tr>
<th>AR= Neg</th>
<th>AD=Neg</th>
<th>Anx=high</th>
<th>PS=Neg</th>
<th>RT=low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>Post</td>
<td>Prior</td>
<td>Post</td>
<td>Evidence</td>
</tr>
<tr>
<td>0.25</td>
<td>0.2427</td>
<td>0.5026</td>
<td>0.949</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4902</td>
<td>0.5026</td>
<td>0.9479</td>
<td>0.5</td>
</tr>
<tr>
<td>0.75</td>
<td>0.7425</td>
<td>0.5026</td>
<td>0.9469</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.5026</td>
<td>0.9458</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 4.13 Prior and posterior probabilities for various prior probability values of $Anx= high$.

<table>
<thead>
<tr>
<th>AR= Neg</th>
<th>AD=Neg</th>
<th>Anx=high</th>
<th>PS=Neg</th>
<th>RT=low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>Post</td>
<td>Prior</td>
<td>Post</td>
<td>Evidence</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4832</td>
<td>0.4856</td>
<td>0.9464</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4867</td>
<td>0.4856</td>
<td>0.9472</td>
<td>0.25</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4902</td>
<td>0.4856</td>
<td>0.9479</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4936</td>
<td>0.4856</td>
<td>0.9486</td>
<td>0.75</td>
</tr>
<tr>
<td>0.5</td>
<td>0.497</td>
<td>0.4856</td>
<td>0.9494</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5158</td>
<td>0.4893</td>
<td>0.0467</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4832</td>
<td>0.4893</td>
<td>0.9464</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>NA</td>
<td>0.4856</td>
<td>0.9458</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>NA</td>
<td>0.4856</td>
<td>0.9461</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>NA</td>
<td>0.4856</td>
<td>0.9493</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>NA</td>
<td>0.4856</td>
<td>0.9461</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Table 4.14 Prior and posterior probabilities of $Anx= high$ for various prior probability values of $AR= Neg$ given $AD= neg$.

<table>
<thead>
<tr>
<th>$AR= Neg$</th>
<th>$AD=Neg$</th>
<th>Anx=high</th>
<th>$PS=Neg$</th>
<th>RT=low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>Post</td>
<td>Prior</td>
<td>Post</td>
<td>Evidence</td>
</tr>
<tr>
<td>0.25</td>
<td>0.2419</td>
<td>1</td>
<td>NA</td>
<td>0.5037</td>
</tr>
<tr>
<td>0.5</td>
<td>0.489</td>
<td>1</td>
<td>NA</td>
<td>0.5075</td>
</tr>
<tr>
<td>0.75</td>
<td>0.7417</td>
<td>1</td>
<td>NA</td>
<td>0.5114</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>NA</td>
<td>0.5154</td>
</tr>
</tbody>
</table>

Table 4.15 Prior and posterior probabilities of $AR= Neg$ for various prior probability values of $Anx=high$ given $AD= neg$.

<table>
<thead>
<tr>
<th>$AR= Neg$</th>
<th>$AD=Neg$</th>
<th>Anx=high</th>
<th>$PS=Neg$</th>
<th>RT=low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>Post</td>
<td>Prior</td>
<td>Post</td>
<td>Evidence</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4812</td>
<td>1</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4852</td>
<td>1</td>
<td>NA</td>
<td>0.25</td>
</tr>
<tr>
<td>0.5</td>
<td>0.489</td>
<td>1</td>
<td>NA</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4929</td>
<td>1</td>
<td>NA</td>
<td>0.75</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4966</td>
<td>1</td>
<td>NA</td>
<td>1</td>
</tr>
</tbody>
</table>
The purpose of the current study was two-fold; (1) to simulate performance on the dot probe task under high and low levels of anxiety using a neural network (NN) and, (2) to provide a new method of analyzing the results of the dot probe and allow emergence of probabilistic relationships among variables using a Bayesian network (BN) model. Results of various simulations are interpreted and discussed below; a rationale is provided for the findings and future research directions are proposed.

5.1 Neural Network Model

The basic aim of the NN model was to test three potential mechanisms explaining attentional bias as understood from the findings of the dot probe and to provide a common ground to compare results obtained from the three. The study also represented the first known attempt to simulate performance on the dot probe using a NN. Despite our lofty expectations, the model fell short of providing an accurate simulation of human performance on the dot probe task. However, the model was able to simulate the basic empirical findings of the dot probe. Although final results could not be depicted in terms of response time (RT), as was the original intent of the study, RT differences could still be estimated using the number of iterations required to compute the output for each mechanism. Various aspects of the NN and the explanations to the failure of the model are discussed below.
5.1.1 Weights of the Network.

Appendix A lists the weights of the NN after training for each of the mechanisms. The network was initialized by randomly assigning weights between 0 and 1 to all connections and biases. Training for each mechanism started from the same initial state and ended when the mean squared error (MSE) was within tolerance (<0.0001) or the network completed 500 epochs, whichever was earlier. For all training mechanisms, except the interaction mechanism, the network managed to achieve MSE within the specified limits.

Cohen et al. (1997) explained Stroop effects on the basis strength of processing (SOP), which refers to higher activation levels of the some of the units of the network either due to strengthened connections between the units or due to a higher resting activation of the unit. However, the concept is not directly applicable for every mechanism employed in the current study because the NN lacks the pathways that were a basic element of the Cohen model.¹ So even though the some units have higher activation levels due to increased strength of connections, interpreting the increased activation for the dot probe is not straightforward. An interesting observation is that the connection weights for Exposure 5x are not necessarily greater in magnitude than the Exposure 3x mechanism, leading to the conclusion that repeated exposure to a subset of the patterns does not increase the connection strength of all the connections but increases the net effect of the stimuli related to those patterns. In other words, the same activation pattern results in a higher activation at the output level even though not all the weights related to that pattern are strengthened. This can also be viewed as an increase in efficiency of the

¹ Recall that the Cohen model was a partially connected NN, with connections forming separate pathways for each of the two tasks, color naming and word reading. See Chapter 2 for details.
network, where weights are adjusted to handle the specific training patterns more efficiently.

Weights can also be used to explain the small range of the number of iterations required to compute the output (which led to a failure of obtaining equations to convert the response iterations to RT). Setting the value of the MSE to 0.0001 caused the network weights to reach an asymptotic value and as a result, it produced very similar output activations for all training mechanisms. Thus, even for the mismatch conditions in the interaction and intensity mechanisms, although the network does not produce the correct output for high anxiety, the number of response iterations remains the same. An obvious method to correct the problem is to set a higher value of the MSE as the stopping criterion during training (say 0.001). Doing so would have two main effects, (1) it would not strengthen the weights to the extent that they reach asymptotic values and so result in almost the same activation for a given input pattern under all training mechanisms, and (2) it would allow a higher margin of error in the network and so lead to a higher variability of performance. A higher margin of error would also allow greater increase in performance with training. Another possible correction is to modify the training algorithm; the current study trained the network using the activation values of the output units and tested it using the mechanism of collecting evidence employed by Cohen et al. (1990). Using the same mechanism for both training and testing could possibly lead to a larger differential in response times.

5.1.2 Performance of the Training Mechanisms

Of the three basic mechanisms used (exposure, intensity, and interaction), only the exposure mechanism produced results consistent with empirical findings of the dot probe task. Between the two exposure mechanisms, Exposure 3x fit the empirical findings
better than Exposure 5x. Thus, it may be argued that of the three mechanisms tested in the current model, repeated exposure to negative stimuli best explains the etiology of attentional bias as neither of the other two training mechanisms was able to successfully elicit behavior consistent with empirical findings of the dot probe task.

The above explanation can be used as the basis for the following conclusion, “Given sufficient exposure to stimuli perceived highly threatening should cause an attentional bias towards those stimuli regardless of anxiety level”. The statement is consistent with the models of Beck (1976) and Bower (1980) who suggested that repeated exposure to a high threat stimulus could strengthen the connections for the stimulus, thereby lowering the threshold of the activation required to elicit the attentional bias. The view is also consistent with Lang’s biphasic theory (2000), which states that a stimulus that is rated sufficiently high in threat will cause the same effects in all individuals, regardless of anxiety level. For high anxiety individuals, the threshold is lower than that for low anxiety individuals. According to this explanation, repeated exposure to certain stimuli could lead to the lowering of the threshold and so produce an attention bias.

The two intensity mechanisms were not able to correctly simulate the dot probe task as they yielded an incorrect output for the mismatch conditions under high anxiety. However, both intensity mechanisms were able to produce higher activation in the output unit representing the negative picture, suggesting the existence of a bias towards the negative image. This can be interpreted to mean that high anxiety individuals perceive the stimulus to have a higher threat value, which causes the attentional bias. Further, it is clear that a higher intensity stimulus makes a larger demand on attention.
The intensity and exposure mechanisms can be combined to form a new potential explanation of the etiology of attentional bias; assume all individuals are low anxiety to begin with but certain individuals are repeatedly exposed to stimuli that they perceive as highly threatening. Repeated exposure to the high threat stimuli strengthens the connections salient to that class of stimuli, effectively increasing the perceived intensity with exposure. Such a view could explain the origin of disorders such as posttraumatic stress disorder (PTSD), where individuals are repeatedly exposed to highly threatening stimuli in a high stress environment. In this explanation, intensity appears to be the more critical element of the two causes as individuals have been known to be afflicted by PTSD after being involved in one stressful situation, like an accident. Possibly the high intensity of the stimulus causes the initial breach to attract the individuals attention. If the stimulus is sufficiently intense, it could lead to PTSD. If not, repeated exposure to the same stimulus could likely cause the onset of the disorder.

A possible reason of failure of the intensity and interaction mechanisms maybe an inadequate attention mechanism encoded in the network, rather than shortcomings in the mechanisms themselves. Attention mechanism in this case refers to the mechanism used by the network so the network eventually attends towards the side on which the probes appear. Indeed, if an attention mechanism can be developed that causes the network to respond with the side on which the dot appears, the two mechanisms will exhibit behavior consistent with the existence of an attentional bias.

Further investigation using a revised attention mechanism is needed before any conclusions regarding the viability of the two mechanisms can be made. However, one unexpected observation is the similarity in the results obtained using intensity and the
interaction mechanisms, which implies that the two mechanisms actually work in the same manner.

5.2 Bayesian Network

The Bayesian network (BN) model was intended to serve as a proof of concept in being able to model data pertaining to attentional bias obtained from the dot probe. The model fell short of expectations and did not yield probabilistic relationships among the variables. Also, it was marred by problems in modeling certain conditions and interpreting the results obtained. However, the model provided some important lessons that can be relied upon to build a better model in the future. Three primary observations emerging from the network are as follows:

1. The conditional probability table (CPT) for attention direction (AD) was set up successfully and provided a starting point in the attempt to quantify relationships among variables.

13. The interpretation of the probability values of variables was not clearly defined.

14. The network could not be tested against actual data as was originally intended, in part due to very broad dichotomies used to represent the states of the variables and partly due to (2) above. As such the accuracy of the probabilistic relationships could not be verified.

5.2.1 The Conditional Probability Tables

As stated in Chapter 4, one of the most critical elements of the BN is the CPT. The CPT encodes the probabilistic relationships among the variables. Of the five variables modeled in the BN, the most important and the only one not evident from the dot probe task is Attention Direction (AD). In fact, the primary objective of the dot probe task is to infer the attention-direction of the participant, typically done from the response time. The current study used the same principle to derive the CPT for AD. The CPTs so obtained
represent the first attempt to probabilistically quantify the relationship among the variables of the dot probe task.

The CPTs for AD highlight the difference in probability of attending towards the negative picture when the picture appears on the left as opposed to when the picture appears on the right. The main relationships represented in the CPTs of Table 4.9 are reproduced below in Table 5.1.

Table 5.1. Conditional Probability Tables for variable AR for the Bayesian network.

<table>
<thead>
<tr>
<th>AR= Neg</th>
<th>Anx = High</th>
<th>Case 1: All data</th>
<th>Case 2: Data from left only</th>
<th>Case 3: Data from right only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AD= Neg</td>
<td>AD= Ntrl</td>
<td>AD= Neg</td>
<td>AD= Ntrl</td>
</tr>
<tr>
<td>1</td>
<td>0.5035</td>
<td>0.4965</td>
<td>0.5416</td>
<td>0.4583</td>
</tr>
<tr>
<td>1</td>
<td>0.5068</td>
<td>0.4932</td>
<td>0.5211</td>
<td>0.4789</td>
</tr>
</tbody>
</table>

The first row of the CPT can be stated as follows, “an individual with high anxiety attended to the negative picture appearing on the left with probability 0.5416 and with probability 0.4638 to a negative picture appearing on the right hand side”. Statistically speaking, this may not be as authoritative as stating the main effect and interaction effects that the picture side has with the reaction time, but it is certainly more discrete and quantifiable. Such a representation offers an advantage over the more traditional method of significance testing by presenting numbers that can be compared directly against numbers obtained in similar fashion from other studies.

Despite the fact that the probabilistic relationships encoded in the CPT could not be tested against real data, the manner in which they were computed is logical and straightforward, and lends itself to discretization of the relationships. The relationships established were used to infer the probability of $AD=$Neg for an individual with known anxiety. For row 3 of Table 4.10 (a), (reproduced below as Table 5.2) the probability of $AD=$Neg increases from 0.5026 to 0.9506 indicating the likelihood that an individual
known to be highly anxious will attend towards the negative image given $PS=\text{Neg}$ (probe replaces negative image) and $RT=\text{low}$. The arousal rating of an image required to grab the participants attention 95% of the time is given by the posterior probability of $AR=\text{Neg}=0.503$ (up from 0.5). Rows 3 through 6 (shaded) in Table 5.2 depict the two possible conditions ($PS=\text{Neg}, RT=\text{low} & PS=\text{Ntrl}, RT=\text{low}$) for $Anx=\text{high}$ and $Anx=\text{low}$. The rows can be interpreted as follows:

2. **Row 3.** Given the observations that the probe replaces the negative image and the recorded $RT$ is low, an individual with high anxiety will attend to a negative picture with normalized arousal rating $= 0.503$ with probability $= 0.9506$.

15. **Row 4.** Given the observations that the probe replaces the neutral image and the recorded $RT$ is low, an individual with high anxiety will attend to a negative picture with normalized arousal rating $= 0.4969$ with probability $= 0.0506$. The decrease in the probability of $AD=\text{Neg}$ is inferred from an apparent attention towards the neutral image. The inference may not appear to be consistent with the findings of the dot probe task (an individual with high anxiety attending to the neutral picture) but it should be noted that the network reflects the changes in probabilities given the observations. The observation in this case already points to a higher response time ($RT=\text{high}$) for the dot replacing the negative image. Thus, this condition can be used to answer the question, “What is the probability that an individual with high anxiety attends to the negative image given that the individual records a high RT for when the dot replaces the negative image? Also what is the normalized arousal rating of the negative image?”

16. **Row 5.** Represents the same observations as Row 3 but for a low anxiety individual. As can be seen, the probability of $AD=\text{Neg}$ decreased slightly, coupled with a decrease in the posterior probability of $AR=\text{Neg}$.

Similarly, rows 7 and 8 show inference of attention direction anxiety level of the individual given a high arousal rating. It should be noted that the values oscillate around 0.95 and 0.5 as these were the values set for $RT$ given $AD$ and $PS$ in the CPT for $RT$ (see Table 4.15).
Table 5.2 Posterior probabilities of various variables computed given the evidence (in bold) using the CPT derived from all data

<table>
<thead>
<tr>
<th>AR= Neg</th>
<th>AD=Neg</th>
<th>Anx=high</th>
<th>PS=Neg</th>
<th>RT=low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>Post</td>
<td>Prior</td>
<td>Post</td>
<td>Evidence</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5023</td>
<td>0.5026</td>
<td>0.9505</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4977</td>
<td>0.5026</td>
<td>0.0505</td>
<td>0.5</td>
</tr>
<tr>
<td>(row 3) 0.5</td>
<td>0.503</td>
<td>0.5026</td>
<td>0.9506</td>
<td>1</td>
</tr>
<tr>
<td>(row 4) 0.5</td>
<td>0.4969</td>
<td>0.5026</td>
<td>0.0506</td>
<td>1</td>
</tr>
<tr>
<td>(row 5) 0.5</td>
<td>0.5016</td>
<td>0.5026</td>
<td>0.9503</td>
<td>0</td>
</tr>
<tr>
<td>(row 6) 0.5</td>
<td>0.4984</td>
<td>0.5026</td>
<td>0.0503</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>NA</td>
<td>0.5026</td>
<td>0.951</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>NA</td>
<td>0.5026</td>
<td>0.051</td>
<td>0.5</td>
</tr>
</tbody>
</table>

5.2.2 Interpretation of Probability Values

A major shortcoming of the model was the lack of rules to interpret the probability values for all of the variables. Interpretation of the independent variables of the dot probe task was fairly straightforward; anxiety represented the normalized score obtained on the STAI-Trait, while *Arousal Rating* depicted the normalized IAPS arousal rating of the image. The ambiguity in interpretation arose in *Attention Direction*; for example, what does a value of $AD(neg) = 0.75$ specify? The participant devoting 75% attentional resources to the stimulus or does it predict the probability of the participant attending to the negative stimulus of given arousal rating? The latter is akin to establishing a likelihood of attention direction towards the negative stimulus. In either case, the interpretation must be clearly established before the results of the modeling with the BN can be used to establish robust relationships.

5.2.3 Testing Against Actual Data

All the variables used in the network had only two possible states (true and false) resulting in an overall lack of accuracy as adding evidence to the network became virtually impossible. It may be recalled that adding evidence to a variable in a BN means that the state of the variable is known with certainty (i.e., the probability of the variable
being in the given state is 1). Clearly, using broad dichotomies, such as a variable being very high negative or neutral in arousal, or an individual being only high or low in anxiety does not reflect actual conditions. For instance, if the median anxiety score is 37, classifying an individual with anxiety score of 38 as high anxious may not be accurate for the individual. It was believed that the network could be tested for any probability of a given variable by setting its prior probability. However, when the prior probability was set in such a fashion, the network computed a posterior probability of the variable as well. The posterior probability represented a change in the state of the variable, which negates the fact that the variable is fixed in a given state. For clarification, consider Table 4.11 (a) and Table 4.11 (b). Table 4.11(a) represents an attempt to compute the posterior probability of $\text{Anxiety} = \text{high}$ and $\text{AD} = \text{Negative}$ for various values of $\text{AR}$ of the image. As can be seen, the networks computes the posterior probability $\text{AR}$ also, which makes the results ambiguous; it cannot be stated that the values of $\text{AD}$ and $\text{Anxiety}$ are representative of an image of arousal rating 0.25. There is no clear cause and effect. Similarly, in Table 4.11 (b), the same happens with $\text{Anxiety}$.

An obvious solution is to increase the number of possible states of the variables of the network. For example, $\text{Anxiety}$ could be specified in terms of quartiles or deciles of the STAI score. This would allow adding evidence to the network and observing a clear effect (change in $\text{AD}$ and $\text{AR}$) of the cause (specifying $\text{Anxiety}$). Continuing with this line of reasoning, setting states for $\text{AD}$ and its implication must be clearly formed.

### 5.3 Statement of Limitations

Both the NN model and the BN model suffered from two limitations, one due to their nature (and therefore not unique to the current study), while the other directly referred to the extent to which the networks modeled the dot probe task.
The first limitation of the NN stemmed from the fact that NN are capable of modeling any problem domain when given the appropriate training. As a result, appropriate training is a critical aspect of a modeling using a NN. The current study attempted to model human performance on the dot probe task. A limitation of the study was in encoding the attention mechanism into the network. Essentially, the network was allowed to learn the attention mechanism from increased training to attend to the side on which the probe appeared. This training does not reflect the attentional mechanism used by human participants; where an individual actively and deliberately searches for the dot probe when the probe does not appear in the spatial location being attended to. Another limitation was the lack of ability to present the images for a certain amount of time and have the dot appear following image offset. This potentially limits the extent to which the task is modeled by the NN as the network applies the activation due to probes to the network without removing the activation due to the stimuli, which is contrary to the actual task. Additionally, the NN was limited in its ability to produce generalizable results due to its inability to present results in terms of RT.

The main inherent limitation of the BN model was they are based in probability theory, and so must obey the laws of probability. This limits the application of the network, specifically when supplying the conditional probability for *Arousal Rating*. Although $AR=Ntrl$ was possible in the network, the BN could not be configured to handle such a condition as one image of the network was assumed to be neutral and the other was assumed to be negative. AR represented the normalized arousal rating of the negative image; so $AR=Ntrl$ meant that both the displayed images were neutral. This information was inconsistent with information listed in the CPTs of other variables.
5.4 Future Research

The current study represents the first application of NN and BN to the dot probe task. Although both models represent a significant first step in the process of establishing robust models to model the dot probe task, much work is required before they can yield relationships among the variables of the task and insights into its mechanisms.

For the NN, the first step should be to set a reliable attentional selection mechanism. The training processes used to train the network also need to be revised so as to yield a larger difference in the number of response iterations. Once these conditions have been met and a working NN is in place, it can be used to derive a relationship between the anxiety and RT by using normalized anxiety scores to set the activation for the anxiety units in much the same that anxiety scores were used to set the conditional probabilities of the BN.

Unlike the NN, which has been accepted into psychological research, BN are as yet untested in the area. They also suffer from the disadvantage of being associated with the stigma of the word “probability”. Consequently, research is required to highlight the ease of use of such models along with their potential advantages. The first step in highlighting the potential advantages involves using the network to set probabilistic relationships among the variables of the task by extending the model used in the current study and adding finer divisions of states for the variables. Once such relationships are established, more variables can be added to the BN. For example, a BN can be developed to account for the relationship between arousal rating of a picture and its perceived arousal rating based on the anxiety of the individual. At this point, it should be reiterated that BN will not yield conclusions about data, instead a BN will yield potential relationships that can then be tested by actual studies.
5.5 Summary and Conclusion

Both the NN and the BN models met with mixed success and shortcomings but provided important insights into the mechanisms of attentional bias as measured by the dot probe task. Also, neither model was completely accurate, with the NN being riddled with problems with the attentional mechanism while the BN was plagued by a broad categorization of variables states.

The NN was not able to establish a robust set of relationships to translate the number of iterations required to produce the response to the RT. However, the number of response iterations themselves was representative of RT consistent with the hypothesis for the NN.

The BN model also could not establish a robust set of probabilistic relationships between the variables. However, the model strongly indicated the existence of such relationships among the variables. Additionally, the model highlighted the importance of establishing the relationships and further, verifying their accuracy.

Summarizing, the two models provided important insights into the potential mechanisms and relationships that exist among the variables of the network. The NN offers a simple and elegant medium to simulate performance and estimate the inner workings from the connection weights and performance of the network. The BN lends itself well to form a common ground to compare results from various studies and so form robust relationships. Last but not least, it should be mentioned that modeling with two networks is becoming increasingly simple with the availability of software packages (such as PDP++ for NN and Microsoft’s MSBNx and the Bayes Net Toolbox for Matlab by Kevin Murphy) and for the two available free of cost for research purposes.
### APPENDIX

**WEIGHTS AND BIASES OF THE NEURAL NETWORK MODEL**

<table>
<thead>
<tr>
<th></th>
<th>Input 1</th>
<th>Input 2</th>
<th>Input 3</th>
<th>Input 4</th>
<th>Input 5</th>
<th>Input 6</th>
<th>Input 7</th>
<th>Input 8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hid 1</strong></td>
<td>-0.8194</td>
<td>-0.4376</td>
<td>-0.4414</td>
<td>-0.5418</td>
<td>0.0397</td>
<td>0.2172</td>
<td>0.8931</td>
<td>0.3169</td>
</tr>
<tr>
<td><strong>Hid 2</strong></td>
<td>2.5209</td>
<td>1.7427</td>
<td>2.0858</td>
<td>-0.0680</td>
<td>-0.2193</td>
<td>-2.8738</td>
<td>-2.4553</td>
<td>-2.1877</td>
</tr>
<tr>
<td><strong>Hid 3</strong></td>
<td>5.9539</td>
<td>5.5501</td>
<td>5.0023</td>
<td>0.4513</td>
<td>0.6929</td>
<td>-5.8293</td>
<td>-5.0641</td>
<td>-4.8331</td>
</tr>
<tr>
<td><strong>Hid 4</strong></td>
<td>-0.9535</td>
<td>-1.1615</td>
<td>-0.8433</td>
<td>0.1623</td>
<td>-0.3398</td>
<td>0.3045</td>
<td>0.6285</td>
<td>0.8757</td>
</tr>
<tr>
<td><strong>Hid 5</strong></td>
<td>-2.4844</td>
<td>-2.4792</td>
<td>-2.0240</td>
<td>0.2296</td>
<td>0.1312</td>
<td>2.3250</td>
<td>2.1204</td>
<td>1.5140</td>
</tr>
</tbody>
</table>

Table A-1. Weights between the input and hidden units after training for baseline mechanism

<table>
<thead>
<tr>
<th></th>
<th>Input 1</th>
<th>Input 2</th>
<th>Input 3</th>
<th>Input 4</th>
<th>Input 5</th>
<th>Input 6</th>
<th>Input 7</th>
<th>Input 8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hid 1</strong></td>
<td>-2.1930</td>
<td>-0.8497</td>
<td>-3.2215</td>
<td>-1.2682</td>
<td>-0.2872</td>
<td>0.9026</td>
<td>1.3947</td>
<td>0.5475</td>
</tr>
<tr>
<td><strong>Hid 2</strong></td>
<td>4.9266</td>
<td>2.2117</td>
<td>5.9993</td>
<td>-1.2224</td>
<td>-0.9730</td>
<td>-7.4268</td>
<td>-5.6760</td>
<td>-3.7501</td>
</tr>
<tr>
<td><strong>Hid 4</strong></td>
<td>-1.8200</td>
<td>-1.1174</td>
<td>-2.5289</td>
<td>-1.2678</td>
<td>-0.8611</td>
<td>-9.8070</td>
<td>-0.2389</td>
<td>0.6141</td>
</tr>
<tr>
<td><strong>Hid 5</strong></td>
<td>-6.6953</td>
<td>-3.7033</td>
<td>-7.9756</td>
<td>0.6873</td>
<td>0.5719</td>
<td>6.9302</td>
<td>6.1968</td>
<td>3.7711</td>
</tr>
</tbody>
</table>

Table A-2. Weights between the input and hidden units after training for exposure 3x mechanism

<table>
<thead>
<tr>
<th></th>
<th>Input 1</th>
<th>Input 2</th>
<th>Input 3</th>
<th>Input 4</th>
<th>Input 5</th>
<th>Input 6</th>
<th>Input 7</th>
<th>Input 8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hid 1</strong></td>
<td>-0.7570</td>
<td>0.1512</td>
<td>-0.8906</td>
<td>-2.3080</td>
<td>-0.3927</td>
<td>-2.9407</td>
<td>-1.1205</td>
<td>-0.5194</td>
</tr>
<tr>
<td><strong>Hid 2</strong></td>
<td>4.4564</td>
<td>1.1325</td>
<td>5.8263</td>
<td>0.0009</td>
<td>-0.5831</td>
<td>-7.7683</td>
<td>-4.6599</td>
<td>-1.6034</td>
</tr>
<tr>
<td><strong>Hid 4</strong></td>
<td>1.7674</td>
<td>0.2543</td>
<td>1.7297</td>
<td>-1.2275</td>
<td>-0.8017</td>
<td>-5.4881</td>
<td>-3.5621</td>
<td>-0.9221</td>
</tr>
<tr>
<td><strong>Hid 5</strong></td>
<td>-5.1677</td>
<td>-3.7056</td>
<td>-5.4234</td>
<td>-0.1136</td>
<td>-0.4983</td>
<td>5.8968</td>
<td>5.9031</td>
<td>0.6683</td>
</tr>
</tbody>
</table>

Table A-3. Weights between the input and hidden units after training for exposure 5x mechanism

Table A-4. Weights between the input and hidden units after training for interaction mechanism
Table A-5. Weights layer 2 (between the hidden and the output units) after training for baseline mechanism

<table>
<thead>
<tr>
<th></th>
<th>Hid 1</th>
<th>Hid 2</th>
<th>Hid 3</th>
<th>Hid 4</th>
<th>Hid 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Out 1</td>
<td>-1.4613</td>
<td>3.1089</td>
<td>10.0525</td>
<td>-1.8728</td>
<td>-4.0868</td>
</tr>
<tr>
<td>Out 2</td>
<td>1.4282</td>
<td>-3.7617</td>
<td>-9.9330</td>
<td>1.9455</td>
<td>4.1715</td>
</tr>
</tbody>
</table>

Table A-6. Weights layer 2 (between the hidden and the output units) after training for exposure 3x mechanism

<table>
<thead>
<tr>
<th></th>
<th>Hid 1</th>
<th>Hid 2</th>
<th>Hid 3</th>
<th>Hid 4</th>
<th>Hid 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Out 1</td>
<td>-3.3076</td>
<td>7.29696</td>
<td>17.1209</td>
<td>-2.1178</td>
<td>-10.4691</td>
</tr>
<tr>
<td>Out 2</td>
<td>1.4928</td>
<td>-6.46894</td>
<td>-17.5494</td>
<td>1.18319</td>
<td>7.48194</td>
</tr>
</tbody>
</table>

Table A-7. Weights layer 2 (between the hidden and the output units) after training for exposure 5x mechanism

<table>
<thead>
<tr>
<th></th>
<th>Hid 1</th>
<th>Hid 2</th>
<th>Hid 3</th>
<th>Hid 4</th>
<th>Hid 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Out 1</td>
<td>-1.05497</td>
<td>5.75858</td>
<td>14.7933</td>
<td>1.9961</td>
<td>-7.01226</td>
</tr>
<tr>
<td>Out 2</td>
<td>1.51735</td>
<td>-4.04257</td>
<td>-16.831</td>
<td>-1.3786</td>
<td>6.70004</td>
</tr>
</tbody>
</table>

Table A-8. Weights layer 2 (between the hidden and the output units) after training for interaction mechanism

<table>
<thead>
<tr>
<th></th>
<th>Hid 1</th>
<th>Hid 2</th>
<th>Hid 3</th>
<th>Hid 4</th>
<th>Hid 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Out 1</td>
<td>-8.74957</td>
<td>6.29904</td>
<td>9.06092</td>
<td>-0.79573</td>
<td>-8.92881</td>
</tr>
<tr>
<td>Out 2</td>
<td>8.65513</td>
<td>-6.47485</td>
<td>-8.91653</td>
<td>0.873517</td>
<td>8.85659</td>
</tr>
</tbody>
</table>

Table A-9. Biases for hidden units for all training mechanisms

<table>
<thead>
<tr>
<th></th>
<th>Hid 1</th>
<th>Hid 2</th>
<th>Hid 3</th>
<th>Hid 4</th>
<th>Hid 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>-0.7642</td>
<td>-0.1195</td>
<td>0.6912</td>
<td>-0.4025</td>
<td>0.0630</td>
</tr>
<tr>
<td>Exposure 3x</td>
<td>-3.91226</td>
<td>-2.66741</td>
<td>-1.50587</td>
<td>-5.3248</td>
<td>-0.38503</td>
</tr>
<tr>
<td>Exposure 5x</td>
<td>-6.56968</td>
<td>-1.56829</td>
<td>2.30081</td>
<td>-5.47391</td>
<td>-0.73735</td>
</tr>
<tr>
<td>Interaction</td>
<td>3.5541</td>
<td>-4.67681</td>
<td>1.02403</td>
<td>-2.56079</td>
<td>-1.29189</td>
</tr>
</tbody>
</table>

Table A-10. Biases for the output units for all training mechanisms.

<table>
<thead>
<tr>
<th></th>
<th>Out 1 (Left)</th>
<th>Out 2 (Right)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>-2.8754</td>
<td>2.7568</td>
</tr>
<tr>
<td>Exposure 3x</td>
<td>-12.0748</td>
<td>3.25915</td>
</tr>
<tr>
<td>Exposure 5x</td>
<td>-10.2105</td>
<td>8.34891</td>
</tr>
<tr>
<td>Interaction</td>
<td>4.10617</td>
<td>-4.09197</td>
</tr>
</tbody>
</table>
LIST OF REFERENCES


Reiger, Schotte, Touyz, Beumont, Griffiths and Russell (1998)


BIOGRAPHICAL SKETCH

I began my professional educational career at the Regional Engineering College (REC) at Jalandhar in India, pursuing my Bachelor of Technology in computer science and engineering. REC did not have any basketball facilities. I had always considered myself an athlete and a basketball player before a student, so I transferred to Guru Nanak Dev University (GNDU), Amritsar, India, where I led the University basketball team to second place at the All India Inter University Basketball Championship.

While at GNDU, I came upon a book on sport psychology that featured an article on the use of imagery (specifically a technique called visuomotor behavior rehearsal (VMBR). The book was my first brush with the field of Sport Psychology. I graduated from GNDU at the top of my class in 1999.

In Fall of 2000, I started my master’s program in computer engineering. Here I met Dr. Anand Rangarajan and my fascination with neural networks and other tools for reasoning under uncertainty flourished. At the same time, I did not feel at home in the field of computer engineering. In Spring of 2002, I met with Dr. Christopher Janelle to explore my options for doing a Master’s in motor learning and control. The rest, as they say, is history. I found the missing element from my career. I am now looking forward to graduating and develop software to enhance human performance and better lifestyle.