SIMULATION AND MODEL CALIBRATION WITH SENSITIVITY ANALYSIS FOR THREAT DETECTION IN THE BRAIN

Keegan Lowenstein[(a)], Brian Leventhal[(b)], Kylie Drouin[(c)] , Robert Dowman[(c)], Katie Fowler[(b)], Sumona Mondal[(b)]

[(a)]Clarkson University, Department of Computer Science
[(b)]Clarkson University, Department of Mathematics
[(c)]Clarkson University, Department of Psychology

[(a)] keegan.lowenstein@gmail.com , [(b)]leventhbc@clarkson.edu, kfowler@clarkson.edu, smondal@clarkson.edu, [(c)]rdowman@clarkson.edu , drouinke@clarkson.edu

ABSTRACT
In this study we use optimization techniques and sensitivity analyses to provide more rigorous, quantitative connectionist models of the functional interactions between the brain areas involved in detecting and orienting attention towards threat. A toolkit has been developed using flexible neural network modeling with automated parameter estimation. A sensitivity analysis is provided within the framework to identify significant model parameters and better understand model dependencies. These studies emphasize the importance of fitting the models to behavioral reaction time and brain activation data. They also show that the specific architecture of the model, and the numerical precision in the model parameters is important in determining an acceptable fit of experimental data, and that it is not the case that any model will work given the appropriate set of connection strength parameters.

Keywords: Neural Network, Least-Square Minimization, Analysis of Variance

1. INTRODUCTION

Our survival depends in part on being able to detect a threatening stimulus that occurs outside the focus of attention and to redirect attention towards the threat so it can be dealt with (Bishop 2008; Corbetta & Shulman 2008; Norman & Shallice 1989). The ability of threats to capture and hold attention can also have an impact on mental health. For example, hypervigilance towards threat is thought to play an important role in the etiology and maintenance of many anxiety disorders (Bishop 2008; MacLeod et al. 2004; Mogg & Bradley 2004). Yet despite its importance to survival and mental health there remain significant gaps in our understanding of this fundamental cognitive process (Corbetta & Shulman 2002; Öhman 2000; Phelps & LeDoux 2005).

Several studies have shown that threats are better at capturing and holding attention than non-threatening stimuli (Bar-Haim et al. 2007; Bishop 2008; Cisler et al. 2009; Öhman 2000, 2005). The enhanced ability of threats to capture attention is evidenced by faster behavioral reaction times for threatening than non-threatening stimuli. Interestingly, the bias in attentional capture has been more difficult to demonstrate than the ability of threats to hold attention (see Bar-Haim et al. 2007; Bishop 2008; Cisler et al. 2009; Wyble et al. 2008). The neurophysiological underpinning of the attentional bias towards threat has also been the focus of a number of studies. Although several key brain areas have been implicated, such as the amygdala and insula, (Öhman 2000, 2005; Phelps & LeDoux 2005), a detailed understanding of how these areas interact with the brain areas involved in perception, response generation, and attention is lacking.

One approach to investigating the interactions between these brain areas is to combine experimental work with computational modeling. In this approach the computational models provide rigorous tests of hypotheses generated by the experimental work, and importantly, should provide novel predictions that can be tested in future work (e.g., Yeung et al. 2004). Little work has been done applying this approach to studying the attentional bias towards threats, and the few that have relied on qualitative fits of the experimental data rather than quantitative fits (Armony & LeDoux 2000; Dowman & ben-Avraham 2008; Wyble et al. 2008). In this study we describe our efforts at applying a connectionist model to quantitatively fit behavioral and brain activation data obtained in our studies of the attentional bias towards threats to the body (somatic threats).

An important focus of this work involved comparing model architectures simulating the different functional interactions between the brain areas thought to be involved in threat detection and orienting. To accomplish this we applied optimization techniques and sensitivity analyses to allow more rigorous, quantitative comparison of the different architectures and to explore the properties of the parameter space.

We explain the experimental setting in section 2 and follow in section 3 with the modeling. We present two different brain architectures and the model calibration results in section 4 with the follow-up sensitivity analysis in section 5. We end with a discussion of future directions.

2. EXPERIMENTATION

In our somatic threat studies (Dowman 2007a, 2007b) subjects performed two tasks: a visual color discrimination task and a somatic intensity discrimination task alternating in random order within the same session. The visual discrimination task consisted of indicating whether a red or a yellow LED was lit, and the somatosensory discrimination task consisted of indicating whether a high or low intensity electrical stimulus was delivered to the sural nerve at the ankle. A symbolic cue given at the beginning of each trial signaled which of the two tasks was forthcoming. The target stimulus was correctly cued on a randomly determined 75% of the trials (validly cued condition) and incorrectly cued on the remaining 25% of the trials (invalidly cued). The subject was instructed to focus his/her attention on the cued target stimulus, but to respond to the target regardless of whether or not it was correctly cued. Note that in the validly cued condition the target stimulus was presented within the subject's focus of attention, and in the invalidly cued condition the target was presented outside the focus of attention.

Two different sural nerve electrical stimulus intensities were used. In Dowman (2007a) both were strong and threatening (one at pain threshold and the other moderately painful), and in Dowman (2007b) both were weak and non-threatening. The attentional bias towards the somatic threat was evidenced in our experimental studies by the reaction time difference between the validly and invalidly cued conditions (validity effect) being smaller for the threatening somatic than the non-threatening somatic or visual target stimuli (Dowman & ben-Avraham 2008). (Reaction time differences due to stimulus intensity and sensory modality precluded a direct comparison between the threatening and non-threatening target stimuli). The smaller validity effect is consistent with the idea that threat is better able to capture and shift attention than non-threatening stimuli.

Electrophysiological measurements obtained during these experiments revealed three brain areas that appear to play an important role in detecting and orienting attention towards somatic threats. That is, for the threatening sural nerve target stimuli these brain areas exhibited greater activation when they were presented outside the focus of attention (invalidly cued) than when they were presented within the focus of attention (validly cued) (Dowman, 2007a, 2007b; Dowman & ben-Avraham 2008).

The electrophysiological data suggest that somatic threats are detected by somatic threat detectors located in the dorsal posterior insula. The threat detector activity is in turn monitored by the medial prefrontal cortex, which then signals the lateral prefrontal cortex to shift attention towards the threat (Dowman & ben-Avraham 2008). The greater activation of the somatic threat detectors in the invalidly than the validly cued condition suggests that that the ability of somatic threats to capture attention is greater when they are presented outside the focus of attention, and is consistent with the smaller reaction time validity effect for threatening targets observed in the reaction time data.

3. MODELING

We further examined the threat detection and orienting hypothesis using artificial neural network modeling (Dowman & ben-Avraham, 2008). The model was based on the work of J.D. Cohen and co-workers on response conflict (Botvinick et al. 2001; Yeung et al. 2004). The response conflict modeling studies, in conjunction with behavioral and functional imaging measurements, have provided convincing evidence that the medial prefrontal cortex is involved in monitoring situations that require a change in attentional control (e.g., response errors, response conflict, unattended threats) and signals the lateral prefrontal cortex to make the change. We modified the Yeung et al. (2004) model by replacing the response conflict component with threat detectors.

We compared several different model architectures in order to test the different physiologically feasible functional interactions between the brain areas responsible for detecting and orienting attention towards somatic threats. The architecture that provided the best qualitative fits of the reaction time and brain activation data is shown in Figure 1.

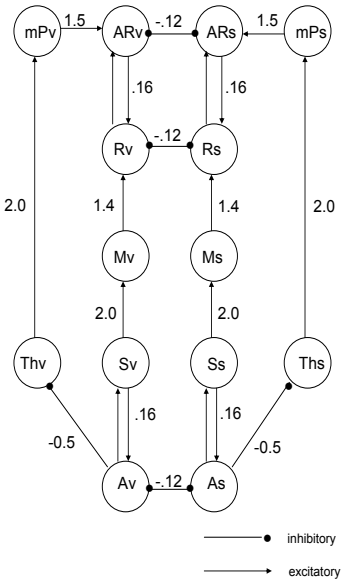

Figure 1: Artificial Neural Network Model of the Threat Detection and Orienting Hypothesis.

The model includes two stimulus-response pathways corresponding to the two tasks, where the Ss-Ms-Rs nodes and their connections simulate the somatosensory intensity discrimination task, and the Sv-Mv-Rv nodes and connections simulate the visual color discrimination task. The S and M nodes correspond to brain areas involved in early and late sensory processing, respectively, and the R nodes correspond to brain areas involved in the response. Note that the

model does not attempt to simulate the discrimination task performance in each stimulus modality, but rather only the reaction time differences for threatening vs. non-threatening target stimuli. The lateral prefrontal cortex areas controlling attention are simulated by attention nodes, one for each of the visual and somatosensory sensory nodes (Av and As, respectively), and one for each of the visual and somatosensory response nodes (ARv and ARs, respectively). The threat detectors for the visual and somatosensory systems are simulated by the Thv and Ths nodes, respectively, and the medial prefrontal cortex is simulated by the mPs and mPv nodes for the visual and somatosensory systems respectively.

The activation for each node was computed over 55 cycles. During the first 5 cycles external inputs were added to the As or Av nodes to simulate the allocation of attention as directed by the cue. The validly cued condition was simulated by adding an external input of 1.0 to the somatosensory attention node (As) and 0.0 to the visual sensory attention node (Av). The invalidly cued condition was simulated by adding an external input of 0.0 to the somatosensory attention node (As) and 1.0 to the visual sensory attention node (Av). During the stimulus cycles (cycles 6-10) external inputs were added to the somatosensory sensory node (Ss) to simulate the presentation of the somatosensory target stimulus. A threatening somatosensory stimulus was simulated by also adding external input to the somatosensory threat detector node (Ths) during the stimulus cycles. Due to the symmetry in the model, targets were only presented on the somatosensory side.

In the remaining 45 cycles, activation was allowed to spread through the model. For a model with *M* nodes, the activation levels of the nodes were computed using the following activation function:

$$A_i = \frac{1}{1+e^{(4-N_i)}}, \quad (1)$$

where A is a column vector containing M elements which represent the activation levels for all nodes in the model for the *i*th cycle. N_i was defined as:

$$N_i = N_{i-1} + (W * A_{i-1}) - (N_{i-1} * \delta), \quad (2)$$

N is also a column vector of size M, δ is a scalar decay constant, and W is an M x M weighted connection matrix. N_{i-1} represents the value of N during the cycle prior to i. Similarly, A_{i-1} represents the value of A during the preceding cycle. The product of the weighted connection matrix and the last known activation values $(W * A_{i-1})$ accounts for the input to each node due to its incoming connections. A_0 and N_0 were both null vectors initially.

The reaction time was defined as the cycle where the response node activation equaled 0.2. Reaction time was converted to milliseconds using the following function:

$$Reaction\ Time = 20c + 500, \quad (3)$$

where c is the cycle at which the activation level of the response node equals 0.20, 20 is an estimate of the number of milliseconds per cycle (based on the brain activation data), and 500 is a constant that accounts for perceptual and decision processes that are not accounted for by the neural network model (Dowman & ben-Avraham, 2008).

Dowman & ben-Avraham (2008) tested a number of different model architectures simulating different functional interactions between the brain areas thought to be involved in detecting and orienting attention towards somatic threats. As noted above, the architecture shown in Figure 1 demonstrated the best qualitative fits with the experimental reaction times and brain activations. Interestingly, this architecture led to the prediction that the attentional bias towards somatic threats will only be observed when the threat is presented outside the focus of attention (invalidly cued) and not when it is presented within the focus of attention (validly cued). As noted earlier, we could not directly test this hypothesis using the sural nerve stimuli because of the stimulus intensity confound. However, this prediction was recently confirmed in a study using pictures of somatic threats (Dowman et al. 2010), where the neutral and somatic threat target stimuli were matched for stimulus intensity and hue.

Dowman & ben-Avraham (2008) used the same set of connection strengths for all of the architectures (see Figure 1). These connection strengths were based on those published by Yeung et al. (2004), and were modified manually to provide acceptable qualitative fits of the data. The comparisons were straightforward given that many of the architectures could not simulate the direction of change in both the reaction time and brain activation data. It is possible of course, that had we chosen a different set of connection strength parameters that another architecture would have fit the data better. Therefore, a much better approach for model architecture comparison would be to use optimization techniques to find the best fit connection strength parameters for each of the architectures. Optimization techniques have the added advantage of allowing us to search for the best quantitative fit of the experimental data, something that is not feasible when the connection strengths are adjusted manually.

It is also important to perform sensitivity analyses to explore the parameter set. Of particular interest is determining whether the fit is dependent on a small range of connection strength values, or whether a wide range of combinations produce a good fit. Together, the optimization techniques and sensitivity analyses will provide a more rigorous quantitative comparison of the different architectures. Importantly they will allow us to determine if the architecture is important in fitting the data, or whether any architecture can be made to fit the data given the right set of connection strength parameters. Clearly the former outcome is of much greater interest in using the models to help determine the functional interactions between these brain areas.

4. MODEL CALIBRATION

Our previous effort involved manually adjusting connection strength parameters to provide acceptable fits, and then using these parameters to compare the different model architectures. This process was slow, tedious and at best led to rough qualitative fits of the data. More recently we developed a MATLAB®-based toolkit that provides automated calibration of connection strengths using optimization techniques (Lowenstein 2010). The optimization involved minimizing

$$J(W) = \frac{1}{P}\sum_{i=1}^{P}\left(\frac{e_i - m_i}{n_i}\right)^2, \quad (4)$$

where W is the matrix of connection strengths, P corresponds to the number of statistics that are being fit, e_i is the experimental value of a given statistic, m_i is the corresponding modeled value (which depends on W), and n_i is a normalization factor which ensures that each statistic contributes equally to the cost function. For our purposes, this normalization can be accomplished by setting n_i equal to e_i.

Within the toolkit, the Nelder-Mead simplex method is used for the minimization of Eq. (4) (Nelder & Mead 1965). Nelder-Mead has previously been shown to be effective in parameterizing connectionist models (Bogacz & Cohen 2004). A benefit of Nelder-Mead is that no gradient information is needed and minimization is based solely on function evaluations using a simplex that changes at each iteration based on the best point found. The function to be minimized can be non-differentiable, non-convex, or even discontinuous. This is an attractive feature of the toolkit because it allows for a general framework for the model calibration. Thus changes made to the simulation tool itself will have little, if any impact on the calibration process.

For each model, the five optimal parameters sought were *asr*, *at*, *in*, *smr*, and *tmar* (see Figure 2). Nelder-Mead is well known to be a local optimization method that can be highly dependent on an initial simplex. Consequently, multiple optimizations are usually required to better search the design space. Here 20 optimization runs were obtained for each model. To find the starting values for each Nelder-Mead optimization run, 1000 connection strength parameter sets were randomly chosen and the fits calculated. The set with the lowest $J(W)$ value was used as the starting values.

4.1 Numerical Results

First we determined whether the original architecture (Figure 1) provides a good, quantitative fit of the reaction time and brain activation data. As noted above, the stimulus intensity confound prevented us from directly comparing behavioral reaction times obtained for the non-threatening and threatening somatic target stimuli. For modeling purposes we approximated the stimulus intensity confound-free invalidly cued threatening somatic target reaction time by multiplying the increase in the invalidly cued reaction time relative to the validly cued condition for the threatening somatic target (i.e., [invalidly cued – validly cued]/validly cued) to the validly cued non-threatening somatic target reaction time. Also, owing to the uncertain relationships between the scalp potentials used to measure the brain activations, the underlying brain activity, and the activation function in the model, we used the percent change in the electrophysiological measurements in the invalidly cued condition relative to the validly cued condition (i.e., {[invalidly cued – validly cued]/validly cued} * 100). As described in detail in Dowman & ben-Avraham (2008) the electrophysiological measures of the threat detector activation also include activation of the adjacent sensory area. Hence, this activity was modeled by combining the activations of the Ss and Ths nodes. The scalp potential measurements do not provide acceptable isolation of the medial and lateral prefrontal cortex activities, hence they were not included in the modeling studies.

The original model was able to provide excellent fits of the reaction time data (least-square error ~ 1.0e-9), but could not also fit the activation data (least-square error = 1.0). The failure of the original model to fit reaction time and activation data was because it could not account for the lack of change in the Ss node across the validly and invalidly cued conditions for the non-threatening somatic targets, as was originally pointed out by Dowman & ben-Avraham (2008). Rather, in that model the Ss node was smaller in the invalidly than the validly cued conditions. Our experimental studies have reported that brain areas involved in later sensory processing do show this attention effect (Dowman 2007a). Hence, we altered the model architecture by connecting the sensory attention nodes (As) to the middle layer (Ms), where the latter simulates the later stage of sensory processing. This model is shown in Figure 2 below. The result was a much better fit of the reaction time and activation data. Specifically, of the 20 best-fit parameter sets obtained from the Nelder-Mead algorithm, three gave a least-square fit to within the measurement error (i.e., least-square error ~ 1.0e-5, modeled reaction times within 5 milliseconds).

Dowman & ben-Avraham (2008) also compared architectures where the threat signal from the medial prefrontal cortex (mPs) to the response attention node (ARs) to one that has the threat signal going to both the sensory (As) and response attention nodes. This architecture is shown in Figure 3. The latter architecture is more consistent with the known anatomical connections between the medial and lateral prefrontal cortices (Miller & Cohen 2001). Dowman & ben-Avraham (2008) could not find any difference between the two architectures. We re-ran this simulation using the best-fit connection strength parameters for each architecture to see if this would make a difference, and

indeed it did: the architecture sending the threat signal to both the sensory and response attention nodes provided a noticeably better quantitative fit of the reaction time and activation data.

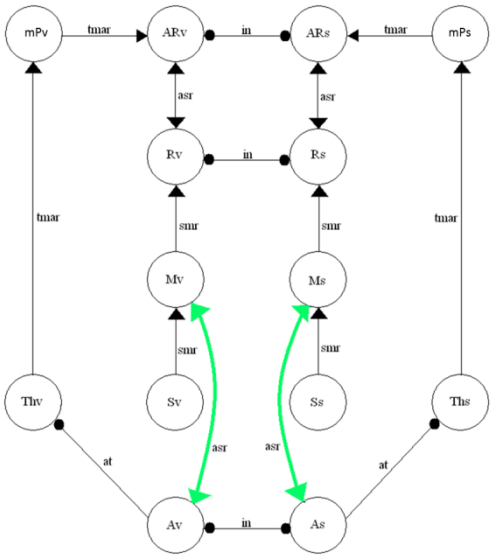

Figure 2: Modified Architecture to Sensory Attention Applied to the Late Sensory Processing Stage.

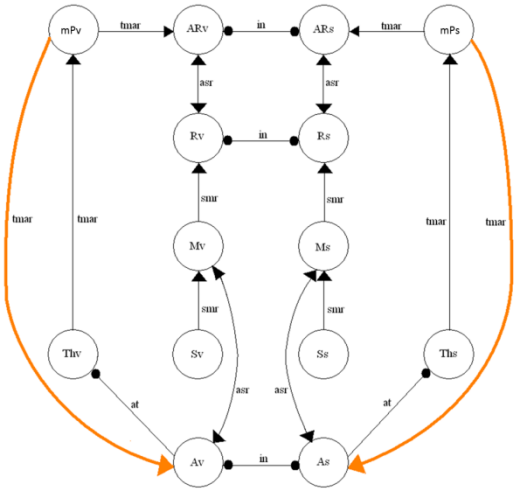

Figure 3: Architecture to Account for the Threat Going to Both the Sensory and Response Attention Nodes

For this model, 7 of the 20 optimization runs produced acceptable fits. The least-square errors for these parameters were an order of magnitude smaller than the previous model (2.0702e-6 vs. 6.8900e-5, respectively) (Table 1), where the modeled reaction times were within 1 millisecond of the experimental data (Table 2) and the modeled percent change in brain activations equaled the experimental data.

We also examined one of the architectures that Dowman & ben-Avraham (2008) reported was unable to fit the reaction time and activation data. In this architecture the threat signal from the medial prefrontal cortex was sent to the sensory attention node instead of the response attention node. The same result was obtained here when trying to quantitatively fit the reaction time and threat detector/sensory activation data (least-square error ~1.0e1).

Table 1: Best Fit Model Parameters and Least-Square Error for Architecture in Figure 3

Parameter	Optimal Value
asr	0.5640
at	-0.2119
in	-0.4327
smr	0.0256
tmar	0.2073
J(w)	2.0702e-6

Table 2: Comparison on Experimental and Model Reaction Times.

Experimental Condition	Reaction Time (milliseconds)
Valid Non-Threat Exp.	694.2
Valid Non-Threat Model	694.8
Invalid Non-Threat Exp.	826.1
Invalid Non-Threat Model	826.1
Valid Threat Exp.	694.2
Valid Threat Model	693.4
Invalid Threat Exp.	770.5
Invalid Threat Model	770.7

Interestingly, we consistently found that almost all of the architectures that we tested could provide excellent fits of the reaction time data when the best fit connection strengths were used (least-square errors $\leq$ 1.0e-8). However, only the 2 architectures described here provided acceptable fits of the reaction time and brain activation data. Clearly, the brain activation data appears to provide critical constraints on the model.

5. SENSITIVITY ANALYSIS

The optimization results described above show that the model architecture is critical in obtaining fits of the reaction time and brain activation data. We next sought to determine the range of connection strength parameters that produced acceptable fits of the data. The mean $\pm$ SD of the connection strength parameters for the architecture producing the best fit of the data (see Figure 3) is shown in Figure 4. The 7 optimization runs that produced acceptable fits (least-square errors = 2.1–7.5e-6) were all tightly clustered around the same values. This was not the case for the 13 runs that

resulted in unacceptable fits (least-square error = 6.6-9.1e-3). Recall that the starting values for the Nelder-Mead optimization were determined by 1000 iterations of randomly selecting parameter values and computing the least-square error, and using the values that produced the best fit as the starting point in the optimization. This strategy reduces the probability that the optimization will always converge on the same local minimum. Hence, the tight coupling of acceptable fit parameters around the same values strongly suggests that range of best-fit connection strengths is very narrow.

This result was confirmed with a sensitivity analysis. When developing and studying mathematical models, it is common in practice to perform a sensitivity analysis to gain a deeper understanding of the model behavior, regardless of whether optimization is part of the design process. Analysis of variance (ANOVA) is one approach to studying the impact of changes in model parameters on model output. Specifically, ANOVA may reveal that some parameters have little effect on the overall model while others have a profound effect. In such cases, certain insignificant parameters can be set to a reasonable value while optimization can be done to fit the sensitive parameters and thereby reduce the problem size for the least-squares problem. This approach can also determine the specificity of the connection strength parameters. That is, does each connection have to be within a tight range for the model to work, or can changes in one connection be offset by a change(s) elsewhere in the model. This analysis can have a significant impact on interpreting the functional significance of the connection strength values. A benefit of ANOVA is that only sets of parameters and output are required as opposed to needing any derivative information.

ANOVA compares the ratio of the variation between sample means to the variation within each sample. The starting point for the procedure is to sort each parameter into groups. Analysis is done by considering changes in a response (here the least-squares error) as the group changes. Specifically, ANOVA is a hypothesis test with null hypothesis,
$H0 : \mu 1 = \mu 2 = \mu 3 = \cdots = \mu k$, where k is the number of experimental groups. Each μ represents the mean of the single parameter, often called a factor, that is being found by the values in each experimental group. When rejecting the null hypothesis, the alternative hypothesis states that at least one mean is different from another, however it does not specify which one. The experimental groups are different equally spaced intervals for a single variable. The ANOVA examines the source of variation by finding the sum of squares of deviation from the mean for each of these groups. Using a statistical F test, the procedure is able to determine whether or not at least one mean is deviating from the others. The F test will produce a p-value; If this value is below a significance of 0.05 then the null hypothesis is rejected.

The model calibration experiments described above revealed that small changes in even the third decimal place of the connection strengths could strongly impact the overall least-square error and result in a poor model fit. For the sensitivity study presented here, tight bounds were place on each parameter based on the best point found. We provide the details of the sensitivity analysis for the architecture in Figure 3, since it provided the overall best fit to the experimental data. For the sensitivity analysis, we chose the best of the 7 parameter sets that provided acceptable fits. The bounds are shown in the second and third columns of Table 3 below. For the analysis, each parameter was divided into 8 equally spaced groups and 500 values of each parameter were chosen via a Latin hypercube sampling, giving 2,500 parameter sets.

The Kolmogorov-Smirnov normality test was applied to the response variable (here the least-squares error) and we found the data was not generated from a normally distributed population. Thus, the non-parametric ANOVA method, the Kruskal-Wallis test, was used to calculate the corresponding p-values, shown in the last column of Table 3. Three parameters had values close to zero (*asr*, *at*, and *in*) indicating that they are significant in the modeling process. However, *smr* has a p-value of 0.052, which is very close to our level of significance 0.05, and we still can consider it to be a sensitive parameter. The parameter *tmar* was identified as insensitive and this was also evident in the values identified by the optimizer for the seven best parameter sets found during optimization. For the significant parameters, the standard deviation was always less than 0.03 but for *tmar* it was 0.07, indicating that a range of values would still lead to reasonable fitting to the data. These results intuitively make sense because *tmar* and *smr* are feed-forward connections, the rest are bidirectional. Clearly the positive feedback associated with a bidirectional connection will make it much more sensitive to change than a feed-forward connection. Furthermore, *at* is particularly sensitive since it is largely and only responsible for the brain activation fit.

Table 3: Lower and Upper Bounds for Parameter Study on Architecture in Figure 3

Parameter	Lower Bound	Upper Bound	p-value
asr	0.5	0.6	≈0
at	-0.3	-0.2	≈0
in	-0.5	-0.3	≈0
smr	0.02	0.03	0.052
tmar	0.2	0.29	0.493

To this end, an interval plot can provide a deeper insight into how the response values are distributed. We show these for *smr* and *in* in Figures 5 and 6. Here, the shaded, red dots correspond to points values of *J(W)*

while the average for each group is shown with a blue '⊕' . For *smr*, the average values are relatively constant, with small fluctuations across the groups, but there is actually a broad range of response values within a group. For *in*, the average values change significantly across the groups, which is expected since *in* was identified as a sensitive parameter. It is important to note that of the 2,500 parameter sets randomly chosen from within the bounds given in Table 3 the least-square errors were greater than 1.0e-2, which is four orders of magnitude greater than the best-fit values.

The sensitivity analysis for *in* produced an unexpected result. Sensitivity analysis is often used to guide the starting parameter set values for optimization. That is, the factor (parameter range) showing the best fit values is chosen in the optimization. However, sensitivity analysis suggests that the best fit point(s) would lie within factor 7 whereas the optimal value for *in* actually falls within factor 2 (-0.4327). The interval plots (Figure 6) reveal that the neighborhood around this point is considerably small thus requiring high accuracy in the optimization process. This implies that caution must be applied when interpreting the sensitivity analysis results for models with a very narrow range of best-fit values. An important clue that the sensitivity analysis results may not provide meaningful information on selecting starting points and/or bound constraints for the optimization was that even the best $J(W)$ values were four orders of magnitude greater than the optimal value.

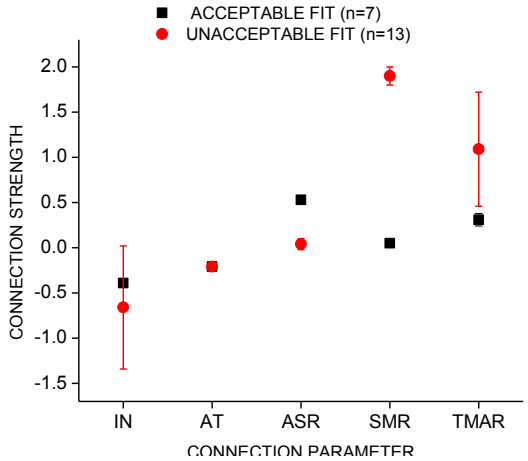

Figure 4: Mean ± SD Connection Strength Parameters For the Best-Fit Architecture.

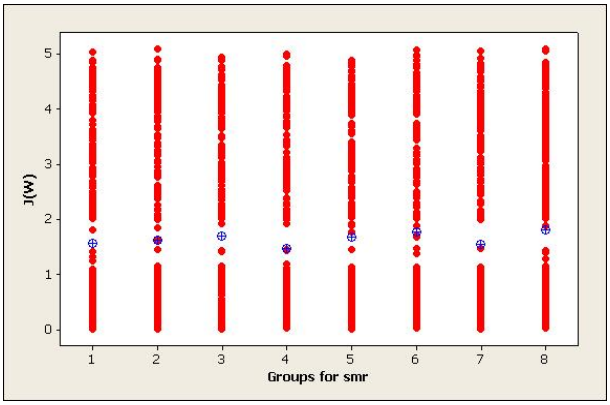

Figure 5: Range of Response Values Across Groups for *smr*

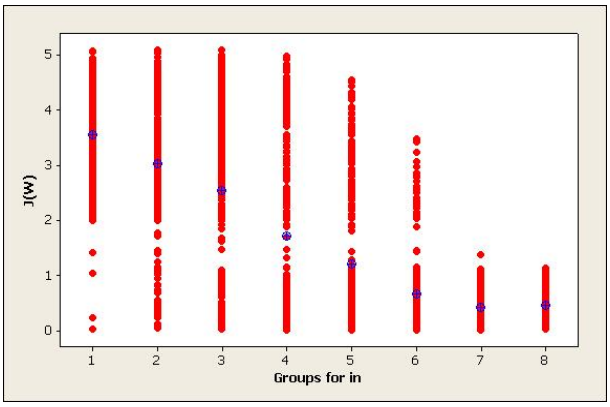

Figure 6: Range of Response Values across Groups for *in*

6. DISCUSSION

We have developed a flexible toolkit to develop and test artificial network models of the brain mechanisms for detecting and orienting attention towards threats to the body. Using optimization techniques were able to provide excellent quantitative fits of behavioral reaction time and brain activation data. These studies demonstrate that the model architecture is critical in producing good quantitative fits of the reaction time and brain activation data. Indeed, of the several models examined by Dowman & ben-Avraham (2008), only 2 provided acceptable fits. However, essentially all of the architectures could fit the reaction time data given optimal set of connect strength parameters. Clearly, including the brain activation data is critical in obtaining meaningful results in this type of work.

The sensitivity analysis suggests that only a very narrow range of connection strength parameters will fit the data. This implies that for fits of reaction time and brain activation at least, it is not case that the acceptable fits are an artifact of having a large number of parameters to fit the data. These results strongly suggest that the sensitivity analysis should not be used to determine the starting parameter values and ranges when the range of optimal values is very narrow. Future

work will include understanding the interaction of parameters.

Future experimental studies are aimed at testing predictions derived from the model. Of most interest is the prediction that the attentional bias towards threats are only seen when the threat is presented outside the focus of attention (Dowman et al. 2010). We are also performing modeling studies to determine if the model can simulate the attentional bias towards threats that have been reported using other experimental paradigms (e.g,. Koster et al. 2007).

ACKNOWLEDGMENTS

This research work was supported by National Science Foundation UBM (Undergraduate Biology-Mathematics) Grant DBI-0926568 and the Clarkson University Honors Program.

REFERENCES

Armony, J.L., & LeDoux, J.E. (2000) How danger is encoded: Toward a systems, cellular and computational understanding of cognitive-emotional interactions in fear. In: M.S. Gazzaniga (Ed.) *The New Cognitive Neurosciences*, Cambridge Massachusetts: The MIT Press, pp. 1067-1079.

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M.J. & van LJzendoorn, M.H. (2007) Threat-related attentional bias in anxious and non-anxious individuals: A meta analytic study. *Psychological Bulletin, 133,* 1-24.

Bishop, S.J. (2008) Neural mechanisms underlying selective attention to threat. *Annuals of the New York Academy of Sciences, 1129,* 141-152.

Bogacz, R. & Cohen, J.D. (2004) Parameterization of connectionist models. Behavior Research Methods, Instruments, & Computers. 36, 732-741.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., & Cohen, J.D. (2001) Conflict monitoring and cognitive control. *Psychological Review, 108,* 624-652.

Cisler, J.M., Bacon, A.K. & Williams, N.L. (2009) Phenomenological characteristics of attentional biases towards threat: A critical review. *Cognitive Therapy and Research, 33,* 221-234.

Corbetta, M. & Shulman, G.L. (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 31,* 201-215.

Corbetta, M., Patel, G., & Shulman, G.L. (2008) The reorienting system of the human brain: From environment to theory of mind. *Neuron, 58,* 306-324.

Dowman, R. (2007a). Neural Mechanisms Of Detecting and Orienting Attention Towards Unattended Threatening Somatosensory Targets. I. Modality Effects. *Psychophysiology, 44,* 407-419.

Dowman, R. (2007b). Neural Mechanisms Of Detecting and Orienting Attention Towards Unattended Threatening Somatosensory Targets. II. Intensity Effects. *Psychophysiology, 44,* 420-430.

Dowman, R., Quin, J., & Sieg, E. (2010) Mechanisms Underlying the Capture of Attention by Somatic Threats. American Psychological Society 22nd Annual Meeting, Boston MA May 27-30.

Dowman, R., & ben-Avraham, D. (2008) An artificial neural network model of orienting attention towards threatening somatosensory stimuli. *Psychophysiology, 45,* 229-239.

Koster, E.H.W., Crombez, G., Verschuere, B., Vanvolsem, P., & De Houwer, J. (2007) A time-course analysis of attentional cueing by threatening scenes. Experimental Psychology, 54, 161-171.

Lowenstein, K. (2010) *A Toolkit for Developing Neural Network Models of How the Brain Detects Threat*. Honors Thesis. Clarkson University.

MacLeod, C., Campbell, L., Rutherford, E., & Wilson, E. (2004) The causal status of anxiety-linked attentional and interpretive bias. In: *Cognition, Emotion and Psychopathology. Theoretical, Empirical and Clinical Directions* (J. Yiend, Ed.) New York: Cambridge University Press, 172-189.

Miller, E.K., & Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Reviews of Neuroscience, 24,* 167-202.

Mogg, K., & Bradley, B.P. (2004). A cognitive-motivational perspective on the processing of threat information in anxiety. In: *Cognition, Emotion and Psychopathology. Theoretical, Empirical and Clinical Directions* (J. Yiend, Ed.) New York: Cambridge University Press, 68-85.

Nelder, J.A., & Mead, R. (1965). A Simple Method for Function Minimization. *Computer Journal*, 7,308-313.

Norman, D.A., & Shallice, T. (1986) Attention to action. In: R.J. Davison, G.E. Schwartz, & D. Shapiro (Eds.) *Consciousness and Self-Regulation. Advances in Research and Therapy.* Plenium Press: New York, pp. 1-18.

Öhman, A. (2000) Fear and anxiety: evolutionary, cognitive, and clinical perspectives. In: M. Lewis & J.M. Haviland-Jones (Eds.) *Handbook of Emotions*, New York: Guilford Press, pp. 573-593.

Ohman, A. (2005) The role of the amygdala in human fear: automatic detection of threat. *Psychoneuroendocrinology, 30,* 953-958.

Phelps, E.A., & LeDoux, J.E. (2005) Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron, 48,* 175-187.

Wyble, B., Sharma, D., & Bowman, H. (2008) Strategic regulation of cognitive control by emotional salience: A neural network model. Cognition and Emotion, 22, 1019-1051

Yeung, N., Botvinick, M.W. & Cohen, J.D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111,* 931-959.