MODELLING THE SLOW EMOTIONAL STROOP EFFECT: SUPPRESSION OF COGNITIVE CONTROL

BRADLEY WYBLE[A] DINKAR SHARMA[B] HOWARD BOWMAN[A]

[A] *Centre for Cognitive Neuroscience and Cognitive Systems, Computing Laboratory, University of Canterbury Canterbury, Kent CT2 7NP, UK*

[B] *Centre for Cognitive Neuroscience and Cognitive Systems, Psychology Department, University of Canterbury Canterbury, Kent CT2 7NP, UK*
bw5@kent.ac.uk, D.Sharma@kent.ac.uk, hb5@kent.ac.uk

Our connectionist model provides a theoretical explanation for the existence of slow and fast emotional Stroop effects, and depicts them as independent but interacting phenomena. We build upon previous modelling work by [1] and [2] among others, and incorporate data that suggest a functional division of the Anterior Cingulate Cortex (ACC) into Cognitive and Affective Divisions. This work suggests that slow emotional Stroop effects are caused by activation of the affective portion of the ACC, which inhibits the Cognitive division, reducing top-down cognitive control on the subsequent trial.

1. Introduction

1.1. *Modelling the Stroop Effect*

Stroop interference refers to increases in reaction time to name the ink color of a word when that word is itself the name of a different color [3]. There are three types of trials in the Stroop paradigm: congruent, in which the ink and color word match, neutral in which the word is not a color name, and incongruent, in which the ink and color name do not match. Incongruent trials show interference relative to neutral trials, while congruent trials show facilitation. In this paper we propose a connectionist model of the Stroop effect, built upon previously existing models [1,2]. that extends their functionality to include fast and slow emotional Stroop effects in a manner that is consistent with recent imaging data about the functionality of the Anterior Cingulate Cortex (ACC).

1.2. *Anterior Cingulate Cortex and Cognitive Control*

The Anterior Cingulate Cortex(ACC) is thought to assist in the mediation of cognitive tasks, especially those that involve mutually incongruent stimuli pairs [4]. Furthermore, it is thought that the ACC is able to detect conflict between the incompatible responses of such tasks and elevates cognitive control from one trial to the next to improve performance in such difficult tasks [2].

There also exists evidence that the rostral extreme of the ACC is activated in emotional states and by affective stimuli. Furthermore the two portions of the ACC, one mediating cognitive and the other affective operations, are mutually inhibitory [5,6]. Our model proposes that it is the interaction between these two halves of the ACC that is responsible for the slow emotional Stroop effect as will be discussed below.

1.3. *Slow Emotional Stroop Effect*

The emotional Stroop task involves presenting an emotionally charged word that is written in various ink colours. Subjects are asked to ignore the word and respond to the ink colour. The typical finding is one of longer response latencies to name the ink colour when the word is emotional, usually negative, rather than a control neutral word [7]. Such attentional biases have been argued to reflect the capture of attention by emotional stimuli. In general this is thought to reflect a shift in processing resources towards emotional stimuli. In addition, such processing is often described as automatic, fast acting and even preattentive ([8]). Evidence using paradigms other than the modified Stroop task have also been used to support these claims [9].

Although much of the research has tended to describe the impact of emotional stimuli in these terms, more recently there has emerged evidence in support of the conclusion that there are relatively slow effects of emotionally negative stimuli. One such data set comes from emotional Stroop studies that have mixed negative and neutral trials [10]. They presented seven Stroop trials: a negative stimulus is shown on the first trial followed by six trials of neutral stimuli. They found that emotional trials are performed at the same speed as neutral trials. However, the following trial had a larger reaction time despite being neutral. This effect returned back to baseline by the second post-emotional word trial. This research highlights the idea that negative words do not capture attention in a relatively fast manner but produce their effects more slowly on attentional components subsequent to detection. Slow effects that operate across trials can also explain the typically larger interference found for emotional stimuli when these stimuli are blocked compared to when they are mixed randomly with other neutral stimuli [11].

Emotional Stroop effects that depend on inter-trial effects can be called slow effects, while those that depend on intra-trial effects can be called fast effects[10]. Our model, derived from previous work in the field, demonstrates

both fast and slow components of the emotional Stroop effect, by assuming that fast effects are due to automatic vigilance [12] and that slow effects are the result of a temporary redistribution of attentional resources in response to the emotional input. We call this latter process salience sensitive control, and view it as an automated process by which the mind of the subject is reconfigured for several hundred milliseconds in order to more readily detect threatening information in the environment.

2.Methods

2.1. *Neural Representation*

Our model is an extension and modification of the framework used by the conflict monitoring and cognitive control model of [2]. Several components are added to allow this network to model emotional Stroop effects, both fast and slow. The recurrent nature of the inter-layer links are changed to a purely feed-forward pattern to simplify the dynamics of the model.

Our implementation, pictured in Figure 1, follows the general pattern of the connectionist architecture for Stroop models proposed by [1]. Word and color information are processed in parallel along different streams, which converge on a common set of nodes. Strong weights within the word stream give word reading an inherent dominance over color naming. This default tendency is overcome through a task-demand system that provides an excitatory bias to the color naming stream. In our model, this task demand system also suppresses competing information in the word stream. This implementation is suggested by imaging work from [13] which shows that a decrease in left extra-striate cortex accompanied increased activation of dorsal cingulate cortex, an area thought to be involved in the processing of word form information [14].

Our model uses simple rate code representations. Outputs from each neural element are derived from thresholds combined with a sigmoid-like function, described in (2). Connections between elements are excitatory or inhibitory and are not modifiable. There are no synaptic delays. The activation function for the membrane potential of all neural elements is given by (1).

$MP_{(i,j,t)}$ is the membrane potential for neuron *i* of layer *j* at time *t*, affected by *Bias*, *Excite*, *Inhib* and *Leak* with a time constant *DT_VM*. *EE*, *EI* and *EL* represent reversal potentials.

$$MP_{(i,j,t)} = MP_{(i,j,t-1)} + DT\_VM_{(i)} * ((Bias_{(i)} + Excite_{(i,j,t-1)}) * (EE_{(i)} - MP_{(i,j,t-1)}) + Inhib_{(i,j,t-1)} * (EI_{(i)} - MP_{(i,j,t-1)}) + Leak_{(i)} * (EL_{(i)} - MP_{(i,j,t-1)})) \quad (1)$$

This membrane potential is used in the following output function which simulates a sigmoid function that bounds a neuron's output to the range [0,1), where θ and γ represent threshold and scaling parameters respectively.

$$Out_{(i,j,t)} = \frac{[MP_{(i,j,t)} - \theta_{(j)}]_+ * \gamma_{(j)}}{[MP_{(i,j,t)} - \theta_{(j)}]_+ * \gamma_{(j)} + 1} \quad \text{where } [x]_+ = \begin{matrix} x & if \quad x > 0 \\ 0 & otherwise \end{matrix} \qquad (2)$$

2.2. *Connectivity and Functionality*

The model we use is shown in figure 1. Color and word processing are located at the bottom left and right respectively. Nodes in these areas are activated directly by corresponding input units and are intended to represent levels of color and word-form processing. The word naming stream also contains nodes representing neutral and emotional words.

Nodes in these input layers are connected to the category layer, which is intended to represent activity in a semantic workspace. Connections from the word form area to the category layer are 2.5 times the value of those from the color area. This asymmetry insures a dominance of word reading over color naming. The relative strengths of the rest of the connections are not of direct theoretical import and will not be discussed in this limited format. All parameters were identical for all data sets addressed by this model.

Interference occurs within the category and response layers through uniform lateral inhibitory connections. Depending on levels of cognitive control, this inhibition may or may not be sufficient to ensure winner take-all dynamics within a layer.

Only the color nodes of the category layer project to the response layer, in which neurons represent activation of motor plans to commit verbal or button pressing responses. Activity in the response nodes builds slowly over the trial until one of the two crosses threshold, at which point a response is determined and the trial concludes. Activation of all nodes are reset to starting values, with the exception of task demand units and those representing the two halves of the ACC.

The color-naming task demand unit in this model excites color processing nodes, inhibits word form processing, and excites task relevant modality independent color nodes in the category layer. The word-naming task demand unit also foregrounds color processing and has similar but opposite connections but to the ink color and word form nodes. Input from the cognitive portion of the ACC, as well as putatively frontally mediated constant bias selects one of the two task demand units to be active, resulting in suppression of the other.

As in the model of [2], and in accord with studies that show ACC conflict monitoring at response production stages [15], a Hebbian energy product equation (3) over the two response nodes provides this measure of conflict.

$$E_{(i,j)} = \sum a_{(i)} a_{(j)} \quad \text{where} \quad a_{(i)} \text{ is the output of response unit i.} \qquad (3)$$

The output of this equation is multiplied by a weight and then added as an excitatory input to the node representing the cognitive division of the ACC.

Presentation of an emotional word is capable of activating the affective division of the ACC, which inhibits the cognitive portion of the ACC. This effect is slow relative to the performance of a single trial such that task demand fluctuates gradually across trials.

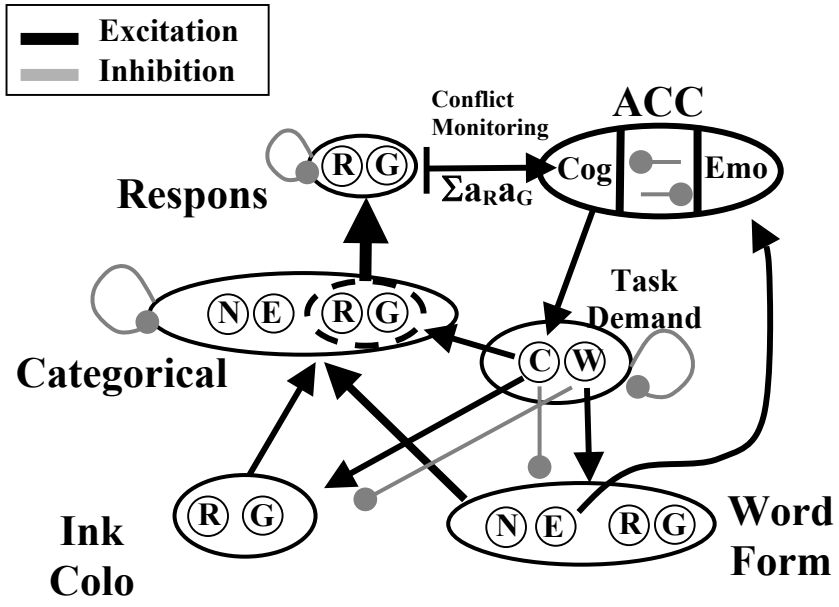

Figure 1. Diagram of the model used in these simulations

3. Results

We demonstrate the validity of this model by fitting a number of data sets.

3.1. *Standard Stroop Data*

This model is able to recreate the standard Stroop effects. Interference caused by incongruent trials produces a larger change in reaction time than does the facilitation caused by congruent trials. This is the classical pattern of asymmetry inherent in practically all standard Stroop studies. Additionally,

word reading tasks are faster than color naming tasks, and exhibit minimal interference or facilitation.

These effects are illustrated in Figure 2, for which human data from [16] are compared to the results from the model. This doubly asymmetrical pattern of results, is the consequence of saturation of the output function of the neurons projecting to the response layer in the case of word reading and congruent color-naming trials as described in [1].

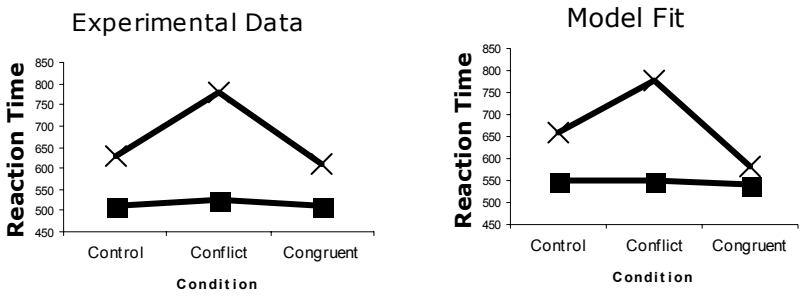

Figure 2. Comparison of simulated to empirical results for the standard Stroop paradigm. The X marked points represent color naming, while black squares represent word naming.

3.2. *Cognitive Control Data*

The data from [17], among others [15], make it clear that Stroop effects cannot be considered invariant across blocked and mixed experimental designs. The composition of a trial block affects how the trials within that block are processed, especially incongruent ones.

In [17], subjects were given blocks with different proportions of color trials and non-color neutral words in steps of 25%. Of the color words, half were congruent and half were incongruent. So a 75% color word block was composed of 37.5% congruent, 37.5% incongruent, and 25% neutral trials. What this work demonstrated is that as the proportion of color words was increased within a block, the amount of interference caused by an incongruent trial decreased, while facilitation remained essentially unaltered.

Likewise in our model, facilitation of congruent trials decreased slightly from 51 to 57 to msec across conditions, while interference decreased from 120 to 87 msec as the frequency of color trial increased from 25% to 75%. Our model exhibits this effect by virture of the conflict monitoring system.

An incongruent trial causes simultaneous activation of both color nodes in the response layer, resulting in a large conflict signal that activates the cognitive division of the ACC. This activation results in an increase in top down cognitive control, further exciting the color-naming area and suppressing the word-form area as the system becomes strongly biased towards color naming.

Therefore, as the proportion of incongruent trials increases from 12.5% to 25% to 37.5%, the chance of any incongruent trial being preceded by another incongruent trial increases, thereby reducing the average amount of interference.

3.3. *Emotional Stroop Effect*

In our framework of interference, fast and slow emotional Stroop effects are qualitatively distinct phenomena, representative of intra and inter-trial time scales respectively.

3.3.1. Fast Effects

In this model, fast effects are obtained by an implementation of automatic vigilance similar to that described by [12]. Emotional word corresponding to a particular anxiety can be given the ability to strongly activate the category layer representation and thereby interfere in the current trial in a similar manner as for incongruent color words. This can be done by either including a threat monitoring unit that adds an excitatory bias to a given emotional word form unit, or increasing the connection strength between these units and category layer representations. The degree of interference obtained can vary greatly by adjusting these parameters, so there is no much to be gained in claiming a fit of a particular data set.

Either by adding a threat monitoring unit that adds a excitatory bias to a given emotional word form unit, or increasing the connection strength between these units and category layer representations, the selected emotional words corresponding to a particular anxiety can be given the ability to strongly activate a category layer representation and thereby interfere in the current trial in a similar manner as for incongruent color words. The degree of interference obtained can vary greatly by adjusting these parameters so there is not much to be gained in claiming a fit of a particular data set.

3.3.2. Slow Effects

Negative emotional word input in this model has a dual effect by exciting both category layer representations and the affective division of the ACC. Even if the emotional input is insufficient to interfere at the category layer, the resulting activation of the affective portion of the ACC will suppress the cognitive division of the ACC, resulting in a slow reduction in cognitive control. On the following trial, the word-form area will be partially released from suppression allowing even neutral words to interfere at the category layer. It is important to note that the emotional word is not causing any direct interference itself, but rather is enabling the subsequent neutral word to interfere. This is the key point in the qualitative distinction between fast and slow emotional Stroop effects.

This model is able to fit results of the [10] both in time course and in effect size as shown in figure 3. Note that this effect is specific to words of a negative emotional valence, rather than a positive emotional valence.

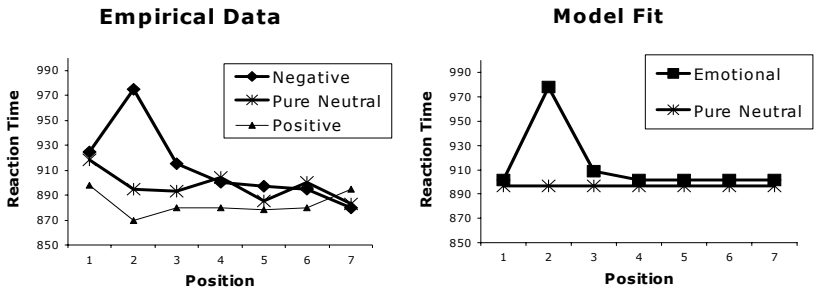

Figure 3. Data from McKenna and Sharma (2004) presented along with the model fit.

4. Predictions

4.1. *Salience Sensitive Control*

The primary prediction of this model and the theory that drives it concerns the fate of information presented to an unattended stream of input following emotional input of any kind. Our conceptualisation of salience sensitive control posits a temporary redistribution of attentional resources across multiple areas of processing following threatening input so as to facilitate the detection of further threatening information. This redistribution comes at the expense of the ongoing task, and the slow emotional Stroop effect is a hallmark of this process.

We predict that if subjects are given a word reading task, and a concurrent audio stream to be ignored, an emotional word presented within the word reading list should create a temporal window of several hundred milliseconds in duration during which input in the unattended audio stream is more deeply processed. This window should give the unattended audio stream an opportunity to prime the word naming task.

4.2. *Cognitive Control of Affective Input*

In this model, processing of word-form information is subject to cognitive control, and this includes processing of emotional salience. Consequently, elevating cognitive control by presenting incongruent trials immediately prior to an emotional word should reduce its ability to induce the salience sensitive control that interferes with task performance on the subsequent neutral trial.

From the data of [10], the sequence of trials: [Neutral, Emotional, Neutral], will result in a slower response time for the second neutral word than the first. Our model demonstrates that finding, and goes on to predict that for the

sequence [Incongruent, Emotional, Neutral], there will be no significant slowing of the neutral word. The increase in cognitive control evoked by the incongruent trial exerts a suppressive influence on the word form processing area, which prevents it from activating the affective portion of the ACC. A failure to find this effect will indicate that excitatory inputs to the affective division of the ACC derives from pre-attentive areas of representation that are immune to suppression by cognitive control.

5. Discussion and Conclusions

We present a theoretical framework that describes the slow emotional Stroop effect as an automated response to a threatening stimulus that exists to reorient attentional and cognitive resources towards detecting further threatening information. This position explains the slow emotional Stroop effect [10], but also fits very well with data describing the mutually inhibitory interaction of the cognitive and affective portions of the ACC [5,6].

The authors of [18] make a strong empirical case that emotional and color Stroop effects result from fundamentally distinct processes. While they have qualitatively dissociated the emotional and classical Stroop effects, all of the experiments, which have done so used a blocked paradigm, which confounds the fast and slow aspects of the emotional Stroop. Thus we propose that the qualitative nature of the differences they have found hinge upon their inclusion of slow effects, which we agree are fundamentally different from conventional Stroop interference. Our theoretical position on the role of the ACC and the model, which embodies it, support their stance for the case of slow (inter-trial) effects but not for fast effects. It should be stated that we are not intent on generalizing this phenomenon to all stressful situations, rather specifically to those in which the stressor is not a source of severe anxiety and is task irrelevant.

We propose that this salience sensitive control is the cortical counterpart of the amygdalar and thalamic fear circuitry [19]. Complementing the pre-attentive automatic vigilance towards threatening stimuli and the generalized arousal reaction following such stimuli is a cortical mechanism mediated by the affective portion of the ACC that temporarily suppresses ongoing cognitive activity to further augment the ability to detect and process threatening stimuli.

References

[1] J. D. Cohen, K. Dunbar and J. L. McClelland (1990) On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychol Rev,97(3),332-61*

[2] M. M. Botvinick, T. S. Braver, D. M. Barch, C. S. Carter and J. D. Cohen (2001) Conflict monitoring and cognitive control. *Psychol Rev,108(3),624-52*

[3] J. R. Stroop (1935) Studies of interference in a letter naming task. *Bulletin of the*

Psychonomic Society,10,31-34
[4] G. Bush, P. J. Whalen, B. R. Rosen, M. A. Jenike, S. C. McInerney and S. L. Rauch (1998) The counting Stroop: an interference task specialized for functional neuroimaging--validation study with functional MRI. *Hum Brain Mapp,6(4),270-82*
[5] W. C. Drevets and M. E. Raichle (1998) Reciprocoal Suppression of Regional Cerebral Blood Flow during Emotional versus Higher Cognitive Processes: Implications for Interactions between Emotion and Cognition. *Cognition and Emotion,12(3),353-385*
[6] G. Bush, P. Luu and M. I. Posner (2000) Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn Sci,4(6),215-222*
[7] F. N. Watts, F. P. McKenna, R. Sharrock and L. Trezise (1986) Colour naming of phobia-related words. *Br J Psychol,77 (Pt 1)97-108.*
[8] K. Mogg and B. P. Bradley (1998) A cognitive-motivational analysis of anxiety. *Behav Res Ther,36(9),809-48*
[9] A. Ohman, A. Flykt and F. Esteves (2001) Emotion drives attention: detecting the snake in the grass. *J Exp Psychol Gen,130(3),466-78*
[10] F. P. McKenna and D. Sharma (2004) Reversing the emotional Stroop effect reveals that it is not what it seems: the role of fast and slow components. *J Exp Psychol Learn Mem Cogn,30(2),382-92*
[11] C. Holle, J. H. Neely and R. G. Heimberg (1997) The effects of blocked versus random presentation and semantic relatedness of stimulus words on response to a modified Stroop task among social phobics. *Cognitive Therapy and Research,21 681-697*
[12] G. Mathews and T. A. Harley (1996) Connectionist Models of Emotional Distress and Attentional Bias. *Cognition and Emotion,10(6),561-600*
[13] C. S. Carter, M. Mintun and J. D. Cohen (1995) Interference and facilitation effects during selective attention: an H215O PET study of Stroop task performance. *Neuroimage,2(4),264-72*
[14] S. E. Petersen, P. T. Fox, A. Z. Snyder and M. E. Raichle (1990) Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. *Science,249(4972),1041-4*
[15] C. S. Carter, A. M. Macdonald, M. Botvinick, L. L. Ross, V. A. Stenger, D. Noll and J. D. Cohen (2000) Parsing executive processes: strategic vs. evaluative functions of the anterior cingulate cortex. *Proc Natl Acad Sci U S A,97(4),1944-8*
[16] K. Dunbar and C. M. MacLeod (1984) A horse race of a different color: Stroop interference patterns with transformed words. *J Exp Psychol Hum Percept Perform,10(5),622-39*
[17] J. Tzelgov, A. Henik and J. Berger (1992) Controlling Stroop effects by manipulating expectations for color words. *Mem Cognit,20(6),727-35*
[18] D. Algom, E. Chajut and S. Lev (2004) A rational look at the emotional stroop phenomenon: a generic slowdown, not a stroop effect. *J Exp Psychol Gen,133(3),323-38*
[19] J. LeDoux (2003) The emotional brain, fear, and the amygdala. *Cell Mol Neurobiol,23(4-5),727-38*