Modelling the Stroop Effect: Dynamics in Inhibition of Automatic Stimuli Processing

Nooraini Yusoff[1], André Grüning[1] and Antony Browne[1]

[1]Department of Computing, Faculty of Engineering and Physical Sciences,
University of Surrey, Guildford, GU2 7XH Surrey, UK
{n.yusoff, a.gruning, a.browne}@surrey.ac.uk

Abstract. In this study, we simulate the dynamics of suppressing an automatic stimulus processing which interferes with a different non-automatic target task. The dynamics can be observed in terms of interference and facilitation effects that influence target response processing time. For this purpose, we use Hopfield neural network with varying attention modulation in a colour-word Stroop stimuli processing paradigm. With the biologically realistic features of the network, our model is able to model the Stroop effect in comparison to the human performance.

Keywords: Neurodynamics, Automatic processing, Stroop effect, Hopfield neural network

1 Introduction

Distraction in processing of a target response can occur not only due to properties of irrelevant stimulus, but also through the relationship of the stimulus properties to the target-finding properties that attract the attention and resulting conflict [1]. Some environments can create automatic responses to certain classes of stimuli resulting in distraction from the intended task. This can also lead to priming effects as results of automatic processing of a stimulus (prime) that would distract or facilitate the processing of the later stimulus. The prime stimulus can be a cue if it is congruent with the later stimulus and facilitates its response; otherwise it is a distractor if both are conflicting, resulting in interference in processing. Having such dynamic effect will lead to variability in response processing time. Whenever interference occurs, the unintended automatic response has to be suppressed, using higher attentional control resources in producing the target response to the intended stimulus. Hence, in this study, we attempt to simulate the dynamics of the priming effect through modulation of attention in a colour-word Stroop paradigm [2].

The Stroop paradigm demonstrates human cognitive processing when one process dominates and inhibits the response of another process. Such environment involves automatic processing of a stimulus, distracting or sometimes facilitating the response to the intended stimulus. For example, in a colour-word Stroop test, the stimuli are the coloured colour-words in three conditions; control (e.g. a word RED written in black or a non colour-word – e.g. CAT written in red), conflicting (e.g. a word RED written in green) and congruent (e.g. a word RED written in red). Subjects are asked either to read the colour-words or to name their colours while the reaction time of performing the task is observed.

The results from Stroop studies show increased reaction time in naming the colour of the printed colour-word in the conflicting condition, while the subject could easily read the word without any significant distraction by the colour it is written in (e.g. [2], [4] and [5]). The conjectured explanation is that the asymmetric processing is due to automaticity of the word stimulus that is always processed first prior to the intended task stimulus, the colour. This consequently leads to the priming effects, whereby to name the colour of the word, one needs to inhibit the prime processed stimulus (the colour-word). The conflicting stimulus will increase the response processing time in colour naming; on the other hand the congruence of both will facilitate the later stimulus reducing the response processing time. For simulating the interference of the stimuli, we use a Hopfield neural network (HNN) with asynchronous pattern update procedures [3]. The Hopfield network is chosen for several reasons; we address the Stroop phenomenon as an association problem, the competition and cooperation of Stroop stimuli meets the pattern processing nature of the Hopfield network and the recall algorithm in Hopfield is biologically realistic.

2 Methods

To simulate the dynamics in inhibition of automated responses, we created a HNN with a memory of four ($M=4$) random 56-bit ($n=56$) patterns. Each pattern represents an association of the colour concept consisting of two congruent components, the colour-word (<WORD>, 16-bit) and the visual colour (<colour>, 16-bit) (based on findings by Folk et al. [1]), and the attentional resource (<Attention>, 24-bit). "<Attention>" models how much a subject attends to a task. If the subject is less attentive less on-bits are inherited from the memory pattern to the test pattern. With such association, we postulate that, each colour concept has some degree of attention resource, cooperating among them simultaneously in recalling a target pattern.

There are two main phases involved in our HNN based Stroop simulation; training phase and testing phase (based on mathematical computation as in [3] and [6]). During the training phase, all memory patterns (a pattern noted as x) are correlated to each other using the equation in 1. The correlation derives a set of weights (W_{56x56}) as a product of pattern vector associations.

$$w_{ij} = \sum_{k=1}^{M} x_i(k)x_j(k),\ i \neq j,\ w_{ii} = 0,\ i,j = \overline{1,n}\ . \qquad \textbf{(1)}$$

Once a set of association weights is obtained, in the testing phase, a test pattern, *xr*, is presented to the system. In our model, *xr*s represent the Stroop stimuli with an amount of attention depending on the task to recall the target colour concept. For Stroop stimuli representations, from each memory set, 20 test patterns are generated to observe the recall performance of a Stroop task (word reading or colour naming). The performances are observed under three conditions of stimuli; control – absence of irrelevant stimulus to the attended task (e.g. for a word reading; <Attention><RED><minimal noise>, 4 test patterns), conflicting – incongruent colour concept (e.g. <Attention><RED><green>, 12 test patterns) and congruent – compatible colour concept (e.g. <Attention><RED><red>, 4 test patterns).

For experimental setup, the initial activation ("on" bits) of a pattern depends on the task; word reading (WR) or colour naming (CN) is assumed to be signalled by another external system. For activation of bits in any components of a pattern, it refers to the percentage of similarity of "on" bits in the correspond memory. For instance, let a <WORD> component in a test pattern inherits *n%* random activations from its memory with 12 bits "on" (out of 16 bits), then there would be *n%x12* "on" bits in the component. For a WR task, the <WORD> component (relevant stimulus) inherits 75% random activation of its memory, while having random activation of the <colour> component (irrelevant stimulus) (ranging from 0% to 25% on-bits, with random uniform distribution). In contrast, for a CN task, the <colour> stimulus inherits 25% random activation from its memory, with random activation of <WORD> component (irrelevant stimulus) (ranging from 0% to 75% on-bits, with random uniform distribution). Greater maximum initial activation of word stimulus in CN is to simulate its automaticity that we predict would cause interference in processing the response to the task. For this reason, the colour stimulus can only survive with higher attention i.e. a test <Attention> that is more similar to the correspond <Attention> of a memory pattern. The initial activation applies to all conditions except the control stimuli with maximum of 25% noise in irrelevant stimulus.

In contrast with some other Stroop models (e.g. [4] and [5]) emphasizing the influence of neuron connection pathways (weights) in response processing, ours considers the influence of attention. In our model, the modulation of attention in a recall process is simulated through a part-pattern (<Attention> bits) completion. For our simulations, we varied the initial activation of attention from 0% (all off-bits) to 100% (identical attention component from correspond memory pattern). The dynamics in HNN asynchronous update would eventually complete the initial random activation of <Attention> vector corresponding to a target memory.

To recall a target memory pattern, at any given time, each bit (representing a neuron in biological system) in *xr*, receives the total activation (*net*, computed using 2) from others through the HNN asynchronous update mechanism:

$$net_i(t) = \sum_{i \neq j} w_{ji} \, (xr_j(t)) \, , \qquad (2)$$

where $net_i(t)$ is the *net* input to neuron i at time t, w_{ji} is the connection strength between neuron j to neuron i, xr_j is the state of neuron j (+1 or -1). In an update cycle, the state of neuron i is readjusted according to 3.

$$xr_i(t+1) = \Theta\ (net_i(t)) = \{+1,\ net_i(t) > 0;\ -1,\ net_i(t) < 0;\ xr_i(t),\ net_i(t)=0\}\ . \qquad (3)$$

The number of bit updates in a cycle with maximum of 300 iterations for a stimulus recalling the closest (measured by Euclidean Distance) target memory is recorded as the reaction time. This simulates the reaction time taken by a subject to perform any of the tasks. Graphs and numbers below are the averages for repetitions of the experiment with 10 different choice of memory sets, and 11 different levels of attention. Asynchronous updates of bits in <Attention>, <WORD> and <colour> vectors simultaneously, simulate the dynamics of cognitive process in Stroop phenomenon for active inhibition, facilitation and interference.

3 Results

As shown in Fig. 1, using the Stroop stimuli to recall their target memories through active inhibition of Hopfield's algorithm, our model predicts the asymmetric stimuli processing in the colour-word Stroop phenomena in comparison to the human performance.

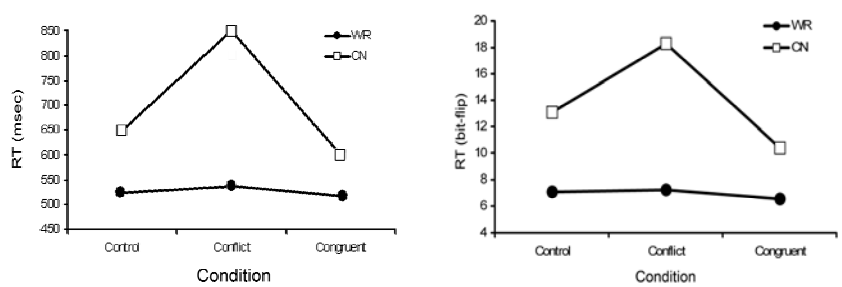

Fig. 1. Performance results for Stroop task averaged over all levels of attention; word reading (*WR*) and colour-naming (*CN*). (Left: Results from empirical study after Dunbar & MacLeod [7], Right: Results of the model's simulation)

For our model, we consider the average of the reaction time (RT) in processing response for both WR and CN at all levels of initial attention activation ranging from 0% to 100%. After running 110 different simulations (10 memory sets for 11 levels of attention), the model shows that the words are always read faster than colours are named with no significant difference of RT (*ANOVA:* $p > 0.05$) on the stimuli conditions ($WR_{(RT,control)}$=7.05, $WR_{(RT,conflict)}$=7.26, $WR_{(RT,congruent)}$=6.53). Meanwhile there is a significant difference (*ANOVA:* $p < 0.01$) found in CN in all conditions ($WCN_{(RT,control)}$=13.05, $CN_{(RT,conflict)}$=18.32, $CN_{(RT,congruent)}$=10.39), except between the control and the congruent stimuli (*ANOVA:* $p_{(control,congruent)} > 0.05$). The interference is obviously observed for CN with increasing time in the conflicting condition whilst the slight benefit of automatic word processing is shown with the decreasing reaction time for congruent condition. In addition to the RT, the frequency (freq) of correct recalls to target responses was also observed. As what we predicted, the same simulations provide consistent results in frequencies of correct recalls with RT, indicating longer processing time leads to higher recall error rate. The correct recalls recorded in WR are as follows; $WR_{freq(control,correct)}$=87.82, $WR_{freq(conflict,correct)}$=87.43, $WR_{freq(congruent,correct)}$=91.87, whilst for CN we obtained; $CN_{freq(control,correct)}$=70.64, $CN_{freq(conflict,correct)}$ =23.25, $CN_{freq(congruent,correct)}$=81.02. However in this study, without also neglecting the recall frequencies results, we only focus on the RT as the determinant of the Stroop effect. The results conclude that high inhibition has occurred in CN especially in the conflicting condition due to the incompatibility of the prime stimulus (the word) with the intended stimulus (the colour) having caused the interference, where this has been demonstrated at any level of pre-selective attention in our simulations. On the other hand, a prime stimulus compatible with the target response of the intended task speeds up the processing time.

4 Conclusions

In our approach, we have used a HNN to demonstrate the dynamics of inhibition by an automatic response over an intended stimulus response. The dynamics can be seen as the influence of priming effects resulting from automatic processing. The results showed that our model is able to model the reaction times in the colour-word Stroop paradigm.

References

1. Folk C. L., Remington R. W., Johnston J. C.: Involuntary Covert Orienting Is Contingent On Attentional Control Settings. Journal of Experimental Psychology: Human Perception & Performance, Vol. 18. (1992) 1030-1044
2. Stroop, J. R.: Studies of Interference in Serial Verbal Reactions. Journal of Experimental Psychology, Vol. 18, (1935) 643-662
3. Hopfield, J.: Neural Networks and Physical Systems with Emergent Collective Computational Abilities. In: National Academy of Sciences of the USA, 9 (2554) (1982)
4. Cohen, J.D., Dunbar, K., McClelland, J.L.: On the Control of Automatic Processes: A Parallel Distributed Account of the Stroop Effect. Psychological Review, 97(3). (1990) 332--361
5. Phaf, R. H., van der Heijden, A. H. C., Hudson, P. T.W.: SLAM: A Connectionist Model for Attention in Visual Selection Tasks. Cognitive Psychology, Vol. 22. (1990) 273-341
6. Popoviciu, N., Boncut, M.: On the Hopfield algorithm. Foundations and examples. General Mathematics, Vol. 13, No. 2. (2005) 35-50.
7. Dunbar, K., MacLeod, C. M.: A Horse Race of a Different Color: Stroop Interference Patterns with Transformed Words. Journal of Experimental Psychology: Human Perception and Performance, Vol. 10. (1984) 622-639