

# How the Stroop Effect Arises from Optimal Response Times in Laterally Connected Self-Organizing Maps

Divya Prabhakaran (divyapr@cs.utexas.edu)<sup>1</sup>, Uli Grasemann (uli.grasemann@gmail.com)<sup>1</sup>, Swathi Kiran (kirans@bu.edu)<sup>2</sup>, Risto Miikkulainen (risto@cs.utexas.edu)<sup>1</sup>

<sup>1</sup>The University of Texas at Austin, Austin, TX 78712 USA

<sup>2</sup>Boston University, Boston, MA 02215 USA

## Abstract

The Stroop effect refers to cognitive interference in a color-naming task: When the color and the word do not match, the response is slower and more likely to be incorrect. The Stroop task is used to assess cognitive flexibility, selective attention, and executive function. This paper implements the Stroop task with self-organizing maps (SOMs): Target color and the competing word are inputs for the semantic and lexical maps, associative connections bring color information to the lexical map, and lateral connections combine their effects over time. The model achieved an overall accuracy of 84.2%, with significantly fewer errors and faster responses in congruent compared to no-input and incongruent conditions. The model's effect is a side effect of optimizing speed-accuracy tradeoffs, and can thus be seen as a cost associated with overall efficient performance. The model can further serve studying neurologically-inspired cognitive control and related phenomena.

**Keywords:** Neural networks; Stroop task; Self-organizing maps; Computational modeling; Cognitive control

## Introduction

Traditionally used in neuropsychological assessments, the Stroop Color and Word Test (SCWT) involves subjects reading words written in colored ink and naming colors (color-word pairs). The SCWT uses congruent and incongruent conditions of color-word naming, where incongruent conditions are represented by mismatched color-words, such as the word "red" in green ink (Stroop, 1935). The Stroop effect refers to cognitive interference in the disruption of processing speed and naming when a stimulus (semantic word) interferes with the concurrent processing of a second stimulus (color), resulting in slower response times and an increased number of naming errors with incongruent stimuli (Scarpina & Tagini, 2017). Stroop task conditions are widely used to assess cognitive control as a measure of processing speed, selective attention, cognitive flexibility, and executive function (Jensen & Rohwer, 1966).

Behavioral studies of the Stroop effect, focused on summary measures such as reaction times and error rates, are not well suited for identifying the underlying mechanisms of interference. Similarly, studies of brain region activations alone cannot reveal precise information about how Stroop interference occurs: control and interference can occur at multiple time points from task initiation to later stages of response production. Computational models can lead to such insights. For instance, existing models (Cohen et al., 1990; Herd et al., 2006; Kaplan et al., 2007; Laeng et al., 2005; Nishikawa et al., 2019; Stafford & Gurney, 2005; Verguts, 2017) produce the Stroop effect based on top-down control mechanisms where processing pathways are modulated based

on task demands. However, recent critique suggests that attention driven by the salience and relevance of stimuli could cause Stroop interference based on bottom-up, input-driven processes (Algom & Chajut, 2019).

The study described in this paper addresses this gap between top-down and bottom-up mechanisms. It demonstrates how the Stroop effect arises mechanistically and naturally from the model's underlying structure of cortically-inspired, self-organized lexical and perceptual representations. Importantly, the model explains not just how the Stroop interference occurs, but also why it represents an optimization of efficient language and perceptual processing. It can thus lead to general insights into executive function and cognitive control by highlighting how attentional constraints and stimulus-driven processing supplement or constrain these higher-order cognitive processes.

## Background

The neural circuitry underlying the Stroop effect is relatively well understood, leading to several efforts to model it computationally.

## Cognitive Neuroscience Models

The Stroop effect is believed to arise from processing in the cortico-striatal network. This network is involved in efficient language processing through the regulation and allocation of cognitive resources such as verbal attention and working memory (Jacquemot & Bachoud-Lévi, 2021). The cortex-basal ganglia-thalamus loop is a crucial part of information routing and selection for cognitive control. The basal ganglia (BG) acts as a cognitive selection mechanism by optimizing information gating, as it receives input from the striatum containing processed signals from the cortex, and projects output through the thalamus back to the cortex (Stafford & Gurney, 2005). The left caudate nucleus of the BG plays an important role in suppressing competing alternatives in the indirect pathway and facilitating selected cortical representations, as well as language selection and control (Abutalebi & Green, 2007; Friederici et al., 2003).

Neuroimaging studies that focused on goal-directed behavior in the prefrontal cortex (PFC) component of the cortico-striatal network have led to a "cascade-of-control" model examining multiple levels of Stroop interference and cognitive control (Banich, 2019). The posterior regions of the lateral prefrontal cortex regions bias processing towards task-relevant and away from task-irrelevant information, the cau-

dal mid-cingulate regions are involved in response selection, and the anterior cingulate cortex (ACC) sends feedback to the lateral prefrontal cortex based on response accuracy.

### Computational Models

The Stroop effect has been replicated computationally through various neural network (NN) models of automaticity and control processes. In one such approach, parallel distributed processing, word reading and color naming pathways run simultaneously with different connection strengths — word reading has stronger connections due to habitual reading while color naming is weaker (Cohen et al., 1990). Task demand units modulate attention in these pathways by enhancing processing in one pathway while reducing activation of the other. Normally, the stronger word reading connections result in correct naming with slower reaction times in the incongruent condition, but increased attention in the color naming units can be used to induce errors.

This model was used to study sensory input factors such as color types. Using trichromatic processing as an input layer, opponent color pairs (blue-yellow as compared to non-opponent color pairs such as blue-red) resulted in reduced interference in incongruent naming conditions (Laeng et al., 2005). The Cohen model was extended with top-down conflict monitoring, where the unit’s activation increases depending on the frequency of past incongruent trials (Botvinick et al., 2001). Following an error, the responses were slower and more accurate. The simulations revealed interference both at the perceptual encoding stage of color processing in the Stroop task, as well as from conflict monitoring at the response level.

This paper builds on existing computational models by implementing self-organizing maps (SOMs) to study the emergence of the Stroop interference from the model structure’s competitive learning mechanisms itself, without explicit top-down control.

### Approach

The Stroop effect is observed in a color naming task, where concurrent linguistic input either competes with or coincides with the visual input. Color naming is a function of the human lexicon, the part of the brain concerned with representing and translating between word forms and their meanings. Thus, to gain insight into the Stroop effect, this paper builds on BiLex, a neural network model of the human lexicon (Grasemann et al., 2019; Grasemann, Peñaloza, Dekhtyar, Miikkulainen, & Kiran, 2021; Kiran et al., 2013; Peñaloza et al., 2019). This section reviews BiLex, extends it with lateral connections to model response times, describes how the model was trained, and how the Stroop effect was measured on it.

### The BiLex Model

BiLex consists of self-organizing maps (SOMs; Kohonen (1982); Kohonen et al. (2001)) that learn to organize word forms and their meanings according to semantic and lexical

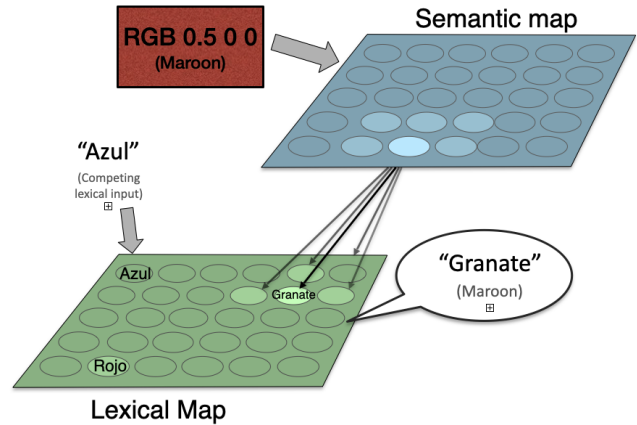


Figure 1: The BiLex model of the Stroop effect. The semantic color input and concurrent lexical input are combined in the lexical map, and a lexical output is generated naming the color (maroon in this case). When the inputs are congruent, the response is fast and reliable; when they are incongruent, slower and error-prone. The errors can be seen as a cost of optimization of overall response times.

similarity. Maps are connected by learned associative connections that are used to translate between alternative semantic or lexical representations of a given word. As an example of a lexical task, picture naming is simulated in BiLex by presenting a semantic word representation to the semantic map and propagating the resulting map activation to a lexical map, which then uses its activation to generate lexical output.

The original BiLex was designed to model Spanish/English bilingual subjects, and therefore had one semantic map and two lexical maps, both modeling spoken words (i.e. phonetic maps). For the Stroop effect, only one lexical map is needed; Spanish was used for the current study. For simplicity, it was used to represent both the concurrent linguistic input and the task output. Figure 1 illustrates the architecture of the model.

### Extension with Lateral Connections

The original BiLex was based on traditional SOMs, where map activation and learning are governed by a predefined neighborhood function, centered around the maximally responding map unit (see Table 1 for the parameters of this model). This paper extends the SOMs with lateral connections, i.e., the map activation settles over several time steps through lateral interactions between map neurons. This settling process is an important extension of the original BiLex: it allows modeling response times, which is crucial for the Stroop effect. Inhibitory lateral connections are uniform, i.e. connection strengths between each pair of units are the same. Excitatory connections focus on a small Gaussian map neighborhood with standard deviation  $\sigma$ . Both the outgoing excitatory and inhibitory connection weights for each map unit sum to one.

When an input  $c \in C$  (a feature vector representing its semantic meaning or its lexical word form) is presented to a

Table 1: SOM Parameter definitions.

Parameter	Definition
$c$	Semantic/lexical feature vector
$C$	Input set
$i$	Map neuron
$\alpha = 0.08$	Learning rate
$\beta = 0.1$	Learning rate (lateral connections)
$r_{\text{lex}} = 0.45$	Lexical input interference strength
$r_{\text{sem}} = 0.05$	Semantic to lexical map influence strength

map, the initial activation for each map neuron  $i$  is based on normalized and scaled similarity calculated as

$$a_{0,i} = \text{softmax}\left(\frac{|u_i - c|^2}{C_m T}\right), \quad (1)$$

where  $u_i$  is the map unit  $i$ 's feature vector,  $c$  is the map input,  $C_m$  is the median Euclidean distance between pairs of vectors in the entire input set  $C$ , and  $T$  is a temperature parameter that controls initial activation focus.

The initial activation is then updated for all units  $i$  over several time steps  $t$ :

$$a_{t+1,i} = \alpha a_{t,i} + \beta \sum_j a_{t,j} (w_{ij}^+ - w_{ij}^-), \quad (2)$$

where  $\alpha$  controls how fast activation fades,  $\beta$  controls the strength of lateral interactions, and  $w_{ij}^+$  and  $w_{ij}^-$  are excitatory and inhibitory connections strengths between units  $i$  and  $j$ . Over time, long-range inhibition and short-range excitation focus the map's activation on a group of neighboring similar map neurons.

At each time  $t$ , the likelihood of a specific map input  $c$  is calculated as

$$L(c,t) = \sum_{i \in M_c} a_{i,t}, \quad (3)$$

where  $M_c$  is the set of map units  $i$  that are closer to  $c$  than to any other input in  $C$  (shown as background colors in Figure 2). Calculating  $L(c,t)$  for all inputs in  $C$  gives a probability distribution whose (negative) entropy  $E_t$  reflects the map's uncertainty at time  $t$  about which input  $c$  is observed. In the reported experiments,  $E_t < 1$  was used as a threshold to stop lateral interactions and produce the maximum-likelihood response  $c$  as the map's output.

## Model Training

A set of 16 words was used for model training, each representing a standard color. RGB values of each color were used as semantic representations (e.g. red was (1, 0, 0)), and lexical representations were generated as in previous BiLex models (Peñalosa et al., 2019), starting with IPA transcriptions of each color name, and transforming them into vectors of phonetic features (e.g. openness for vowels, place and manner for consonants).

Maps were trained similarly to standard SOMs, i.e. by repeatedly presenting the map with a randomly chosen word

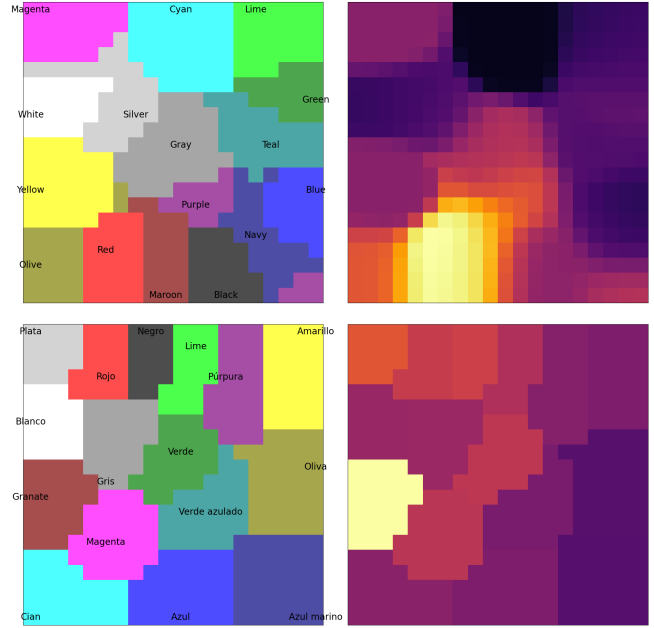


Figure 2: The organization and activation in the BiLex model of the Stroop Effect. The model implements SOMs with lateral connections, with map units representing 16 colors and words. The semantic map (*top*) was trained with RGB values of each color, and the lexical map (*bottom*) with Spanish phonetic representations of words for each color. Initial map activation for the color maroon/granate is depicted in the right column. Similar colors are nearby in the semantic map, and similar words in the lexical map. As a result, the map activations are smooth and continuous.

$c$ , and adjusting a map neighborhood to become more similar to the input. However, instead of using a neighborhood surrounding the map unit most similar to the input, the map was presented with the input pattern (Eqn. 1), and then updated for 30 time steps (Eqn. 2). Each unit's feature vector  $u_i$  was then adjusted to become more similar to the input  $c$ , with more highly activated units changing more:

$$u'_i = (1 - \eta a_i) u_i + \eta a_i c, \quad (4)$$

where  $\eta$  is the map's learning rate. Over time, as with regular SOMs, this process results in neighboring units becoming more similar to each other, and the map as a whole representing the input vectors more and more closely (as shown on the left side of Figure 2).

Both maps in the experiments consisted of 20x20 units. They were trained for 1000 epochs, with the learning rate decreasing exponentially from 0.2 initially to 0.05 at the end of training. The standard deviation  $\sigma$  of excitatory lateral connection weights decreased from 2.0 to 0.25 at the same time.

Associative connections between the semantic and lexical map were trained with Hebbian learning. Both maps were repeatedly presented with the same randomly chosen input color, and the resulting map activations were updated until

the map’s entropy fell below one bit. Connections between semantic unit  $i$  and lexical unit  $j$  were then increased if they responded to the same input:

$$w'_{ij} = w_{ij} + \theta a_i a_j, \quad (5)$$

where  $\theta$  is the learning rate for associative connections. In this way, co-activated neurons were connected more strongly over time. Associative connections were trained for 10 epochs; the learning rate  $\theta$  was set to 0.1. After each epoch, associative connections were normed such that each neuron’s output weights summed to one.

### Simulating the Stroop Task

The Stroop task can be simulated in the model as follows:

1. The semantic representation of a color (its RGB values) is presented to the semantic map (Eqn. 1), resulting in an initial map activation like the one for "Maroon" shown in the top right panel of Figure 2.
2. A congruent or incongruent lexical input is presented to the lexical map, as shown in the bottom right panel of Figure 2. In the 'no input' condition, no lexical input is presented to the lexical map. The strength of the interference (referred to below as  $r_{\text{lex}}$ ) can be varied by multiplying the initial activation by a number between 0 and 1, where 0 reflects no interference.
3. Map activations for both maps are updated using lateral interactions between map units (Eqn. 2), resulting in incrementally more focused activations.
4. Activation is transferred from the semantic to the lexical map through associative connections:  $a_{t,j} = a_{t,j} + \gamma \sum_i a_{t,i} w_{i,j}$ , where  $w_{i,j}$  is the associative connection weight between semantic unit  $i$  and lexical unit  $j$ , and  $\gamma$  determines the strength of between-map interaction.
5. If the entropy  $E_t$  of the lexical map is below a given threshold, the maximum-likelihood lexical output  $c$  (according to Eqn. 3) is produced as the color named by the model. If the maximum number of steps is exceeded, naming fails. Otherwise, return to Step 2.

The Stroop task simulation is counted as successful if the output color produced corresponds to the one presented to the semantic map, independent of the competing lexical input. The number of time steps  $t$  is recorded as the time required to complete the Stroop task. A complete factorial design with 16 colors was used to evaluate the congruent (16 input cases), no input (16 input cases), and incongruent conditions (240 input cases).

An important part of setting up the model is to determine the right level of interference, i.e. the relative strength of routing the semantic and congruent lexical input. Optimal routing parameters were found by running the Stroop task simulation with a baseline  $r_{\text{sem}}$  value of 0.05 and varying the  $r_{\text{lex}}$  value

from 0.0 to 1.0 in 0.05 increments. Error rates and response times across all three conditions were recorded. The quality  $q$  of each  $r_{\text{lex}}$  value was measured as

$$q = \frac{e_i - \min(e_i)}{\max(e_i) - \min(e_i)} + \frac{t_i - \min(t_i)}{\max(t_i) - \min(t_i)} + \frac{t_c - \min(t_c)}{\max(t_c) - \min(t_c)} \quad (6)$$

where  $e_i$  = mean incongruent error rate,  $t_i$  = mean incongruent response time,  $t_c$  = mean congruent response time (there were no errors in the congruent condition and therefore  $e_c$  was not needed). The min and max indicate the smallest and largest observations of these values; min-max normalization converts  $e_i$ ,  $t_i$ , and  $t_c$  to values between 0 and 1. Note that a low  $q$  score can be achieved by a very good performance in one metric, even when the performance is poor in the others. The  $q$  score thus balances the need to respond fast and to reduce errors at the same time. The  $r_{\text{sem}}$  and  $r_{\text{lex}}$  values with the best  $q$  value were chosen for the rest of the experiments.

In the experiments, the threshold for  $E_t$  was always one bit. The strength  $\gamma$  of between-map interactions was 0.05. The temperature  $T$  in Eqn. 1 was 0.7 for the semantic map and 0.5 for the lexical map.  $\alpha$  and  $\beta$  in Eqn. 2 were 0.8 and 0.1. The maximum number of time steps was set to 500, which was never reached in the reported experiments. All parameters were set empirically. Linear mixed model analysis with the semantic target color as a repeated measure was conducted to determine whether the response time differences were statistically significant between the Stroop conditions.

## Results

Lower  $r_{\text{lex}}$  values resulted in slower response times and reduced incongruent errors, while higher  $r_{\text{lex}}$  values allowed completing the color naming task faster, but with more incongruent errors. Normalized values for  $e_i$ ,  $t_i$ , and  $t_c$  were compared. Lower  $r_{\text{lex}}$  values, like between 0.30 and 0.40, have lower  $e_i$  normalized values, but higher  $t_i$ . Higher  $r_{\text{lex}}$  values, e.g. 0.50 to 0.65, have higher  $e_i$  and  $t_i$  normalized values, but low  $t_c$ . Further, the max-min spread for 0.45 is smaller compared to other  $r_{\text{lex}}$  values. The best routing parameters were found to be  $r_{\text{lex}}=0.45$  and  $r_{\text{sem}}=0.05$ , as max-min spread is smaller compared to other  $r_{\text{lex}}$  values, indicating balanced performance (Figure 3).

The simulations demonstrated a strong Stroop effect. Overall, across all three conditions, the model named the colors correctly 84.2% of the time. When presented with congruent inputs, the model was 100% accurate, with incongruent inputs, 83.0% accurate, and with no-input cases, 87.5% accurate.

Similar results were observed in response times as well. The linear mixed model analysis used 229 successful congruent, incongruent, and no-input cases across the 16 target color groups. Group sizes ranged from 2 to 17 cases, with a mean group size of 14.3. Compared to congruent input, the incongruent and no input conditions significantly delayed the output convergence (Figure 4;  $p < 0.001$ ). The incongruent condition had the strongest effect, with accurate color nam-

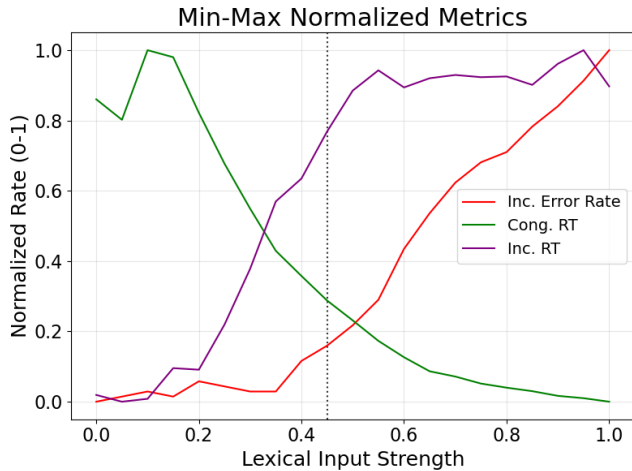


Figure 3: Finding the best values for the  $r_{lex}$  routing parameter. The normalized  $e_i$  (red),  $t_i$  (purple), and  $t_c$  (green) components of the  $q$  metric are depicted for each possible  $r_{lex}$  value. The optimal routing parameter was found to be an  $r_{lex}$  of 0.45, marked by the dashed black line, balancing color naming speed with minimized naming errors.

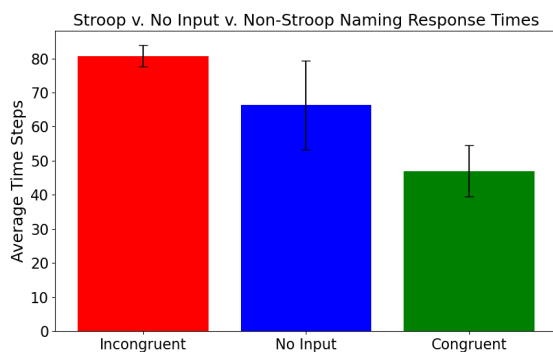


Figure 4: The mean and standard error (SE) of RTs across the three conditions of the Stroop effect: congruent (red), no-input (blue), and incongruent (green). RTs improve significantly for congruent vs. no-input vs. incongruent conditions, indicating that the model exhibits a strong Stroop effect.

ing taking approximately 55 steps longer than the congruent condition ( $\beta = 55.05$ ,  $SE = 6.08$ ,  $95\% CI = [43.1, 67.0]$ ). The no-input cases also converged more slowly, on average taking 36 steps longer than the congruent condition ( $\beta = 35.87$ ,  $SE = 8.58$ ,  $95\% CI = [19.1, 52.7]$ ). Group variance (3485.2) indicated that there was high variation between different semantic inputs. However, the Stroop effect within the combined model persisted across all input patterns for the same task, with statistically significantly slower response times for incongruent vs. no input vs. congruent conditions.

Figure 5 compares model response times to those of human adults in two prior studies (Forte et al., 2024; Wright, 2017), where mean reaction times (RTs, in ms) were measured with a computerized Stroop test. Results indicate that the model

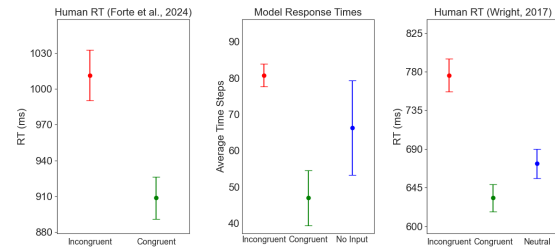


Figure 5: The mean and SE of RTs in the model compared with human performance. Left:  $n = 98$ , ages: 37 - 55 years (Forte et al., 2024); right:  $n = 77$ , ages: 17 - 45 years (Wright, 2017). The vertical scales were adjusted to line up the averages; the SEs are mostly similar and the differences between the conditions are significant, suggesting that the model replicates the human data.

performs similarly to humans. The absolute RT values and variance are different, but in all cases there is a statistically significant differences between conditions.

Figure 6 illustrates the progression of the lexical and semantic map activations in an incongruent input case. Over time, activation is transferred from the semantic map to the lexical map, enabling the lexical map to converge on the correct color output in this case. Figure 7 illustrates the responses over time for incongruent and congruent naming, demonstrating that the rate of convergence on the correct output is faster in the congruent condition.

## Discussion and Future Work

The main result is that the laterally connected BiLex model can account for the Stroop effect. Incongruent inputs take longer to process and result in more errors than congruent inputs. The same is true to a lesser extent for the no-input condition. Importantly, this process arises from the interactions within and between the self-organizing maps, mediated by the lateral and associative connections. Attention mechanisms are implicitly represented through these interactions, demonstrating how performance is optimized.

The current experiments should be improved by modeling transfer in semantic conflict and response conflict effects (Chen et al., 2013), stimulus onset asynchrony (SOA), word reading vs. color naming performance, and practice effects. It may also be possible to model age-related sensitivity, i.e. how the Stroop effect becomes stronger as a result of a reduced capacity for task-inhibition in older age (Milham et al., 2002), or simulate performance in patients with impairments by systematically lesioning the semantic and lexical maps, as implemented with BiLex models of aphasic patients (Grasemann, Peñalozza, Dekhtyar, Hillis, et al., 2021).

The routing parameters that were found to maximize naming speed while minimizing error responses reveal an interesting effect. With competing lexical input, both naming errors and response time increase, while blocking competing input

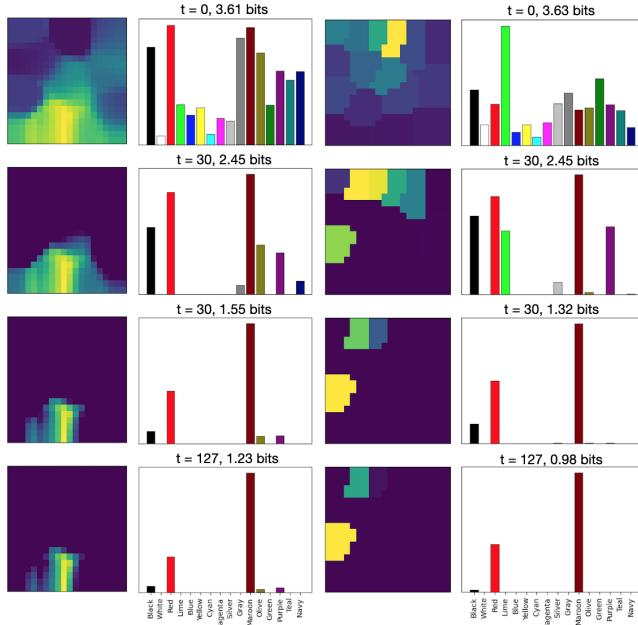


Figure 6: The response of the model over time. Lexical map activation (*right*) is shown over time for incongruent inputs. Initially (at  $t = 0$ ), map activation (*left*) only reflects the lexical input. However, as time progresses, the correct input (maroon) is transferred from the semantic to the lexical map and gains more weight ( $t = 30$ ), eventually causing the lexical map to converge on the correct output ( $t = 127$ ). It is this competition over time, mediated by the lateral connections, that results in the Stroop effect.

leads to more accurate and faster responses. Naming is faster on average when the lexical input and transfer strength are above zero; however, the cost of this improvement is that the Stroop effect emerges with incongruent cases. Data-driven selective attention towards the most salient inputs is a crucial part of human cognition (Algom & Chajut, 2019); the Stroop effect thus arises from the intention to use available inputs optimally. By optimizing lexical input strength ( $r_{\text{lex}} = 0.45$ ), the model can achieve full accuracy in the congruent condition and minimize error in the incongruent condition, while achieving an efficient run time. Given that inputs in the real world are more likely to be congruent, Stroop errors can be seen as a natural side effect of a system that performs well in most cases.

In the current model, routing was simply reduced to two parameters, but it is a complex mechanism in its own right. A more accurate modeling of routing in BiLex could lead to further insights into cognitive control. One promising approach is the conditional routing model (CRM; Stocco et al., 2010), in which the basal ganglia implement an adaptive information routing system that modulates signals between pairs of cortical regions. Similar to CRM, the routing rules in BiLex can be learned through contrastive Hebbian learning (CHL; Dayan & Abbott, 2001; Rolls & Treves, 1998), which uses

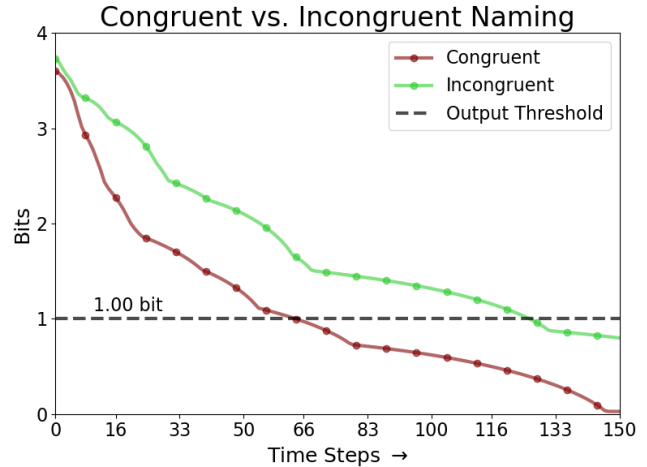


Figure 7: Comparing the model response to the congruent and incongruent inputs. When the lexical map is given incongruent input (*lime*), the map takes significantly longer to focus on the specified color and cross the output threshold for uncertainty below 1.0 bit than when the inputs are congruent (*maroon*).

only locally available information to adapt connections and represent routing rules within the striatal part of the network.

While the original CRM executes one interaction between a pair of cortical regions at a time, the new model would continually update (and possibly reevaluate) all routing parameters within the lexicon, creating an adaptive, task-dependent dynamic system rather than a binary routing mechanism. Unlike previous bilingual Stroop models (Yoon, 2011), the new model would demonstrate how bilingual lexical function in humans is optimized over time. It is therefore ideally suited to explain how bilingual Stroop interference arises from stimulus-driven attention combined with top-down control. In addition to the Stroop effect, BiLex with conditional routing can help understand language acquisition and lexical access in L1 and L2, priming effects, response timing, true and false cognates, and language and task switching effects. This expansion follows logically from the Stroop effect experiments in this paper and constitutes a most compelling direction of future work.

## Conclusion

The BiLex self-organizing map model of the lexical/semantic system was adapted to simulate the Stroop effect. When extended with lateral connections, the model could account not only for the increased errors in the incongruent condition but also for the differences in response time between congruent, no-input, and incongruent conditions. Interestingly, the Stroop effect in the model can be seen as a necessary cost for optimizing overall response times with accuracy. The model can thus serve as a foundation for studying cognitive control and the effects that arise from it.

## Acknowledgments

The authors wish to thank Andrea Stocco and Arturo Hernandez for valuable discussions. This research was supported in part by NIH under grant 5R01DC020653-02.

## References

- Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, *20*(3), 242–275. doi: 10.1016/j.jneuroling.2006.10.003
- Algom, D., & Chajut, E. (2019). Reclaiming the stroop effect back from control to input-driven attention and perception. *Frontiers in psychology*, *10*, 1683.
- Banich, M. T. (2019). The Stroop Effect occurs at multiple points along a cascade of control: Evidence from cognitive neuroscience approaches. *Frontiers in Psychology*, *10*. doi: 10.3389/fpsyg.2019.02164
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. doi: 10.1037/0033-295X.108.3.624
- Chen, Z., Lei, X., Ding, C., Li, H., & Chen, A. (2013). The neural mechanisms of semantic and response conflicts: An fmri study of practice-related effects in the stroop task. *NeuroImage*, *66*, 577–584. doi: https://doi.org/10.1016/j.neuroimage.2012.10.028
- Cohen, J., Dunbar, K., & McClelland, J. (1990, 07). Cohen jd, mclelland jl, dunbar k. on the control of automatic processes: a parallel distributed processing account of the stroop effect. *psychol rev* *97*: 332–361. *Psychological review*, *97*, 332–61. doi: 10.1037/0033-295X.97.3.332
- Dayan, P., & Abbott, L. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems* (Vol. 15). MIT Press.
- Forte, G., Troisi, G., Favieri, F., & Casagrande, M. (2024). Inhibition changes across the lifespan: experimental evidence from the stroop task. *BMC Psychology*, *12*, 336. doi: 10.1186/s40359-024-01844-0
- Friederici, A. D., Rüschemeyer, S.-A., Hahne, A., & Fiebach, C. J. (2003, feb). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cerebral Cortex*, *13*(2), 170–177. doi: 10.1093/cercor/13.2.170
- Grasemann, U., Miikkulainen, R., Peñaloza, C., Dekhtyar, & Kiran, S. (2019). Evolutionary optimization of neural-network models of human behavior. In *Proceedings of the international conference on cognitive modeling*. Montreal, Canada.
- Grasemann, U., Peñaloza, C., Dekhtyar, M., Hillis, A. E., Martin, R. C., Shafto, P., & Rapp, B. (2021). Predicting language treatment response in bilingual aphasia using neural network-based patient models. *Scientific Reports*, *11*(1), 10497. doi: 10.1038/s41598-021-89443-6
- Grasemann, U., Peñaloza, C., Dekhtyar, M., Miikkulainen, R., & Kiran, S. (2021). Predicting language treatment response in bilingual aphasia using neural network-based patient models. *Scientific Reports*, *11*, 10497.
- Herd, S. A., Banich, M. T., & O’reilly, R. C. (2006). Neural mechanisms of cognitive control: An integrative model of stroop task performance and fmri data. *Journal of cognitive neuroscience*, *18*(1), 22–32.
- Jacquemot, C., & Bachoud-Lévi, A.-C. (2021). Striatum and language processing: Where do we stand? *Cognition*, *213*, 104785. (Special Issue in Honour of Jacques Mehler, Cognition’s founding editor) doi: https://doi.org/10.1016/j.cognition.2021.104785
- Jensen, A. R., & Rohwer, W. D. (1966). The stroop color-word test: A review. *Acta Psychologica*, *25*, 36–93. doi: https://doi.org/10.1016/0001-6918(66)90004-7
- Kaplan, G. B., Şengör, N., Gürvit, H., & Güzeliş, C. (2007). Modelling the stroop effect: A connectionist approach. *Neurocomputing*, *70*(7), 1414–1423. (Advances in Computational Intelligence and Learning) doi: https://doi.org/10.1016/j.neucom.2006.05.009
- Kiran, S., Grasemann, U., Sandberg, C., & Miikkulainen, R. (2013). A computational account of bilingual aphasia rehabilitation. *Bilingualism: Language and Cognition*, *16*(2), 325–342.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*, 59–69. doi: 10.1007/BF00337288
- Kohonen, T., Schroeder, M. R., & Huang, T. S. (2001). *Self-organizing maps* (3rd ed.). Berlin, Heidelberg: Springer.
- Laeng, B., Låg, T., & Brennen, T. (2005, 06). Reduced stroop interference for opponent colors may be due to input factors: Evidence from individual differences and a neural network simulation. *Journal of experimental psychology. Human perception and performance*, *31*, 438–52. doi: 10.1037/0096-1523.31.3.438
- Milham, M. P., Erickson, K. I., Banich, M. T., Kramer, A. F., Webb, A., Wszalek, T., & Cohen, N. J. (2002). Attentional control in the aging brain: Insights from an fmri study of the stroop task. *Brain and Cognition*, *49*(3), 277–296. doi: https://doi.org/10.1006/brcg.2001.1501
- Nishikawa, T., Hashimoto, Y., Minami, K., Watanuki, K., Kaede, K., & Muramatsu, K. (2019). Examination of the brain areas related to cognitive performance during the stroop task using deep neural network. In S. Fukuda (Ed.), *Advances in affective and pleasurable design* (pp. 94–101). Cham: Springer International Publishing.
- Peñaloza, C., Grasemann, U., Dekhtyar, M., Miikkulainen, R., & Kiran, S. (2019). Bilex: A computational approach to the effects of age of acquisition and language exposure on bilingual lexical access. *Brain and Language*, *195*, 104643. doi: https://doi.org/10.1016/j.bandl.2019.104643
- Rolls, E. T., & Treves, A. (1998). Neural networks and brain function. In *Neural networks and brain function*.

- Scarpina, F., & Tagini, S. (2017). The stroop color and word test. *Frontiers in Psychology, 8*. doi: 10.3389/fpsyg.2017.00557
- Stafford, T., & Gurney, K. (2005). The basal ganglia as the selection mechanism in a cognitive task. *Modeling Natural Action Selection, 77–83*.
- Stocco, A., Lebiere, C., & Anderson, J. R. (2010). Conditional routing of information to the cortex: A model of the basal ganglia's role in cognitive coordination. *Psychological Review, 117*(2), 541–574. doi: 10.1037/a0019077
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. doi: 10.1037/h0054651
- Verguts, T. (2017). Binding by random bursts: A computational model of cognitive control. *Journal of cognitive neuroscience, 29*(6), 1103–1118.
- Wright, B. C. (2017). What stroop tasks can tell us about selective attention from childhood to adulthood. *British Journal of Psychology, 108*(3), 583–607. Retrieved from <https://doi.org/10.1111/bjop.12230> doi: 10.1111/bjop.12230
- Yoon, H. W. (2011). Computational neural model of the bilingual stroop effect: An fmri study. In *Proceedings of the international conference on ubiquitous computing and multimedia applications (ucma 2011)* (Vol. 151, pp. 60–65). Springer. doi: 10.1007/978-3-642-20998-7\_8