What came first, the color or the letter?: An Investigation of Top-Down Feedback and Cross-Activation Models of Synesthesia

Isabella Longoria-Valenzuela and Peter McDermott

Cognitive, Linguistics, and Psychological Sciences Department, Brown University

CLPS 1492: Computational Cognitive Neuroscience

Dr. Michael Frank

December 21st, 2021

Introduction:

Synesthesia is the perceptual phenomenon in which one sensory stimulation leads to the involuntary stimulation of a second sensory experience. In each type of synesthesia, there is a triggering inducer mapped to corresponding synesthetic concurrents (Grossenbacher & Lovelace, 2001). Synesthesia can occur across two different sensory modalities, or within the same modality (Mills, 1999). There are various types of synesthesia, such as visual letter and color, sound and color, sound and taste, etc.

One of the most common types of synesthesia is called grapheme-color synesthesia, in which an individual sees a letter or number and also sees a color associated with that letter or number (Day, 2005). Some synesthetes report seeing the color, some report knowing the color is there, and other's report seeing the color in their "mind's eye" (Dixon et al., 2004) (Flournoy, 1893) (Ramachandran and Hubbard, 2001). The experience of "seeing" the color in conjunction with the presented grapheme is known as "projector" synesthesia, whereas the experience of "knowing" the color is associated with a grapheme is known as "associator" synesthesia.

The neural basis of grapheme-color synesthesia has been widely debated, and there are two prevalent theories: either direct cross activation or disinhibited (top-down) feedback causes the color pathway to be co-activated with the grapheme pathway (van Leeuwen et al., 2011). Given the anatomical finding that the visual word form area (VWFA) lies adjacent to the color processing region (hV4), researchers have proposed that grapheme-color synesthesia exists due to prenatal connections causing direct cross activation of these two brain regions (Ramachandran and Hubbard, 2001, Grossenbacher & Lovelace, 2001). Alternatively, others posit that sensory information is sent to higher-order brain regions in order to be integrated normally, but disinhibited top-down feedback from the integration region to the color pathway causes the

synesthetic experience (Friston, 2005) (Friston and Kiebel, 2009). Furthermore, neuroimaging studies have investigated the neuroanatomy of "projector" and "associator" synesthetes. One study showed that synesthetes, as compared to non-synesthetes, display increased connectivity in the inferior temporal cortex that may underlie the cross-activation theory, or "projector" synesthesia, and other synesthetes display increased connectivity in higher level association areas, which may underlie the top-down feedback theory (Rouw & Scholte, 2007).

Previous studies have employed the Stroop Task in grapheme-color synesthetes to assess the effect synesthesia has on digit processing (Mills 1999). Mills (1999) used digits in the color that the patient associated with the digit, the patient's "photoism." The study found that when the digit's color was mismatched from the photos, the patient took significantly longer to name the color of the digit as compared to the case in which the digit matched the photoism and the case in which the digit was black. (Mills 1999). However, it remains unknown how Mill's (1999) Stroop task may map onto the biological underpinnings of the disinhibited top-down processing theory or cross-activation theory of synesthesia.

In the present study, we use a neural network in different configurations to represent a control patient, the cross-activation theory, or the top-down processing theory of grapheme-color synesthesia in order to replicate the findings of Mills (1999) that synesthetes will demonstrate the Stroop Effect with incongruent, or mismatched, stimuli as compared to neutral and congruent stimuli.

We hypothesize that both computational models of synesthesia (cross-activation and top-down feedback) will exhibit a stronger Stroop Effect than the computational model of a control patient without synesthesia. Specifically in both synesthesia models, synesthetes will demonstrate slower response times in the incongruent trials of the Stroop Task as compared with

the congruent trials. By contrast, we expect that control patients will not show significant differences in response times over the two Stroop Task conditions.

It is also apparent to us that the control and cross-activation configurations of the network may not be ecologically valid because they have no top-down feedback from higher-order cortical processing areas, as is present in a typical human brain. So, to strengthen the validity of our results and subsequent conclusions, we will examine two additional configurations of the network. The first will be identical to the control condition, but with moderate top-down feedback, and the second will be identical to the cross-activation condition, but with moderate top-down feedback. We predict that the more ecologically valid control condition will result in a stronger Stroop Effect as compared to the control condition, and we predict that the more ecologically valid cross-activation condition will result in a stronger Stroop Effect as compared to the cross-activation condition.

Methods:

Model:

To build this neural network of parallel sensory processing in order to capture the effects of synesthesia, we used the framework of a Simple Recurrent Network (Frank et al., 2021). The conceptual basis for this neural network comes from a parallel processing model, similar to the ObjRec (O'Reilly et al., 2020) model shown in **Figure 1**. This network contained an input layer, a hidden layer, a context layer, and an output layer. We created an additional input and output layer and converted the context layer into a second hidden layer. This resulted in a neural network consisting of one two input layers, two hidden layers, two output layers, and an "Associator layer" as shown in **Figure 2**. This model represents two sensory processing

pathways that are, in their default relative weight scale settings. In our case, the left stream is arbitrarily responsible for deciphering "letters" and the right stream is arbitrarily responsible for deciphering "colors," as indicated by our network labels.
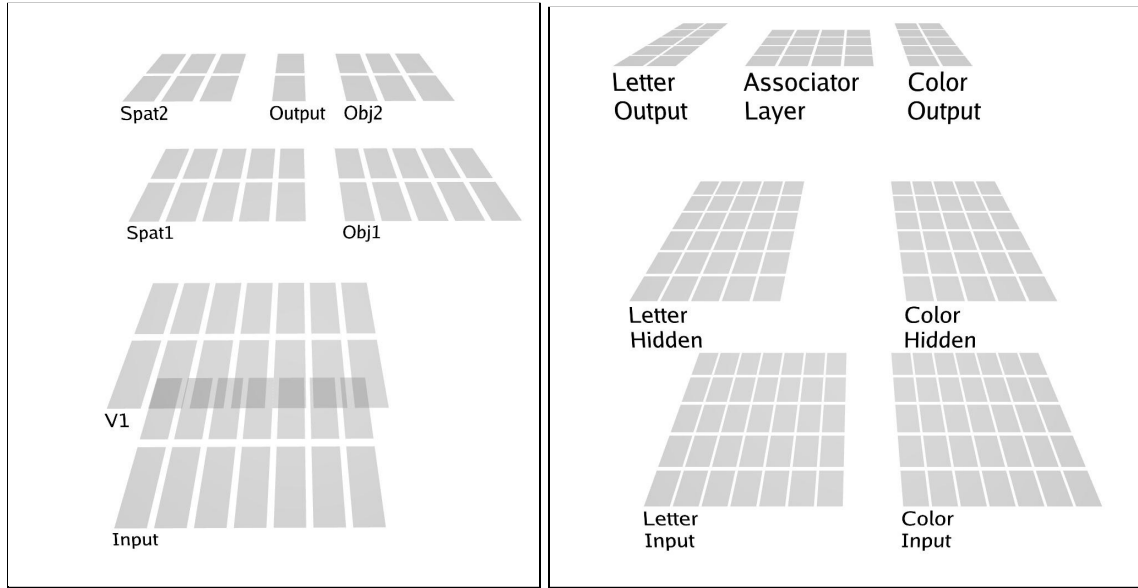

Figure 1: SpatObj Model (O'Reilly et. al, 2020) - Parallel processing object recognition model that served as one of the conceptual and computational bases for our model of synesthesia

Figure 2: Synesthesia Model - Our model consists of two processing streams for color and letter sensory input, as well as a converging output processing level, an "associator layer," that will be activated and investigated during the top-down model of synesthesia.

Data Sets:

We presented the model with two distinct data sets in order to train and subsequently test its performance. The first data set, the "congruent data set," consists of ten distributed input patterns for the letter input layer and ten distributed input patterns for the color input layer, along with the correct output that each output layer should learn to generate. The data set and model work such that a color input and letter input are presented to the network simultaneously. This congruent data set represents the association between colors and letters that a grapheme-color synesthetes may have.

To create this first data set, we used a .dat file from a previous synesthesia final project as a starting point. This file had a 7 x 5 grid for the inputs and a 5x2 grid for the outputs. In this file, the letter"A" (a distributed representation in one grid) was paired with the number "0" (a distributed representation in the other grid), "B" was paired with "1", etc, as shown in **Figure 3**. To modify this file for our project, we changed the column headers to match our pathway streams, respectively labeling them "letter input" and "color input," etc. In order to conceptually model grapheme-color synesthesia, the number inputs represent the color that a synesthete would associate with the letter.

Figure 3: Training Pattern for Congruent Data Set

The second data set, the "incongruent data set," uses the same input patterns as the first data set, however each letter input is paired with a different color input. For instance, if the first dataset presented A with red, B with blue, and C with green, the second data set presented A with blue, B with green, and A with red.

In order to create this second data, we leveraged a .dat file from a previous project. This .dat file used two 5x7 grids as the inputs and two 5x2 grids as the outputs. The inputs were a

distribution of activated units that visually represented letters and numbers. As shown in **Figure 4,** The input for letter A was paired with the number 1, the letter B was paired with the number 2, etc.
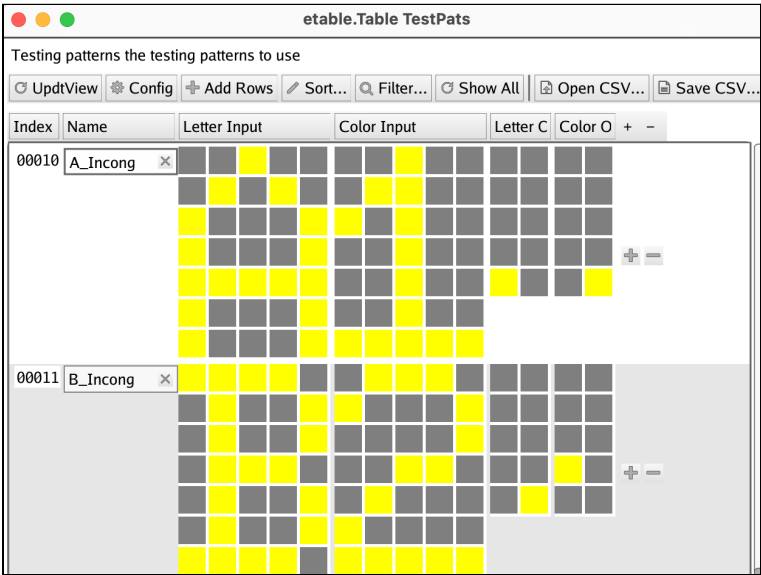

Figure 4: Training Pattern for Incongruent Data Set -

Experimental Conditions:

Our experimental conditions consisted of three different configurations of the network. The first, Isabella's subproject, was a model of synesthesia using lateral connections between the two sensory hidden layers. This model represents activation from one sensory processing stream "bleeding" over into a different processing stream, causing each sensory output to be more heavily influenced by opposing sensory input. The second condition, Peter's subproject, was of a model of synesthesia using a converging output processing layer between the two sensory outputs. This model represents a "higher order" cortical processing layer that provides disproportionate amounts of top-down feedback to the two sensory outputs, causing each sensory output to be more heavily influenced by the other.

Isabella's Subproject: Cross-Activation

In order to create this first experimental condition, I coded bilateral connections between the color and the letter hidden layers. In this way, the representation of one letter will influence the representation of an associated color and vice versa, eventually influencing the output of each processing stream. By having each hidden layer influence the other, I hoped to capture the essence of the cross-activation theory of grapheme-color synesthesia, in that one processing stream would influence a different, distinct processing stream by having activity "bleed" over. So, the implementation of connections between the hidden layers allowed this computational model to closely reflect Grossenbacher & Lovelace's (2001) model of synesthesia.

To create these connections, I repurposed code from the 1492 Midterm that initially set up bidirectional inputs between output and hidden layers and reworked them to do the same between the two sensory hidden layers of our model. I then introduced an adjustable weight scale for these lateral connections and added the scale to the graphical user interface (GUI) to manually strengthen and weaken the weights during testing.

However, after testing the model, I found that bidirectional connections between these hidden layers were causing our network to crash and did not effectively illustrate our intended results. After getting some advice from the CLPS1492 teaching staff, I decided to change the connection to be unidirectional, from color hidden to the letter hidden layer. This change would still result in the same intended experimental design, since one sensory pathway would still be influencing the other.

Peter's Subproject: Top-down Feedback (Activator)

In order to create the second experimental condition, I added another layer that receives and sends information to and from the two output layers. This layer served as a converging output processing layer, a higher order cortical area where information from distinct sensory

processing streams converges. Otherwise known as an "associator" layer, this layer provides top-down feedback onto the distinct sensory processing streams. So, introducing this layer renders our model more human-like. However, the primary reason for the associator layer's inclusion is to explore the "associator" theory of synesthesia, in which an associator layer provides disinhibited feedback to distinct sensory pathways, therefore permitting one sensory stream to have a disproportionately large impact on a different sensory stream (van Leeuwen et al., 2011).

By adding in the associator layer to the code, the two processing streams complete their own transformation of a distributed input representation to a distinct output. But, as those outputs converge, they are transformed into a unique representation in the associator layer, which is then subsequently fed back into the output layers, therefore affecting the final output of each stream. Further, I introduced an adjustable weight scale into the code and into the graphical user interface (GUI) so that we could easily adjust the weight scales between the layers for various experimental conditions.

However, implementing the adjustable weights into the GUI proved to be difficult in the current model. So, I adjusted the connectivity such that there was a unidirectional connection from the Color Output to the Associator Layer and from the Associator Layer to the Letter Output. In this way, I simplified the model such that one stream, the color stream, modifies another stream, the letter stream, but not vice versa. This simplification allowed for me to implement the adjustable relative weight scales into the GUI, allowing me to conduct the tests on the associator condition.

Overview of Network Configurations:

The first configuration of the network is the "control state" where the weight scale of the lateral connections between hidden layers and bidirectional connections between the output layers and converging layers is turned off. The result is a network with two distinct sensory processing pathways with no weights between them.

The second configuration of the network is the "cross-activation state" where the weight scale between the hidden layers is non-zero, but the weight scale from the output layers to the converging layer is still zero. The result of this configuration is a network with two processing pathways that allows activity to bleed from one stream to the other.

The third configuration of the network is the "associator state" where the weight scale between the output layers and the converging layer is non-zero whereas the lateral connections between the hidden layers is zero. The result of this configuration is a network with two distinct processing pathways that complete the transformation of inputs separately, but converge at a downstream point that subsequently influences the output layers' output.

Procedure:

For each configuration or experimental condition, Control State, Cross-Activation State, and Associator State, we present the network with a version of the Stroop Task (Mills, 1999). First, we present the network with the congruent data set, where each "letter" is paired with a particular "color." As previously explained, the inputs for letters and colors are distributed representations.

Following a training period where the network learns to produce the correct output given the corresponding input, we present the network with the congruent and incongruent data sets. So, the network will be presented with neutral stimuli that are associated with no other inputs,

"familiar" combinations, "unfamiliar" combinations, in this case the "letter" will be presented with a different "color" with which it was previously presented.

Data Collection:

We recorded the response time for the network to settle on the correct answer when given familiar combinations of "letters" and "colors" (congruent data set) and when given unfamiliar combinations of "letter" and "colors" (incongruent data set).

To investigate which model of synesthesia may be more biologically plausible, we compared the average "reaction time," as measured by epochs to arrive at zero percent error, of correct response for the congruent and incongruent data sets. This difference in average reaction times will serve as a measure of the "Stroop Effect."

Finally, we compared the "Stroop Effects" for each experimental condition (Control State, Cross-Activation State, and Associator State). From here, we will compare the results to the general Stroop Task data in actual synesthetes, as found in Mill's (1999) study.

Results:

Data Analysis (Stroop Effect calculation):

For our analyses, we calculated the size of the Stroop Effect by subtracting the number of cycles the network took to settle for the congruent pairings from the number of cycles the network took to settle for the incongruent pairings.

Control Condition Model:

To assess whether the control experimental condition displayed a significant Stroop Effect, we set all the adjustable weight scales to 0 (Letter Hidden and Color Hidden layers, Color Output to Associator, and Associator to Letter Output). As seen in **Figure 5**, we found that the

average number of cycles the network took to settle in the incongruent trial type, 14 cycles, was not significantly greater than the average number of cycles the network took to settle in the congruent trial type, 14.2 cycles. Therefore, the control condition did not exhibit a significant Stroop Effect.
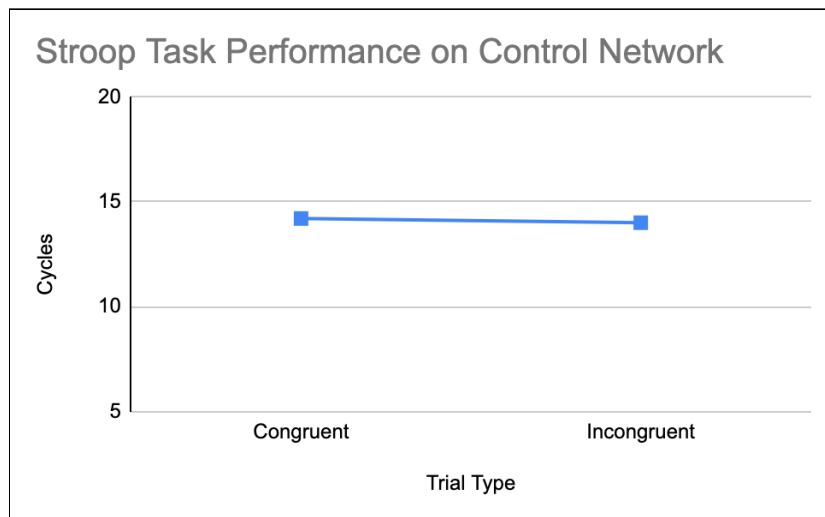

Figure 5: Control Condition Performing the Stroop Task - The graph displays the cycle number for each of the 10 incongruent and 10 congruent letter (A-J) and number (0-9) pairings shown during the control testing condition.

Control Condition with Top-Down Feedback Model:

To investigate whether the control condition with top-down feedback displayed a significant Stroop Effect, we set the weight scale between the Letter Hidden and Color Hidden layers to 0, and set the weight scale between the Letter and Color Outputs and the Associator layer to 0.5, respectively. The presence of the top-down weights was included to mimic a more ecologically-valid sensory pathway where top-down feedback would influence sensory output. We found that the average number of cycles the network took to settle in the Incongruent trial type, 13.9 cycles, was not significantly greater than the average number of cycles the network took to settle in the Congruent trial type, 14 cycles, as illustrated by our graph in **Figure 6**.

Similar to the control condition, the control condition with the added top-down feedback did not exhibit a significant Stroop Effect.
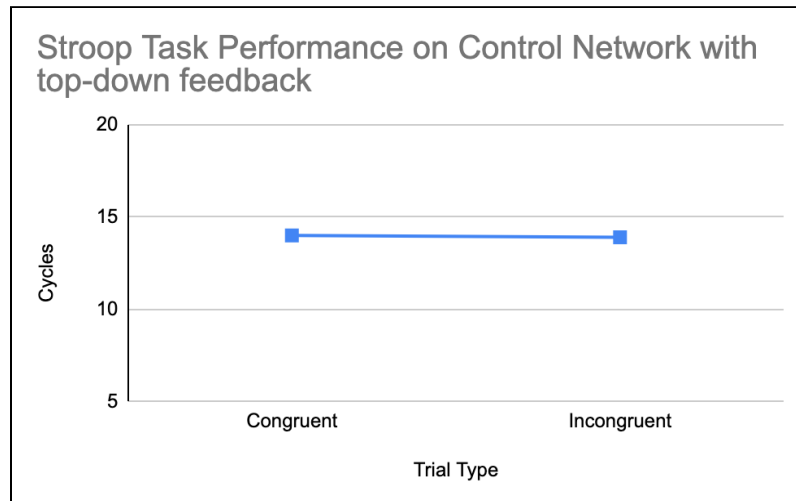

Figure 6: Control Condition Performing Stroop Task with Top-Down Feedback - The graph displays the cycle number for each of the 10 incongruent and 10 congruent letter (A-J) and number (0-9) pairings shown during the control with top-down feedback testing condition.

Isabella: Cross-Activation Model

In order to assess whether the cross-activation model of synesthesia displayed a significant Stroop Effect, I turned on the weight scale between the Letter Hidden and Color Hidden layer to 1, and turned off the weight scale between the Letter and Color Outputs and the Associator layer. As shown in **Figure 7**, my results indicate that the average number of cycles the network took to settle in the Incongruent trial type, 26.2 cycles, was significantly greater than the average number of cycles the network took to settle in the Congruent trial type, 14.4 cycles. Therefore, the cross-activation model of synesthesia exhibited a significant Stroop Effect.
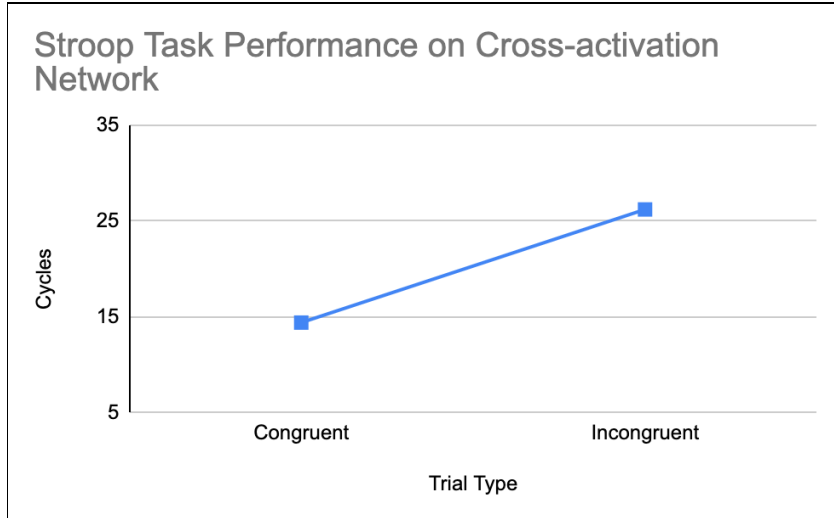
Figure 7: Cross-Activation Model of Synesthesia Performing Stroop Task with Top-Down Feedback - The graph displays the cycle number for each of the 10 incongruent and 10 congruent letter (A-J) and number (0-9) pairings shown during the cross-activation condition testing.

Cross-Activation with Top-Down Feedback Model:

In order to assess whether the cross-activation condition with top-down feedback displayed a significant Stroop Effect, we turned on the weight scale between the Letter Hidden and Color Hidden layers to 1, and turned on the weight scale between the Letter and Color Outputs and the Associator layer to 0.5, respectively. As displayed in **Figure 8**, we found that the average number of cycles the network took to settle in the Incongruent trial type, 17.3 cycles, was not significantly greater than the average number of cycles the network took to settle in the Congruent trial type, 14.8 cycles. Therefore, the cross-activation condition with top-down feedback exhibited a slight Stroop Effect, though it was significantly smaller than that of the
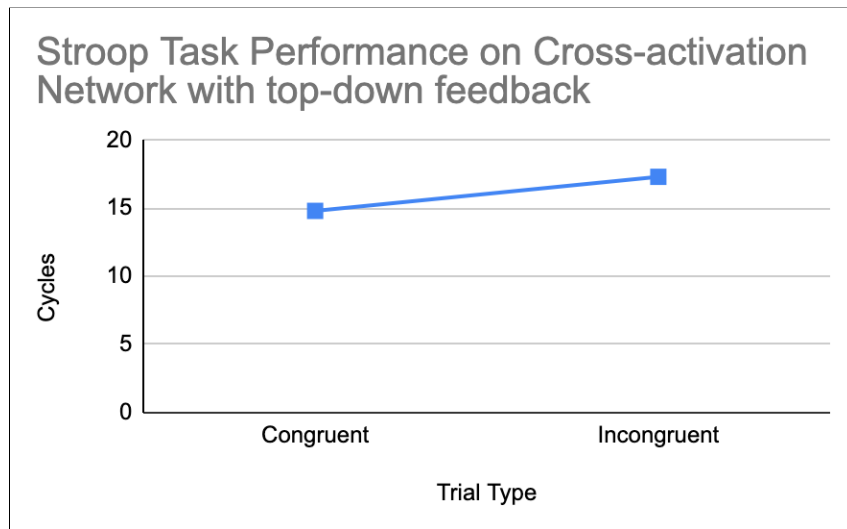
cross-activation condition feedback.

Figure 8: Cross-Activation Model of Synesthesia Performing Stroop Task with Ecologically-Valid Top-Down Feedback - The graph displays the cycle number for each of the 10 incongruent and 10 congruent letter (A-J) and number (0-9) pairings shown during the cross-activation condition with top-down feedback testing.

Peter: Associator Model

In order to assess whether the disinhibited top-down feedback condition displayed a significant Stroop Effect, I turned off the relative weight scale between the Letter Hidden and Color Hidden layers and increased both relative weight scales between the Letter and Color Outputs and the Associator layer to 1. My results illustrate that the average number of cycles the network took to settle in the Incongruent trial type, 16.2 cycles, was significantly greater than the average number of cycles the network took to settle in the Congruent trial type, 14.3 cycles, as demonstrated in **Figure 9**. Therefore, the disinhibited top-down feedback condition exhibited a

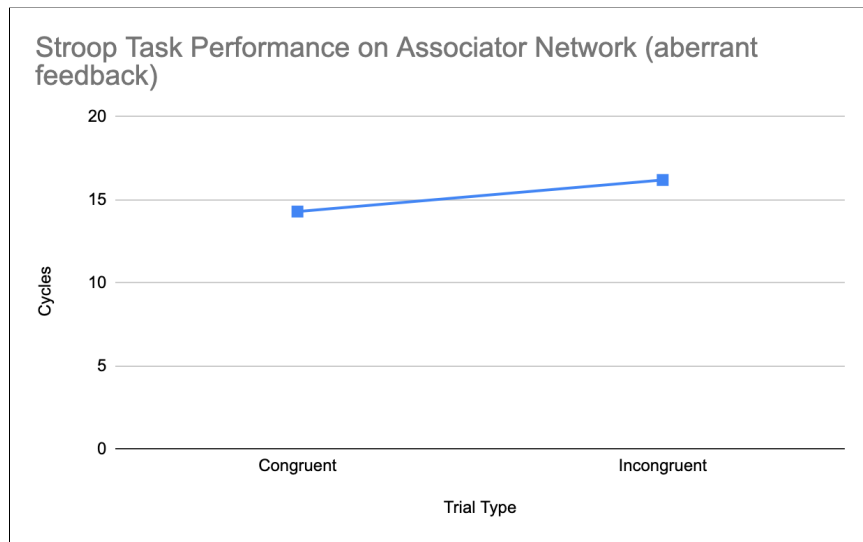slight Stroop Effect compared to the other experimental conditions.

Figure 9: Associator Model of Synesthesia Performing Stroop Task - The graph displays the cycle number for each of the 10 incongruent and 10 congruent letter (A-J) and number (0-9) pairings shown during the associator condition testing.

Finally, in order to compare Stroop Task performance of all experimental conditions, we created a graph that overlaid each condition. As shown in **Figure 10**, the cross-activation condition without feedback exhibited the most drastic Stroop Effect, with the cross-activation model with top-down feedback and the associator model displaying a slight Stroop Effect.
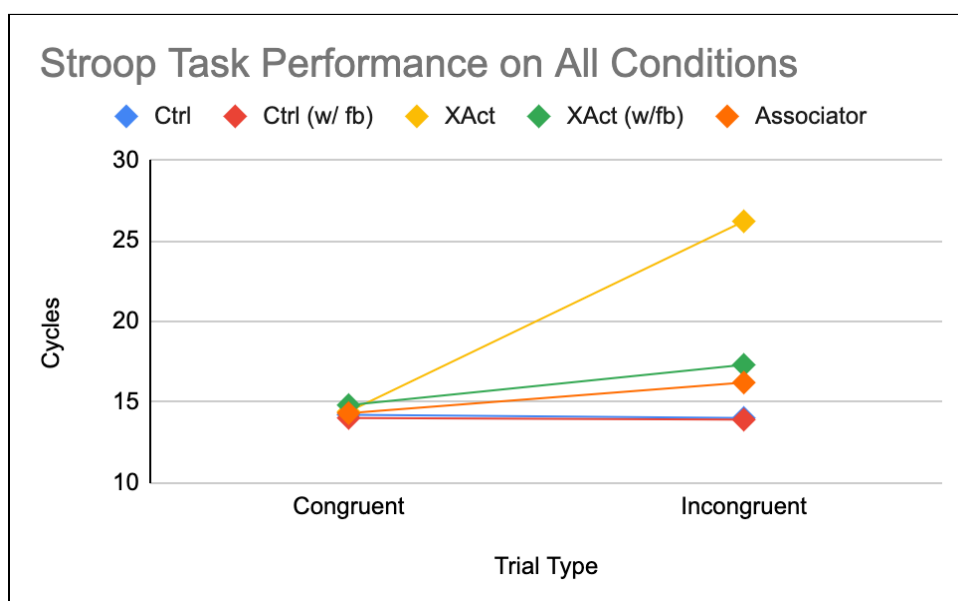
Figure 10: Associator Model of Synesthesia Performing Stroop Task - The graph displays the cycle number for each of the 10 incongruent and 10 congruent letter (A-J) and number (0-9) pairings shown during all five testing conditions.

Discussion:

In the present study, we investigated the performance of two distinct computational models of synesthesia, top-down and cross-activation. Specifically, we explored the potential neurobiological validity of these models and how each would perform on Mill's (1999) Stroop Task as opposed to a control patient without synesthesia. In line with our hypothesis, the cross-activation model of synesthesia showed decreased network performance with incongruent letter-number pairing compared to congruent pairing. Additionally, the control condition did not display a significant Stroop Effect, as expected. To our surprise, the associator model only exhibited a slight Stoop Effect. Furthermore, the cross-activation condition with top-down feedback also only showed a slight Stroop Effect, whereas the control condition with top-down feedback did not show a significant Stroop Effect.

Our computational model provides further evidence that the cross-activation theory of grapheme-color synesthesia can be accurately modeled in a simple neural network involving two separate sensory streams and lateral connections between the two streams. However, our findings indicate that our model was not able to effectively capture the observed performance in synesthetes on Stroop Task performance using the associator model.

Isabella's results demonstrate that the cross-activation model displayed a Stroop Effect that was significantly larger than the control condition. Our data closely mirrored the size of the Stroop Effect found in Mills' (1999) on a patient with synesthesia. From this finding, we suggest that the cross-activation model is a good representation of the biological underpinnings of synesthesia. By contrast, Peter found that the disinhibited top-down feedback condition only

produced a slight Stroop Effect, but one that is still larger than the control condition and the control condition with top-down feedback. However, this slight Stroop Effect was not in line with the findings of Mills (1999).

In the two conditions in which we set the relative weight scales for the top-down feedback pathway to 0.5, the control with feedback and the cross-activation with feedback conditions, the top-down feedback seem to dampen the Stroop Effect that was present in the version of the condition without top-down feedback. However, in the condition where we set the relative weight scale for the top-down feedback pathway to 1, the associator with the disinhibited feedback condition, we observed a slight Stroop Effect which was larger than the control condition. Interestingly, we included the conditions with added moderate top down feedback to make our network more ecologically valid, so the result that moderate top down feedback helped the network perform on the Incongruent trials is in line with our logic on the surface level. However, given how simple our model is, and given how much more complex actual human brains are, we do not believe we accurately captured the top-down activation that normally occurs in the human brain. Further, the slight Stroop Effect we observed in the associator with disinhibited top down feedback is in line with our hypothesis, but it is a slight Stroop Effect as compared to that which was observed in the cross-activation condition.

Additionally, our results perhaps illustrate the processing complexities of synesthesia. Specifically, during testing, we noticed that some letters appeared to take more cycles to learn than others. For example, in the cross-activation model, the letter C took 76 cycles whereas the letter B only took 19 cycles. This could be due to the randomization of the sending weights of input units before training that could subsequently affect learning, or the extent to which random inputs overlapped with one another. However, we believe that this might reveal a methodological

flaw in our study's design. Perhaps the specific representations of certain letters and number combinations used in our study, depending on how many input units they activate, pose more of a challenge for synesthetes than other letters and numbers. Furthermore, the "ease" with which each synesthete learns these associations may be unique, similar to our experimental models. To remedy this, other forms of synesthesia should be explored beyond grapheme-color synesthesia to account for input variability.

If we had more time, we would have found a way to present the network with the control data set in addition to the congruent and incongruent data sets in order to further replicate Mills' findings (Mills 1999). Further, an interesting expansion of this study would be to add another data set that presented letter/number inputs as probabilistic representations. This data set would mimic an environmentally developed form of synesthesia, where an individual develops a slight bias for certain letter/number combinations. In terms of our code, we could create a data set with 100 letters, with 10 of each letter (A-J) but this time, 9 out of 10 of each letter would go with the "congruent" number response but 1 of 10 of each letter would go with the incongruent number response. This would result in a statistical association between letter/number combinations, which ultimately might lead to slight slowed performance on the Stroop Task.

Importantly, these combined findings could help guide future research exploring the origins of synesthesia. If future studies also support the synesthesia models proposed by Grossenbacher and Lovelace (2001) and van Leeuwen et al. (2011), perhaps treatment options that target higher-order cortical areas or connectivity in lower-level and proximally close sensory brain regions should be investigated.

Other reference images of our model with receiving weights between layers highlighted:
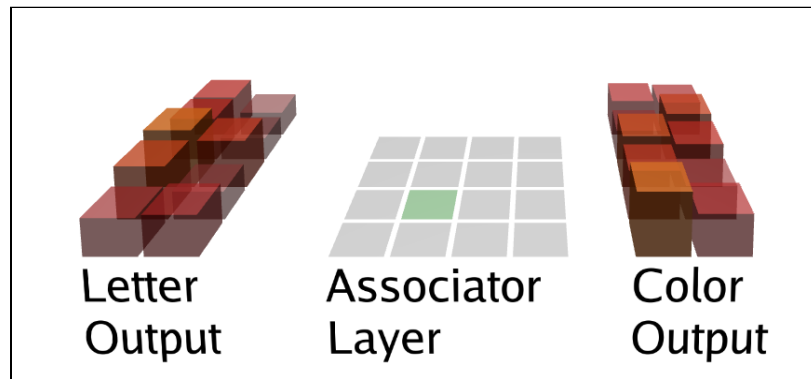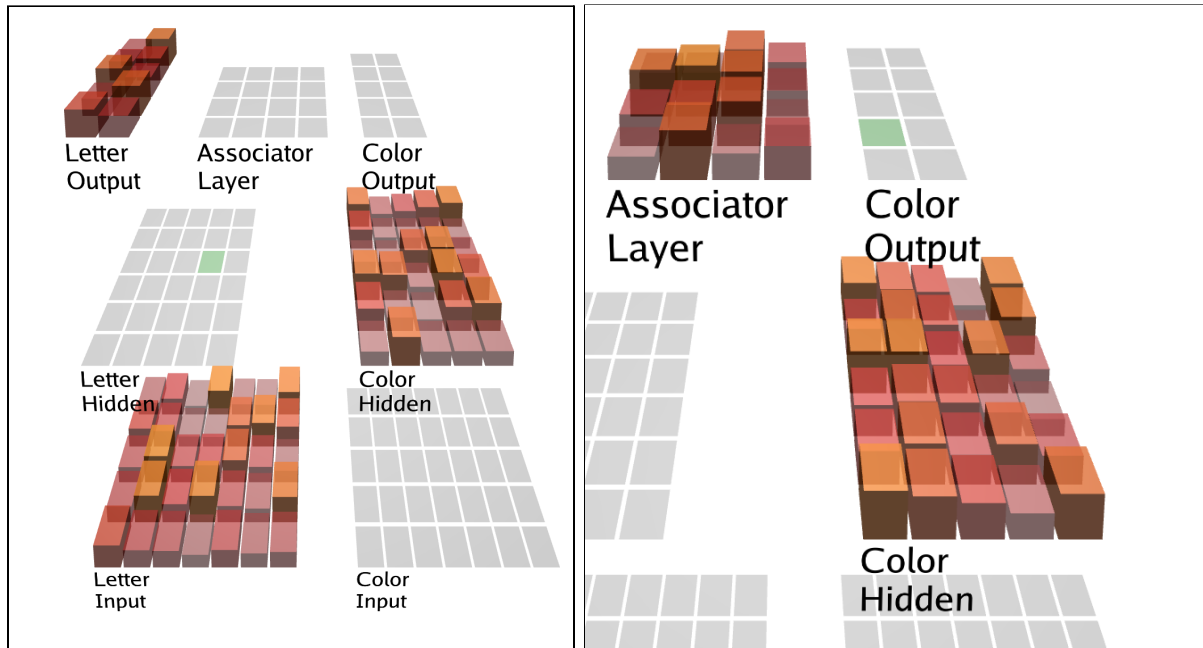
References:

Day, S. (2005). Some Demographic and Socio-cultural Aspects of Synesthesia. In L. C. Robertson & N. Sagiv (Eds.), *Synesthesia: Perspectives from cognitive neuroscience* (pp. 11–33). Oxford University Press.

Dixon, M. J., Smilek, D., & Merikle, P. M. (2004). Not all synaesthetes are created equal: Projector versus Associator Synaesthetes. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(3), 335–343. https://doi.org/10.3758/cabn.4.3.335

Grossenbacher, P. G., & Lovelace, C. T. (2001). Mechanisms of synesthesia: Cognitive and physiological constraints. *Trends in Cognitive Sciences*, *5*(1), 36–41. https://doi.org/10.1016/s1364-6613(00)01571-0

Flournoy, T. (1893). Des phénomenes de synopsie [On the phenomena of synopsia]. *Geneve: Charles Eggimann & Co*.

Friston K, (2009) Predictive coding under the free-energy principle. *Philos Trans R Soc Lond B Biol Sci* 364:1211–1221.

Hubbard, E. M., & Ramachandran, V. S. (2005). Neurocognitive mechanisms of synesthesia. *Neuron*, *48*(3), 509–520. https://doi.org/10.1016/j.neuron.2005.10.012

Kiebel S (2009) Predictive coding under the free-energy principle. *Philos Trans R Soc Lond B Biol Sci* 364:1211–1221.

Mills, C. B. (1999). Digit synaesthesia: A case study using a stroop-type test. *Cognitive Neuropsychology*, *16*(2), 181–191. https://doi.org/10.1080/026432999380951

Ramachandran, V. S., & Hubbard,, E. M. (2002). Synaesthesia -- A window into perception, thought and language. *Journal of Consciousness Studies*, *8*(12), 3–34(2).

Rouw, R., & Scholte, H. S. (2007). Increased structural connectivity in grapheme-color synesthesia. *Nature Neuroscience*, *10*(6), 792–797. https://doi.org/10.1038/nn1906

van Leeuwen, T. M., den Ouden, H. E., & Hagoort, P. (2011). Effective connectivity determines the nature of subjective experience in grapheme-color synesthesia. *Journal of Neuroscience*, *31*(27), 9879–9884. https://doi.org/10.1523/jneurosci.0569-11.2011

Simulations:

Objrec Model: https://github.com/CompCogNeuro/sims/tree/master/ch6/objrec

Simple Recurrent Network (SRN): http://ski.clps.brown.edu/cogsim/SRN_Midterm2021.zip