When forgetting fosters learning: A saliency map for TP computations - Simulations of Giroux & Rey, 2009

Ansgar D. Endress, City, University of London

Abstract

NA

1 Experiments with a basic stream: Words and part-words, tested forwards and backwards

A documented version of this model can be found in [Endress and Johnson, 2021].

In line with [Giroux and Rey, 2009] experiment, we create streams consisting of two three-syllable words and four two-syllable words. These units are randomly concatenated into a familiarization stream so that each unit occurs 143 times. We will then present the network with test-items (see below) and record the total network activation while each item is presented. We hypothesize that the total activation provides us with a measure of the network's familiarity with the unit.

This cycle of familiarization and test will be repeated 100 times, representing 100 participants.

While we keep the parameters for self-excitation constant (α and β in Supplementary Material XXX), we used forgetting rates (λ_{act} in Supplementary Material XXX) between 0, 0.2, 0.4, 0.6, 0.8, 1 and inhibition rates between 0.4, 0.6, 0.8, 1. As forgetting in our model is exponential, a forgetting rate of zero means no forgetting, a forgetting rate of 1 implies the complete disappearance of activation on the next time step (unless a population of neurons receives excitatory input from other populations), and a forgetting rate of .5 implies the decay of half of the activation.

2 Resuls

For each comparison, we will create normalized difference scores to evaluate the model performance:

$$d = \frac{\text{Item}_1 - \text{Item}_2}{\text{Item}_1 + \text{Item}_2}$$

We then evaluate these difference scores against the chance level of zero using Wilcoxon tests.

As in [Giroux and Rey, 2009], there were two types of sub-units resulting from an *ABC* unit: *AB* sub-units and *BC* sub-units. As shown in Figure **??**, when averaging across trials comparing two-syllable units to *AB* and *BC* sub-units, there was a significant preference for units for most parameter sets (except for some simulations with low inhibition rates).

However, as shown in Figure **??**, while units were systematically preferred over *AB* sub-units for most parameter values, *BC* sub-units could be preferred for very low or very high interference rates.

0 ‘***’ 0.001 ’’*** *0.01* ” 0.05 ‘.’ 0.1 ’ ’ 1

To support our contention that the preference for units over sub-units might arise from the interplay between learning (and thus excitation) and inhibition, we plot in Figure **??** weights between different pairs of neurons

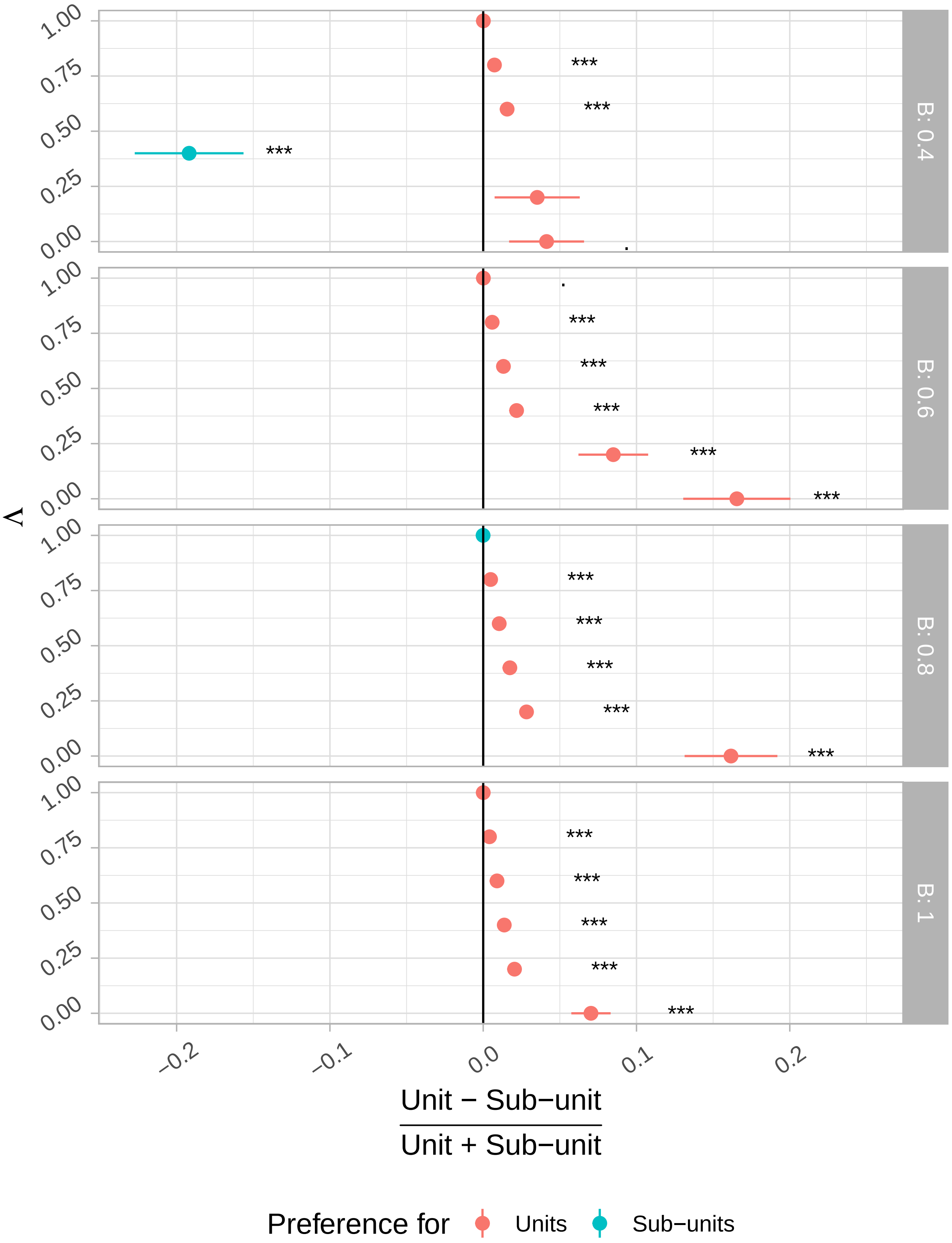

Figure 1: Normalized difference scores of network activations after presentation of entire two-syllable units and different types of two-syllable units (i.e., AB and BC from ABC units), as a function of the forgetting rate (y axis) and the interference rate (rows). The rightmost column is the average of the other columns reported by Giroux2009. Positive values indicate stronger activations for units. Significance stars reflect a Wilcoxon test against the chance level of zero. Units generally elicit greater activation compared to AB subunits and compared to the average; when compared to BC units, the sign of the difference score depends on the parameters.

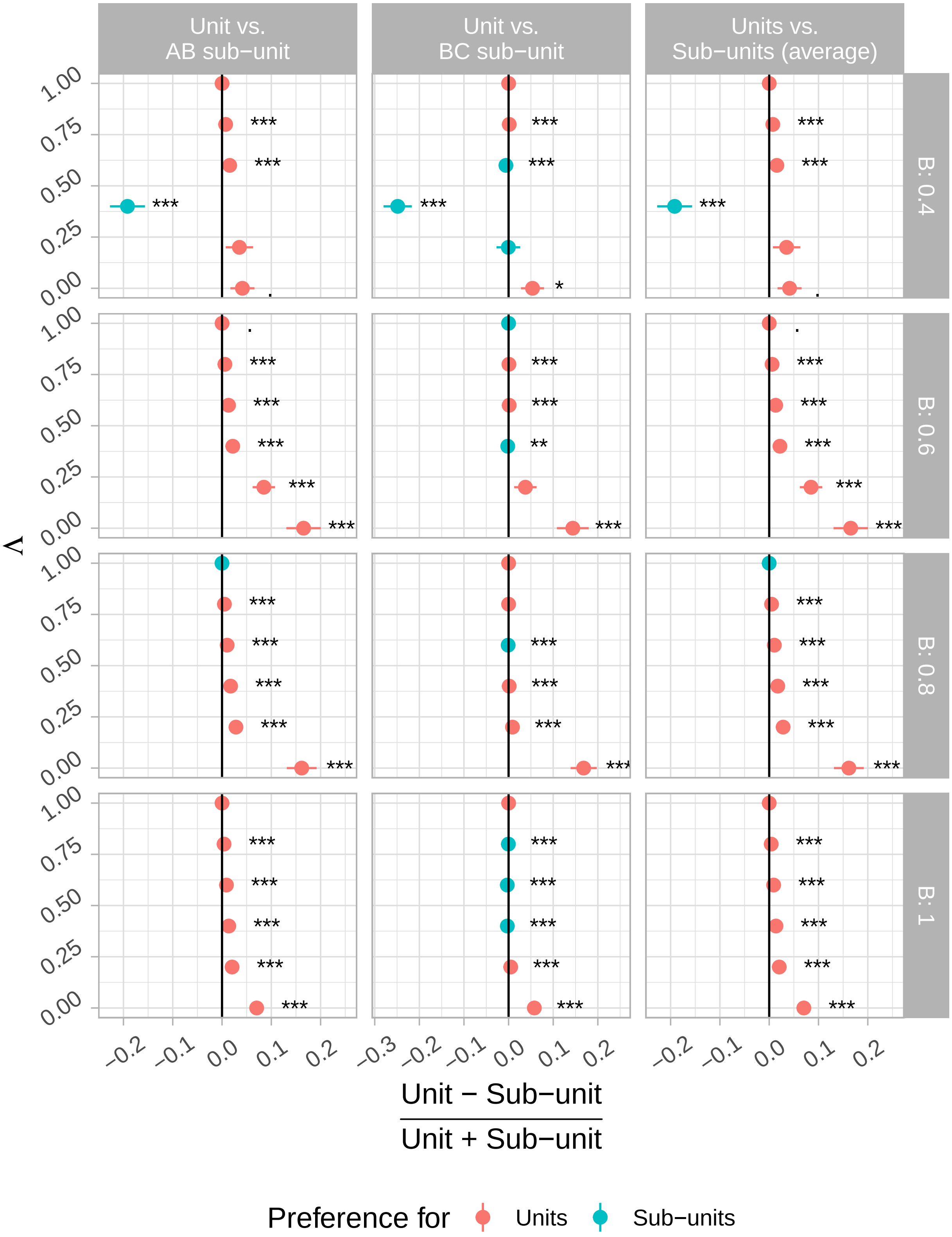

Figure 2: Normalized difference scores of network activations after presentation of entire two-syllable units and different types of two-syllable units (i.e., AB and BC from ABC units), as a function of the forgetting rate (y axis) and the interference rate (rows). The rightmost column is the average of the other columns reported by Giroux2009. Positive values indicate stronger activations for units. Significance stars reflect a Wilcoxon test against the chance level of zero. Units generally elicit greater activation compared to AB subunits and compared to the average; when compared to BC units, the sign of the difference score depends on the parameters.

after learning. As suggested above, the connection between A and C in a three-syllable ABC unit is generally weaker than other connections, and often substantially smaller than the interference rate. Depending on the parameter values, (second order) activation of C might thus suppress activation in AB sub-units, and activation of A might suppress activation in BC sub-units. However, the exact computational mechanisms, as well as the differences in behavior betwee AB and BC sub-units deserve further investigation. For the current purposes, we just conclude that the fact that a simple Hebbian learning model can account for a preference for units over sub-units demonstrates that such results do not provide evidence that units have been placed in memory.

References

Ansgar D Endress and S P Johnson. When forgetting fosters learning: A neural network model for statistical learning. *Cognition*, 104621, 2021. doi: 10.1016/j.cognition.2021.104621.

Ibrahima Giroux and Arnaud Rey. Lexical and sublexical units in speech perception. *Cognitive science*, 33: 260–272, March 2009. ISSN 0364-0213. doi: 10.1111/j.1551-6709.2009.01012.x.

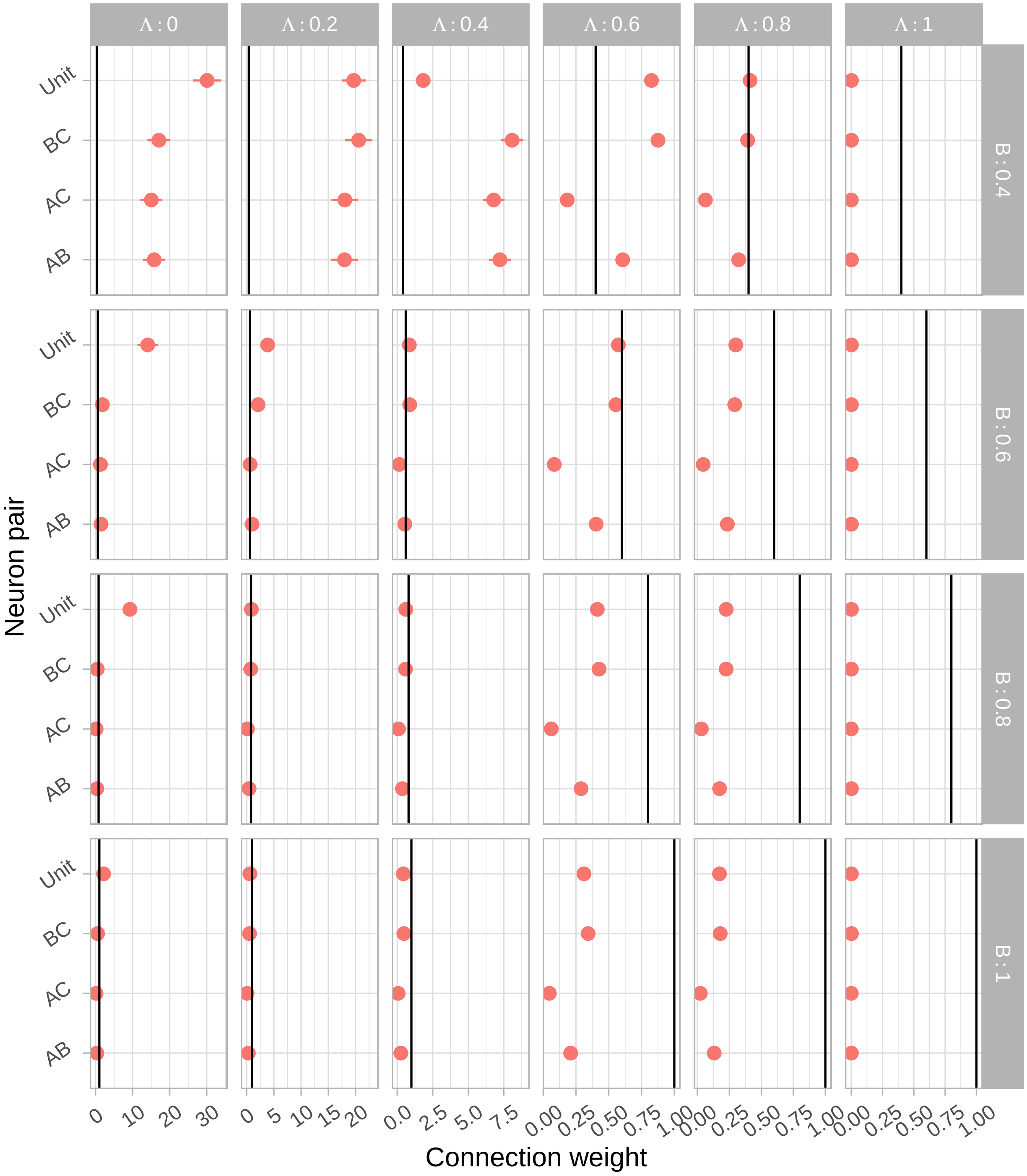

Figure 3: Connection weights between different pairs of neurons as a function of the forgetting rate (columns) and the interference rate (rows). The figure shows connection weights within a trisyllabic unit (ABC) and a bisyllabic unit (BC). The black line represents the interference rate. The A-C connection is generally smaller than the other connections, and often substantially smaller than the interference rate.