Supplementary Material for: When forgetting fosters learning: A neural network model for Statistical Learning

Ansgar D. Endress
Scott P. Johnson
Department of Psychology, City, University of London, UK
Department of Psychology, UCLA

Contents

An R implementation of the model is available at `http://doi.org/10.25383/city.11359376`. This research was support by NIH grant R01-HD073535 to SPJ.

Supplementary Material A
Model definition

The activation of the i^{th} unit $x_i(t)$ is governed by the differential equation.

$$\dot{x_i} = -\lambda_a x_i + \alpha \sum_{j \neq i} w_{ij} F(x_j) - \beta \sum_{j \neq i} F(x_j) + \text{noise}$$

where $F(x)$ is some activation function. (Here we use $F(x) = \frac{x}{1+x}$). The first term represents exponential forgetting with a time constant of λ_a, the second term activation from other units, and the third term inhibition among items to keep the overall activation in a reasonable range.

The weights w_{ij} are updated using a Hebbian learning rule

$$\dot{w}_{ij} = -\lambda_w w_{ij} + \rho F(x_i) F(x_j)$$

λ_w is the time constant of forgetting (which we set to zero in our simulations) while ρ is the learning rate.

A discrete version of the activation equation is given by

$$x_i(t+1) = x_i(t) - \lambda_a x_i(t) + \alpha \sum_{j \neq i} w_{ij} F(x_j) - \beta \sum_{j \neq i} F(x_j) + \text{noise}$$

While the time step is arbitrary in the absence of external input, we use the duration of individual units (e.g., syllables, visual symbols etc.) as the time unit in our discretization because associative learning is generally invariant under temporal scaling of the experiment (Gallistel & Gibbon, 2000). Further, while only excitatory connections are tuned by learning in our model, the same effect could be obtained by tuning inhibition, for example through tunable disinhibitory interneurons (Letzkus et al., 2011). Here, we simply focus on the result that a fairly generic network architecture accounts for the hallmarks of statistical learning that, so far, have eluded explanation.

The discrete updating rule for the weights is

$$w_{ij}(t+1) = w_{ij}(t) - \lambda_w w_{ij}(t) + \rho F(x_i) F(x_j)$$

Simulation parameters are listed in Table A1. An *R* implementation is available at `http://doi.org/10.25383/city.11359376`.

Table A1
Parameters used in the simulations

Symbol	**Function**	**Value(s)**
α	Excitation coefficient	0.7
β	Inhibition coefficient	0.4
λ_a	Forgetting rate — Activation	0, 0.2, 0.4, 0.6, 0.8, 1
λ_w	Forgetting rate — Weights	0
$\sigma_{\text{noise, activation}}$	Standard deviation of activation noise	0.001
$\sigma_{\text{noise, weights}}$	Standard deviation of weight noise	0
ρ	0.05	

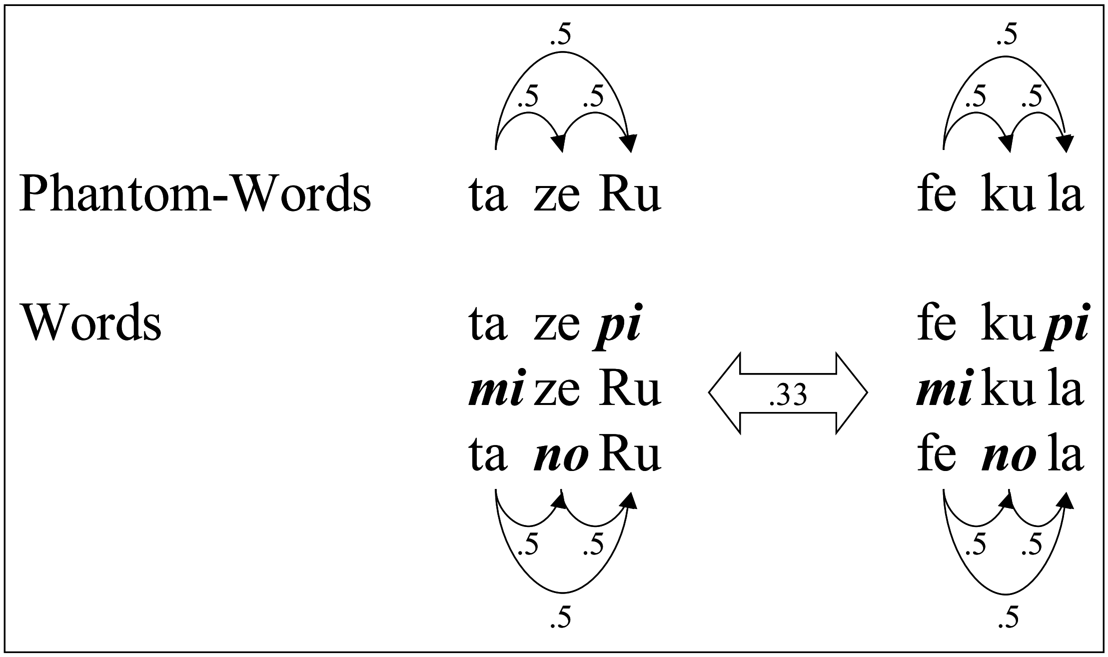

Figure B1. Design of the stimuli used in Endress and Mehler's (2009) experiments. Reproduced from Endress and Mehler (2009).

Supplementary Material B
Design of phantom-words

The design of Endress and Mehler's (2009) experiments is shown in Figure B1. During familiarization, participants listened to continuous speech streams consisting of a concatenation of nonce words.

Critically, the "words" were constructed to equate TPs among syllables in words and in "phantom-words," that is, in items that did not occur in the stream but had the same TPs as words.

As shown in Figure B1, phantom-words were constructed through overlap with actual words in the speech stream. Specifically, each phantom-word shared the first two syllables with one word, the last two syllables with another word, and the first and last syllable with yet another word. (Syllables not shared between words and the corresponding phantom-words are shown in bold face in Figure B1.) As result, TPs among adjacent and non-adjacent syllables were identical in both words and phantom-words. Within-word TPs were 0.5, while TPs among syllables straddling a word boundary were 0.33.

Supplementary Material C
Detailed results

Table C1 provides detailed results for the simulations in terms of descriptive statistics and statistical tests for the simulation testing the recognition of (forward and backward) units, part-units, rule-units and class-units.

Table C2 provides similar results for the simulations testing the recognition of units, part-units and phantom-units.

Table C1

Detailed results for the different forgetting rates and comparisons (Unit vs. Part-Unit: ABC *vs.* BC:D *and* ABC *vs.* C:DE*; Rule-Unit vs. Class-Unit:* AGC *vs.* AGF *and* AXC *vs.* AXF*), for items presented in forward and backward order, and using the global activation as a measure of the network's familiarity with the items.* $p_{Wilcoxon}$ *represents the* p *value of a Wilcoxon test on the difference scores against the chance level of zero.* $P_{Simulations}$ *represents the proportion of simulations showing positive difference scores.*

λ_a	Statistic	ABC vs BC:D	ABC vs C:DE	AGC vs AGF	AXC vs AXF
Forward					
0	*M*	-180×10^{-3}	-113×10^{-3}	-82.7×10^{-3}	-101×10^{-3}
0	*SE*	-18.1×10^{-3}	-11.4×10^{-3}	-8.31×10^{-3}	-10.2×10^{-3}
0	$p_{Wilcoxon}$	95.7×10^{-3}	222×10^{-3}	452×10^{-3}	607×10^{-3}
0	$P_{Simulations}$	470×10^{-3}	540×10^{-3}	490×10^{-3}	570×10^{-3}
200×10^{-3}	*M*	-109×10^{-3}	-72.8×10^{-3}	-92.6×10^{-3}	-87.1×10^{-3}
200×10^{-3}	*SE*	-11.0×10^{-3}	-7.32×10^{-3}	-9.31×10^{-3}	-8.75×10^{-3}
200×10^{-3}	$p_{Wilcoxon}$	120×10^{-3}	118×10^{-3}	152×10^{-3}	134×10^{-3}
200×10^{-3}	$P_{Simulations}$	490×10^{-3}	530×10^{-3}	540×10^{-3}	510×10^{-3}
400×10^{-3}	*M*	3.68×10^{-3}	102×10^{-3}	12.4×10^{-3}	13.2×10^{-3}
400×10^{-3}	*SE*	369×10^{-6}	10.2×10^{-3}	1.25×10^{-3}	1.33×10^{-3}
400×10^{-3}	$p_{Wilcoxon}$	2.92×10^{-12}	3.96×10^{-18}	4.08×10^{-18}	4.08×10^{-18}
400×10^{-3}	$P_{Simulations}$	830×10^{-3}	1.00	990×10^{-3}	990×10^{-3}
600×10^{-3}	*M*	7.65×10^{-3}	50.8×10^{-3}	565×10^{-6}	465×10^{-6}
600×10^{-3}	*SE*	769×10^{-6}	5.10×10^{-3}	56.8×10^{-6}	46.7×10^{-6}
600×10^{-3}	$p_{Wilcoxon}$	3.96×10^{-18}	3.96×10^{-18}	462×10^{-6}	320×10^{-6}
600×10^{-3}	$P_{Simulations}$	1.00	1.00	630×10^{-3}	630×10^{-3}
800×10^{-3}	*M*	9.48×10^{-3}	17.1×10^{-3}	-13.0×10^{-6}	-35.9×10^{-6}
800×10^{-3}	*SE*	953×10^{-6}	1.72×10^{-3}	-1.31×10^{-6}	-3.61×10^{-6}
800×10^{-3}	$p_{Wilcoxon}$	3.96×10^{-18}	3.96×10^{-18}	583×10^{-3}	681×10^{-3}
800×10^{-3}	$P_{Simulations}$	1.00	1.00	590×10^{-3}	470×10^{-3}
1.00	*M*	32.9×10^{-6}	31.9×10^{-6}	23.7×10^{-6}	-64.9×10^{-6}
1.00	*SE*	3.30×10^{-6}	3.21×10^{-6}	2.38×10^{-6}	-6.52×10^{-6}
1.00	$p_{Wilcoxon}$	737×10^{-3}	646×10^{-3}	897×10^{-3}	231×10^{-3}
1.00	$P_{Simulations}$	530×10^{-3}	500×10^{-3}	480×10^{-3}	450×10^{-3}

Table C1

Detailed results for the different forgetting rates and comparisons (Unit vs. Part-Unit: ABC *vs.* BC:D *and* ABC *vs.* C:DE*; Rule-Unit vs. Class-Unit:* AGC *vs.* AGF *and* AXC *vs.* AXF*), for items presented in forward and backward order, and using the global activation as a measure of the network's familiarity with the items.* $p_{Wilcoxon}$ *represents the* p *value of a Wilcoxon test on the difference scores against the chance level of zero.* $P_{Simulations}$ *represents the proportion of simulations showing positive difference scores. (continued)*

λ_a	Statistic	ABC vs BC:D	ABC vs C:DE	AGC vs AGF	AXC vs AXF
Backward					
0	*M*	-125×10^{-3}	-82.7×10^{-3}	-79.9×10^{-3}	-74.8×10^{-3}
0	*SE*	-12.5×10^{-3}	-8.31×10^{-3}	-8.03×10^{-3}	-7.52×10^{-3}
0	$p_{Wilcoxon}$	947×10^{-3}	448×10^{-3}	286×10^{-3}	607×10^{-3}
0	$P_{Simulations}$	620×10^{-3}	560×10^{-3}	480×10^{-3}	560×10^{-3}
200×10^{-3}	*M*	9.35×10^{-3}	5.52×10^{-3}	-75.9×10^{-3}	-91.2×10^{-3}
200×10^{-3}	*SE*	940×10^{-6}	555×10^{-6}	-7.63×10^{-3}	-9.16×10^{-3}
200×10^{-3}	$p_{Wilcoxon}$	753×10^{-3}	730×10^{-3}	160×10^{-3}	92.4×10^{-3}
200×10^{-3}	$P_{Simulations}$	650×10^{-3}	580×10^{-3}	520×10^{-3}	510×10^{-3}
400×10^{-3}	*M*	111×10^{-3}	76.7×10^{-3}	14.9×10^{-3}	16.9×10^{-3}
400×10^{-3}	*SE*	11.2×10^{-3}	7.71×10^{-3}	1.50×10^{-3}	1.70×10^{-3}
400×10^{-3}	$p_{Wilcoxon}$	3.96×10^{-18}	3.96×10^{-18}	7.01×10^{-18}	3.96×10^{-18}
400×10^{-3}	$P_{Simulations}$	1.00	1.00	980×10^{-3}	1.00
600×10^{-3}	*M*	54.9×10^{-3}	32.2×10^{-3}	308×10^{-6}	536×10^{-6}
600×10^{-3}	*SE*	5.52×10^{-3}	3.23×10^{-3}	31.0×10^{-6}	53.9×10^{-6}
600×10^{-3}	$p_{Wilcoxon}$	3.96×10^{-18}	3.96×10^{-18}	239×10^{-3}	14.2×10^{-6}
600×10^{-3}	$P_{Simulations}$	1.00	1.00	550×10^{-3}	660×10^{-3}
800×10^{-3}	*M*	16.4×10^{-3}	12.8×10^{-3}	-22.4×10^{-6}	42.4×10^{-6}
800×10^{-3}	*SE*	1.65×10^{-3}	1.29×10^{-3}	-2.25×10^{-6}	4.26×10^{-6}
800×10^{-3}	$p_{Wilcoxon}$	3.96×10^{-18}	3.96×10^{-18}	985×10^{-3}	463×10^{-3}
800×10^{-3}	$P_{Simulations}$	1.00	1.00	500×10^{-3}	500×10^{-3}
1.00	*M*	-118×10^{-6}	-50.9×10^{-6}	-47.2×10^{-6}	-22.9×10^{-6}
1.00	*SE*	-11.9×10^{-6}	-5.12×10^{-6}	-4.75×10^{-6}	-2.30×10^{-6}
1.00	$p_{Wilcoxon}$	39.6×10^{-3}	278×10^{-3}	358×10^{-3}	709×10^{-3}
1.00	$P_{Simulations}$	410×10^{-3}	460×10^{-3}	490×10^{-3}	490×10^{-3}

Table C2
Detailed results for the different forgetting rates and comparisons, using the global activation as a measure of the network's familiarity with the items. $p_{Wilcoxon}$ *represents the* p *value of a Wilcoxon test on the difference scores against the chance level of zero.* $P_{Simulations}$ *represents the proportion of simulations showing positive difference scores.*

λ_a	Statistic	Unit vs BC:D	Unit vs C:DE	Phantom vs BC:D	Phantom vs C:DE	Unit vs Phantom
0	*M*	-57.8×10^{-3}	-121×10^{-3}	-49.7×10^{-3}	-91.3×10^{-3}	-38.7×10^{-3}
0	*SE*	-5.81×10^{-3}	-12.1×10^{-3}	-5.00×10^{-3}	-9.18×10^{-3}	-3.89×10^{-3}
0	$p_{Wilcoxon}$	876×10^{-3}	385×10^{-3}	865×10^{-3}	835×10^{-3}	133×10^{-3}
0	$P_{Simulations}$	540×10^{-3}	520×10^{-3}	570×10^{-3}	550×10^{-3}	450×10^{-3}
200×10^{-3}	*M*	-53.0×10^{-3}	-164×10^{-3}	-53.5×10^{-3}	-178×10^{-3}	27.6×10^{-3}
200×10^{-3}	*SE*	-5.33×10^{-3}	-16.5×10^{-3}	-5.38×10^{-3}	-17.8×10^{-3}	2.77×10^{-3}
200×10^{-3}	$p_{Wilcoxon}$	761×10^{-3}	120×10^{-3}	979×10^{-3}	111×10^{-3}	544×10^{-3}
200×10^{-3}	$P_{Simulations}$	500×10^{-3}	480×10^{-3}	590×10^{-3}	540×10^{-3}	530×10^{-3}
400×10^{-3}	*M*	76.4×10^{-3}	-27.0×10^{-3}	72.2×10^{-3}	-36.4×10^{-3}	14.3×10^{-3}
400×10^{-3}	*SE*	7.68×10^{-3}	-2.71×10^{-3}	7.25×10^{-3}	-3.66×10^{-3}	1.44×10^{-3}
400×10^{-3}	$p_{Wilcoxon}$	22.7×10^{-3}	819×10^{-3}	6.92×10^{-3}	471×10^{-3}	681×10^{-3}
400×10^{-3}	$P_{Simulations}$	640×10^{-3}	570×10^{-3}	700×10^{-3}	650×10^{-3}	450×10^{-3}
600×10^{-3}	*M*	2.06×10^{-3}	21.8×10^{-3}	2.12×10^{-3}	21.9×10^{-3}	-60.7×10^{-6}
600×10^{-3}	*SE*	207×10^{-6}	2.20×10^{-3}	214×10^{-6}	2.20×10^{-3}	-6.10×10^{-6}
600×10^{-3}	$p_{Wilcoxon}$	296×10^{-12}	3.96×10^{-18}	5.91×10^{-12}	3.96×10^{-18}	654×10^{-3}
600×10^{-3}	$P_{Simulations}$	780×10^{-3}	1.00	820×10^{-3}	1.00	500×10^{-3}
800×10^{-3}	*M*	2.12×10^{-3}	5.21×10^{-3}	2.17×10^{-3}	5.26×10^{-3}	-50.4×10^{-6}
800×10^{-3}	*SE*	213×10^{-6}	524×10^{-6}	218×10^{-6}	529×10^{-6}	-5.07×10^{-6}
800×10^{-3}	$p_{Wilcoxon}$	3.96×10^{-18}	3.96×10^{-18}	3.96×10^{-18}	3.96×10^{-18}	382×10^{-3}
800×10^{-3}	$P_{Simulations}$	1.00	1.00	1.00	1.00	480×10^{-3}
1.00	*M*	17.8×10^{-6}	17.9×10^{-6}	17.5×10^{-6}	17.7×10^{-6}	233×10^{-9}
1.00	*SE*	1.79×10^{-6}	1.80×10^{-6}	1.76×10^{-6}	1.78×10^{-6}	23.4×10^{-9}
1.00	$p_{Wilcoxon}$	5.51×10^{-18}	172×10^{-18}	2.31×10^{-15}	846×10^{-18}	849×10^{-3}
1.00	$P_{Simulations}$	980×10^{-3}	920×10^{-3}	880×10^{-3}	870×10^{-3}	490×10^{-3}

Supplementary Material D
Experiments using the activation in the test-items

Here, we report on experiments where we evaluate the network performance using the activation of only those items that are part of the the test-items instead of the global activation. That is, when an unit *ABC* was presented, we assess the network's familiarity with the items by recording the activation in *A*, *B* and *C*; in contrast, in the simulation above, we recorded the activation in *all* items. Intuitively, one would expect the results to be similar, as the active items will mainly be those that have been stimulated.

D.1 High- vs. low-TP items, tested forwards and backwards

D.1.1 Adjacent and non-adjacent forward TPs. In this section, we seek to demonstrate that the network is sensitive to basic forward TPs among and non-adjacent items. Again, to demonstrate a sensitivity to TPs among adjacent items, the network will be tested on units and part-units. Likewise, the demonstration of a sensitivity to TPs among *non*-adjacent items is inspired by the paradigm by Endress and Bonatti (2007) and will be tested on rule-units vs. class-units, either with a middle item that appear during familiarization or with a novel middle item.

As shown in Figure D1 and D2, the results are very similar to those based on the global network activation reported above: The network fails for very low and very high forgetting parameters, and successeds on all comparisons with intermediate forgetting parameters. Numerically speaking, the results are similar to those used above as well.

D.1.2 Adjacent and non-adjacent backward TPS. Again, we test the network's ability to track backward TPs by familiarizing the network with the same streams as in the previous section, but playing the test-items in reverse order (e.g., *CBA* instead of *ABC*).

As shown in Figures D3 and D4, the results are very similar to those based on the global network activation reported above: The network fails for very low and very high forgetting parameters, and successeds on all comparisons with intermediate forgetting parameters. Numerically speaking, the results are similar to those used above as well.

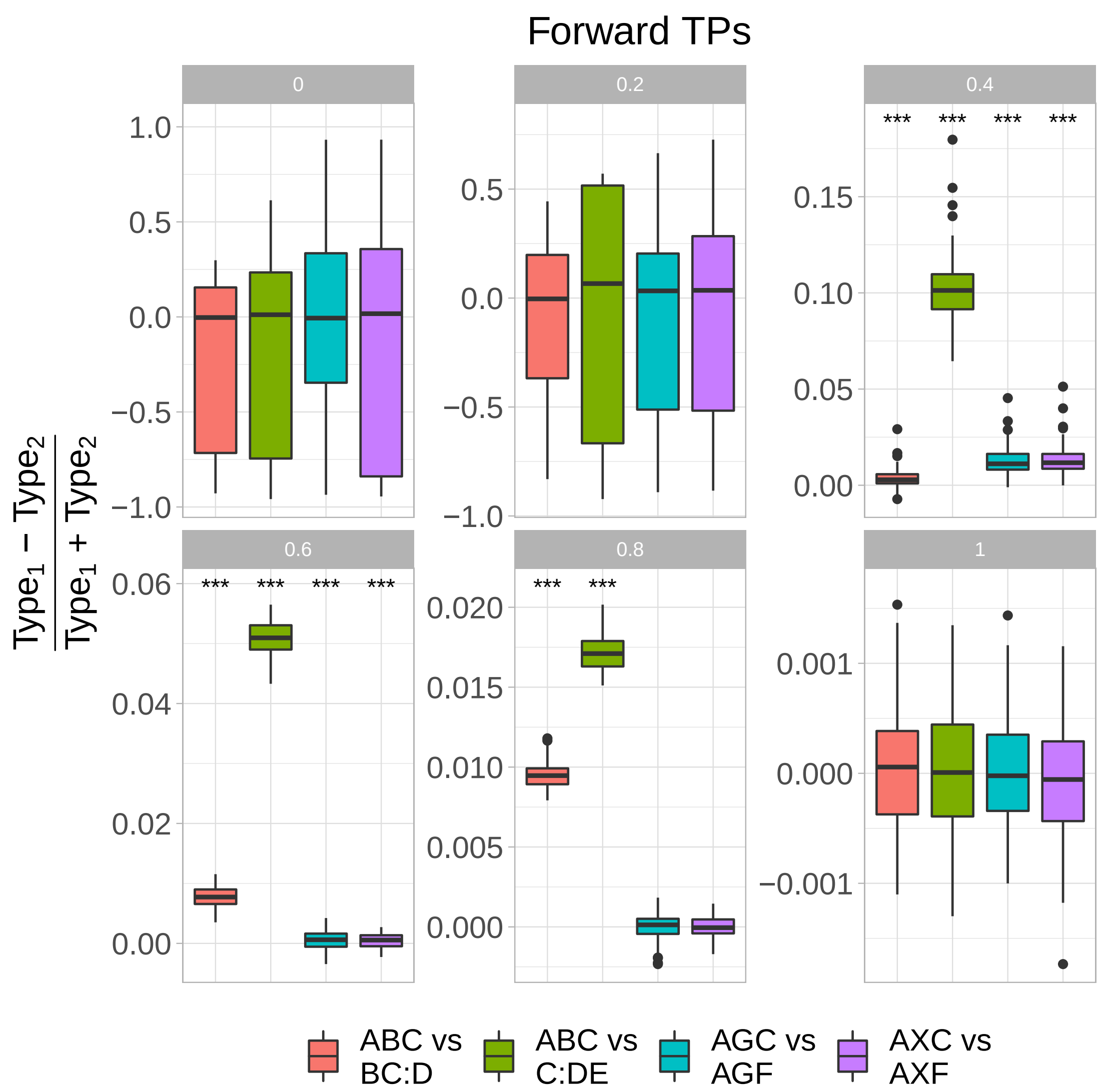

Figure D1. Difference scores for items presented in **forward order**, different forgetting rates (0, 0.2, 0.4, 0.6, 0.8 and 1), and for the different comparisons (Unit vs. Part-Unit: *ABC* vs. *BC:D* and *ABC* vs. *C:DE*; Rule-Unit vs. Class-Unit: *AGC* vs. *AGF* and *AXC* vs. *AXF*). The scores are calculated based the activation in the test items as a measure of the network's familiarity with the items. Significance is assessed based on Wilcoxon tests against the chance level of zero.

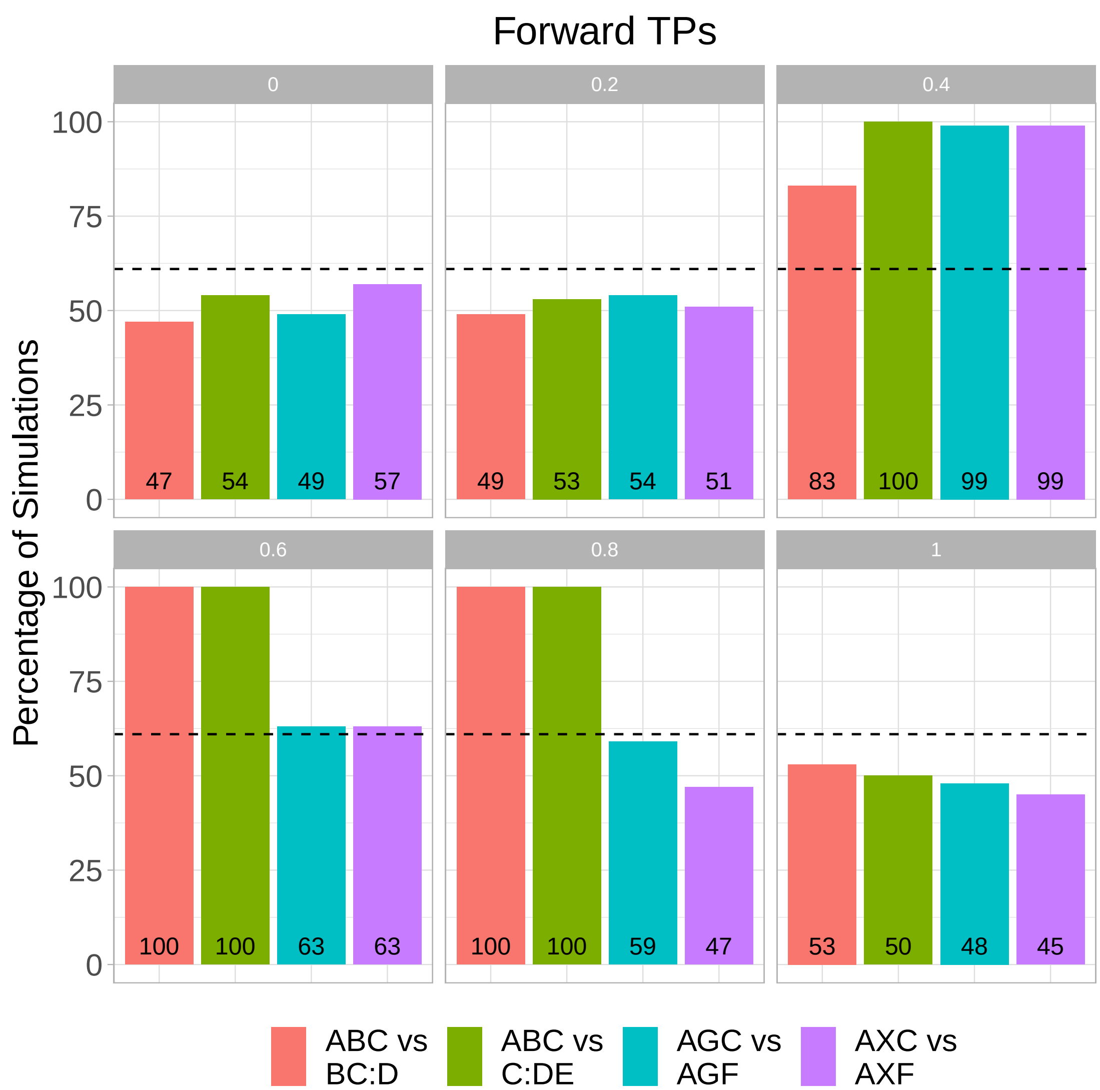

Figure D2. Percentage of simulations with a preference for the target items for items presented in **forward order**, different forgetting rates (0, 0.2, 0.4, 0.6, 0.8 and 1) and for the different comparisons (Unit vs. Part-Unit: *ABC* vs. *BC:D* and *ABC* vs. *C:DE*; Rule-Unit vs. Class-Unit: *AGC* vs. *AGF* and *AXC* vs. *AXF*). The simulations are assessed based on the activation in the test items. The dashed line shows the minimum percentage of simulations that is significant based on a binomial test.

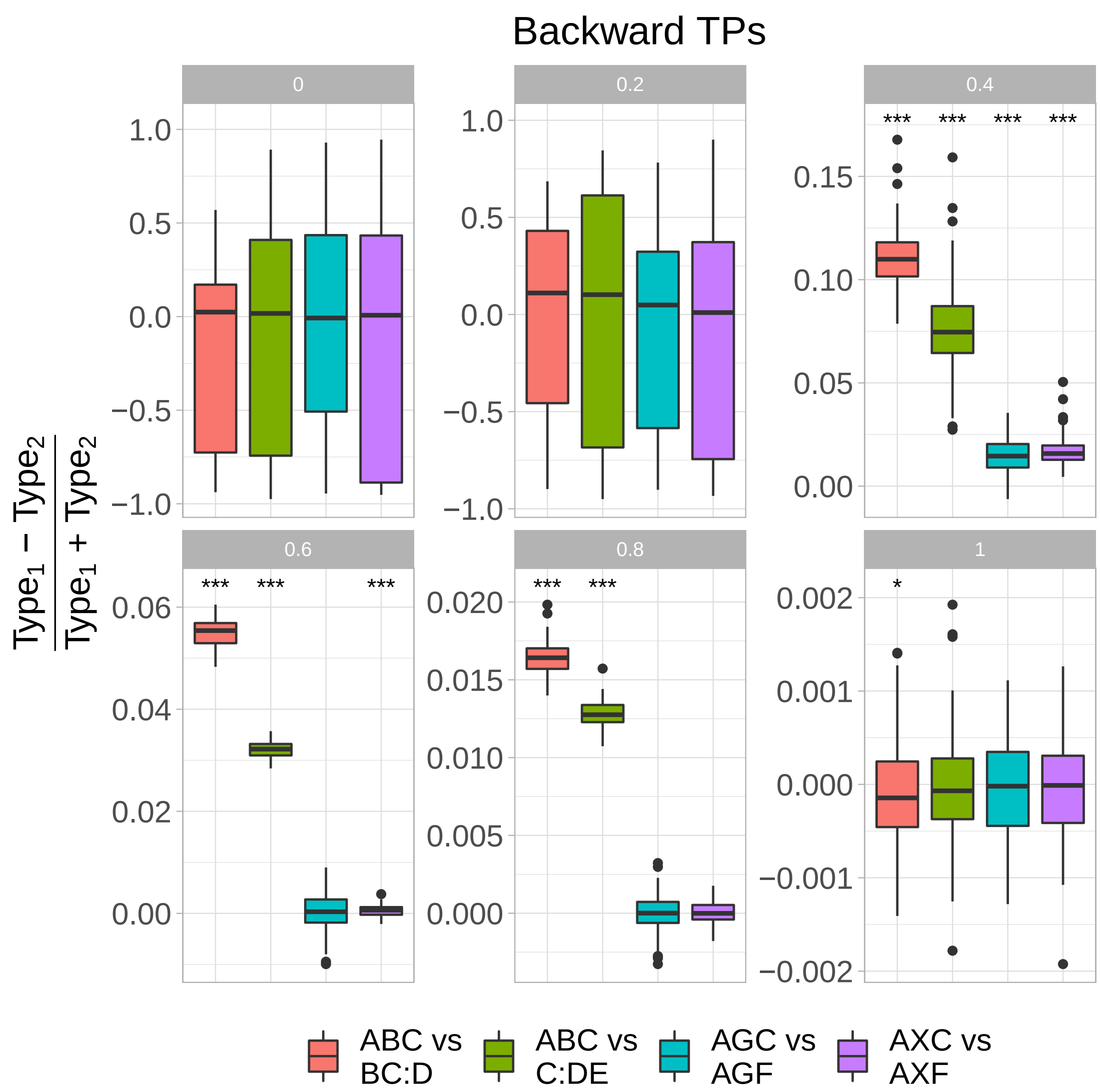

Figure D3. Difference scores for items presented in **backward order**, different forgetting rates (0, 0.2, 0.4, 0.6, 0.8 and 1), and for the different comparisons (Unit vs. Part-Unit: *ABC* vs. *BC:D* and *ABC* vs. *C:DE*; Rule-Unit vs. Class-Unit: *AGC* vs. *AGF* and *AXC* vs. *AXF*). The scores are calculated based the activation in the test items as a measure of the network's familiarity with the items. Significance is assessed based on Wilcoxon tests against the chance level of zero.

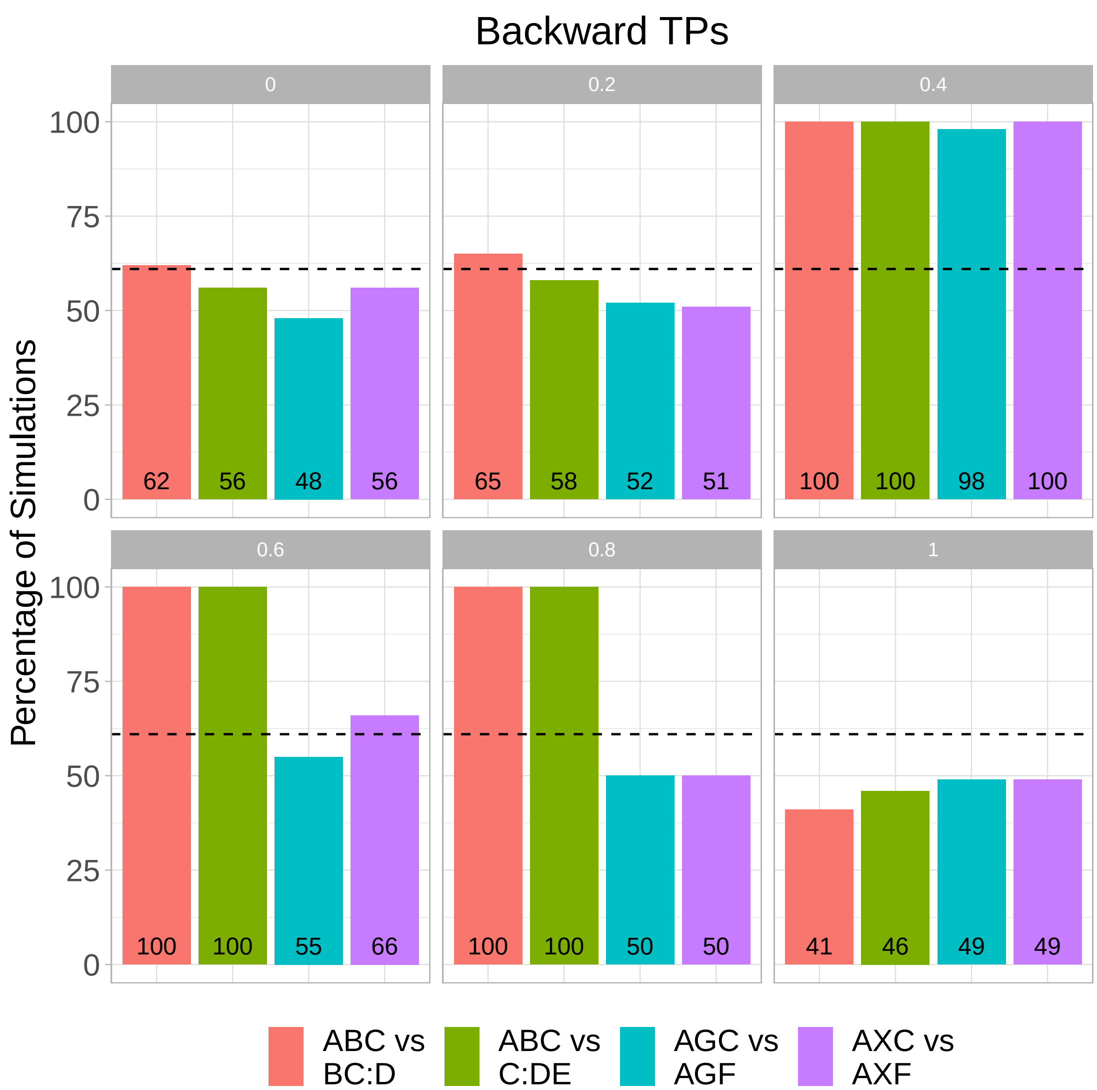

Figure D4. Percentage of simulations with a preference for the target items for items presented in **backward order**, different forgetting rates (0, 0.2, 0.4, 0.6, 0.8 and 1) and for the different comparisons (Unit vs. Part-Unit: *ABC* vs. *BC:D* and *ABC* vs. *C:DE*; Rule-Unit vs. Class-Unit: *AGC* vs. *AGF* and *AXC* vs. *AXF*). The simulations are assessed based on the activation in the test items. The dashed line shows the minimum percentage of simulations that is significant based on a binomial test.

Table D1

Detailed results for the different forgetting rates and comparisons (Unit vs. Part-Unit: ABC *vs.* BC:D *and* ABC *vs.* C:DE*; Rule-Unit vs. Class-Unit:* AGC *vs.* AGF *and* AXC *vs.* AXF*), for items presented in forward and backward order, and using the activation of the elements of the test-items as a measure of the network's familiarity with the items.* $p_{Wilcoxon}$ *represents the* p *value of a Wilcoxon test on the difference scores against the chance level of zero.* $P_{Simulations}$ *represents the proportion of simulations showing positive difference scores.*

λ_a	Statistic	ABC vs BC:D	ABC vs C:DE	AGC vs AGF	AXC vs AXF
Forward					
0	*M*	-180×10^{-3}	-113×10^{-3}	-82.7×10^{-3}	-101×10^{-3}
0	*SE*	-18.1×10^{-3}	-11.4×10^{-3}	-8.31×10^{-3}	-10.2×10^{-3}
0	$p_{Wilcoxon}$	95.7×10^{-3}	222×10^{-3}	452×10^{-3}	607×10^{-3}
0	$P_{Simulations}$	470×10^{-3}	540×10^{-3}	490×10^{-3}	570×10^{-3}
200×10^{-3}	*M*	-109×10^{-3}	-72.8×10^{-3}	-92.6×10^{-3}	-87.1×10^{-3}
200×10^{-3}	*SE*	-11.0×10^{-3}	-7.32×10^{-3}	-9.31×10^{-3}	-8.75×10^{-3}
200×10^{-3}	$p_{Wilcoxon}$	120×10^{-3}	118×10^{-3}	152×10^{-3}	134×10^{-3}
200×10^{-3}	$P_{Simulations}$	490×10^{-3}	530×10^{-3}	540×10^{-3}	510×10^{-3}
400×10^{-3}	*M*	3.68×10^{-3}	102×10^{-3}	12.4×10^{-3}	13.2×10^{-3}
400×10^{-3}	*SE*	369×10^{-6}	10.2×10^{-3}	1.25×10^{-3}	1.33×10^{-3}
400×10^{-3}	$p_{Wilcoxon}$	2.92×10^{-12}	3.96×10^{-18}	4.08×10^{-18}	4.08×10^{-18}
400×10^{-3}	$P_{Simulations}$	830×10^{-3}	1.00	990×10^{-3}	990×10^{-3}
600×10^{-3}	*M*	7.65×10^{-3}	50.8×10^{-3}	565×10^{-6}	465×10^{-6}
600×10^{-3}	*SE*	769×10^{-6}	5.10×10^{-3}	56.8×10^{-6}	46.7×10^{-6}
600×10^{-3}	$p_{Wilcoxon}$	3.96×10^{-18}	3.96×10^{-18}	462×10^{-6}	320×10^{-6}
600×10^{-3}	$P_{Simulations}$	1.00	1.00	630×10^{-3}	630×10^{-3}
800×10^{-3}	*M*	9.48×10^{-3}	17.1×10^{-3}	-13.0×10^{-6}	-35.9×10^{-6}
800×10^{-3}	*SE*	953×10^{-6}	1.72×10^{-3}	-1.31×10^{-6}	-3.61×10^{-6}
800×10^{-3}	$p_{Wilcoxon}$	3.96×10^{-18}	3.96×10^{-18}	583×10^{-3}	681×10^{-3}
800×10^{-3}	$P_{Simulations}$	1.00	1.00	590×10^{-3}	470×10^{-3}
1.00	*M*	32.9×10^{-6}	31.9×10^{-6}	23.7×10^{-6}	-64.9×10^{-6}
1.00	*SE*	3.30×10^{-6}	3.21×10^{-6}	2.38×10^{-6}	-6.52×10^{-6}
1.00	$p_{Wilcoxon}$	737×10^{-3}	646×10^{-3}	897×10^{-3}	231×10^{-3}
1.00	$P_{Simulations}$	530×10^{-3}	500×10^{-3}	480×10^{-3}	450×10^{-3}
Backward					
0	*M*	-125×10^{-3}	-82.7×10^{-3}	-79.9×10^{-3}	-74.8×10^{-3}
0	*SE*	-12.5×10^{-3}	-8.31×10^{-3}	-8.03×10^{-3}	-7.52×10^{-3}
0	$p_{Wilcoxon}$	947×10^{-3}	448×10^{-3}	286×10^{-3}	607×10^{-3}
0	$P_{Simulations}$	620×10^{-3}	560×10^{-3}	480×10^{-3}	560×10^{-3}
200×10^{-3}	*M*	9.35×10^{-3}	5.52×10^{-3}	-75.9×10^{-3}	-91.2×10^{-3}
200×10^{-3}	*SE*	940×10^{-6}	555×10^{-6}	-7.63×10^{-3}	-9.16×10^{-3}

Table D1

Detailed results for the different forgetting rates and comparisons (Unit vs. Part-Unit: ABC *vs.* BC:D *and* ABC *vs.* C:DE*; Rule-Unit vs. Class-Unit:* AGC *vs.* AGF *and* AXC *vs.* AXF*), for items presented in forward and backward order, and using the activation of the elements of the test-items as a measure of the network's familiarity with the items.* $p_{Wilcoxon}$ *represents the* p *value of a Wilcoxon test on the difference scores against the chance level of zero.* $P_{Simulations}$ *represents the proportion of simulations showing positive difference scores. (continued)*

λ_a	Statistic	ABC vs BC:D	ABC vs C:DE	AGC vs AGF	AXC vs AXF
200×10^{-3}	$p_{Wilcoxon}$	753×10^{-3}	730×10^{-3}	160×10^{-3}	92.4×10^{-3}
200×10^{-3}	$P_{Simulations}$	650×10^{-3}	580×10^{-3}	520×10^{-3}	510×10^{-3}
400×10^{-3}	*M*	111×10^{-3}	76.7×10^{-3}	14.9×10^{-3}	16.9×10^{-3}
400×10^{-3}	*SE*	11.2×10^{-3}	7.71×10^{-3}	1.50×10^{-3}	1.70×10^{-3}
400×10^{-3}	$p_{Wilcoxon}$	3.96×10^{-18}	3.96×10^{-18}	7.01×10^{-18}	3.96×10^{-18}
400×10^{-3}	$P_{Simulations}$	1.00	1.00	980×10^{-3}	1.00
600×10^{-3}	*M*	54.9×10^{-3}	32.2×10^{-3}	308×10^{-6}	536×10^{-6}
600×10^{-3}	*SE*	5.52×10^{-3}	3.23×10^{-3}	31.0×10^{-6}	53.9×10^{-6}
600×10^{-3}	$p_{Wilcoxon}$	3.96×10^{-18}	3.96×10^{-18}	239×10^{-3}	14.2×10^{-6}
600×10^{-3}	$P_{Simulations}$	1.00	1.00	550×10^{-3}	660×10^{-3}
800×10^{-3}	*M*	16.4×10^{-3}	12.8×10^{-3}	-22.4×10^{-6}	42.4×10^{-6}
800×10^{-3}	*SE*	1.65×10^{-3}	1.29×10^{-3}	-2.25×10^{-6}	4.26×10^{-6}
800×10^{-3}	$p_{Wilcoxon}$	3.96×10^{-18}	3.96×10^{-18}	985×10^{-3}	463×10^{-3}
800×10^{-3}	$P_{Simulations}$	1.00	1.00	500×10^{-3}	500×10^{-3}
1.00	*M*	-118×10^{-6}	-50.9×10^{-6}	-47.2×10^{-6}	-22.9×10^{-6}
1.00	*SE*	-11.9×10^{-6}	-5.12×10^{-6}	-4.75×10^{-6}	-2.30×10^{-6}
1.00	$p_{Wilcoxon}$	39.6×10^{-3}	278×10^{-3}	358×10^{-3}	709×10^{-3}
1.00	$P_{Simulations}$	410×10^{-3}	460×10^{-3}	490×10^{-3}	490×10^{-3}

D.2 The role of frequency of occurrence

As mentioned above, the experiments presented so far confound TPs and frequency of occurrence: Units do not only have stronger TPs than part-units, but they also occur more frequently. Among the control experiments for this issue (Aslin, Saffran, & Newport, 1998; Endress & Mehler, 2009; Endress & Langus, 2017), our computational experiments are inspired by Endress and Mehler (2009) and Endress and Langus (2017). We thus expose the network to a six unit stream inspired by Endress and Mehler (2009) and Endress and Langus (2017). Following this, we test the network on units, phantom-units and part-units.

As shown in Figure D5 and D6, the results are very similar to those based on the global network activation reported above: The network fails for very low and very high forgetting parameters, and prefers units and phantom-units over part-units roughly

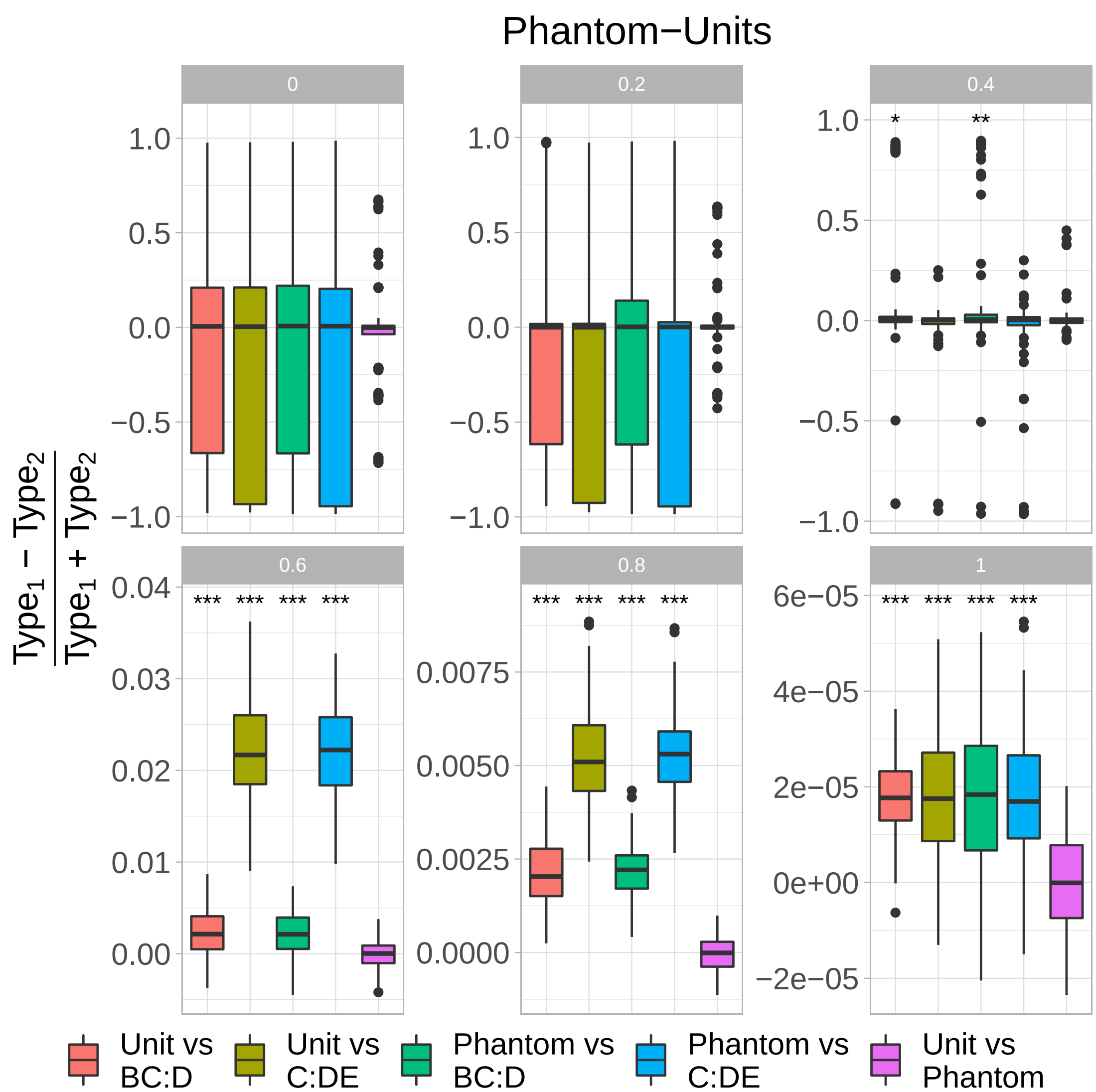

Figure D5. Difference scores for items presented in **forward order**, different forgetting rates (0, 0.2, 0.4, 0.6, 0.8 and 1), and for the different comparisons (Unit vs. Part-Unit: *ABC* vs. *BC:D* and *ABC* vs. *C:DE*; Phantom-Unit vs. Part-Unit: Phantom-Unit vs. *BC:D* and Phantom-Unit vs. *C:DE*; Unit vs. Phantom-Unit). The scores are calculated based the activation in the test items as a measure of the network's familiarity with the items. Significance is assessed based on Wilcoxon tests against the chance level of zero.

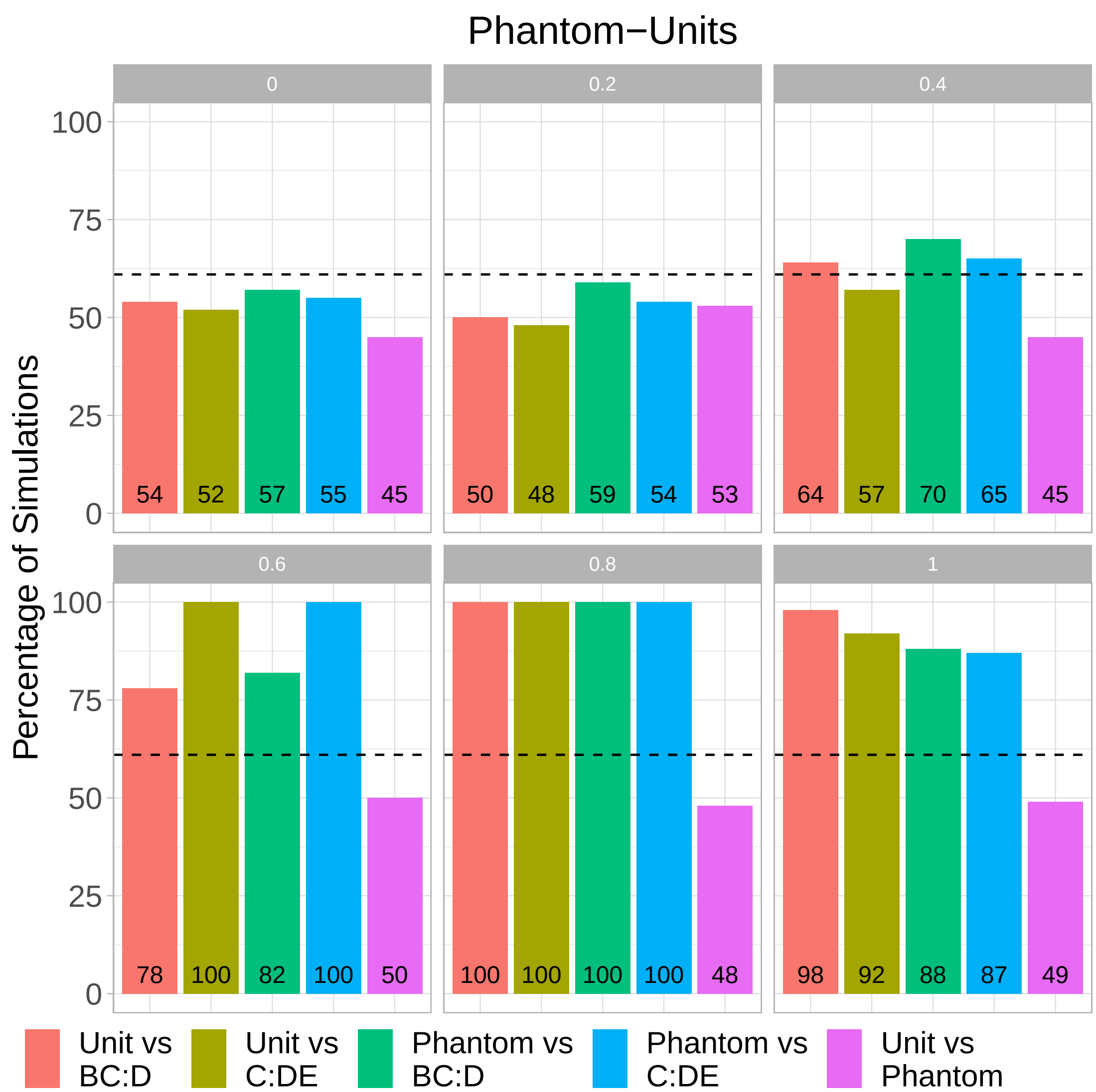

Figure D6. Percentage of simulations with a preference for the target items for items presented in **forward order**, different forgetting rates (0, 0.2, 0.4, 0.6, 0.8 and 1) and for the different comparisons (Unit vs. Part-Unit: *ABC* vs. *BC:D* and *ABC* vs. *C:DE*; Phantom-Unit vs. Part-Unit: Phantom-Unit vs. *BC:D* and Phantom-Unit vs. *C:DE*; Unit vs. Phantom-Unit). The simulations are assessed based on the activation in the test items. The dashed line shows the minimum percentage of simulations that is significant based on a binomial test.

to the same extent for medium and high forgetting rates. As in Endress and Mehler (2009) and Endress and Langus (2017), it thus more sensitive to differences in TPs than to differences in frequency of occurrence. In contrast, the network does not seem to discriminate between units and phantom-units, replicating Endress and Mehler's (2009) and Endress and Langus's (2017) results.

D.3 Detailed results

Table D1 provides detailed results for the simulations in terms of descriptive statistics and statistical tests for the simulation testing the recognition of (forward and backward) units, part-units, rule-units and class-units.

Table D2 provides similar results for the simulations testing the recognition of units, part-units and phantom-units.

Table D2

Detailed results for the different forgetting rates and comparisons, and using the activation of the elements of the test-items as a measure of the network's familiarity with the items. $p_{Wilcoxon}$ *represents the* p *value of a Wilcoxon test on the difference scores against the chance level of zero.* $P_{Simulations}$ *represents the proportion of simulations showing positive difference scores.*

λ_a	Statistic	Unit vs BC:D	Unit vs C:DE	Phantom vs BC:D	Phantom vs C:DE	Unit vs Phantom
0	*M*	-57.8×10^{-3}	-121×10^{-3}	-49.7×10^{-3}	-91.3×10^{-3}	-38.7×10^{-3}
0	*SE*	-5.81×10^{-3}	-12.1×10^{-3}	-5.00×10^{-3}	-9.18×10^{-3}	-3.89×10^{-3}
0	$p_{Wilcoxon}$	876×10^{-3}	385×10^{-3}	865×10^{-3}	835×10^{-3}	133×10^{-3}
0	$P_{Simulations}$	540×10^{-3}	520×10^{-3}	570×10^{-3}	550×10^{-3}	450×10^{-3}
200×10^{-3}	*M*	-53.0×10^{-3}	-164×10^{-3}	-53.5×10^{-3}	-178×10^{-3}	27.6×10^{-3}
200×10^{-3}	*SE*	-5.33×10^{-3}	-16.5×10^{-3}	-5.38×10^{-3}	-17.8×10^{-3}	2.77×10^{-3}
200×10^{-3}	$p_{Wilcoxon}$	761×10^{-3}	120×10^{-3}	979×10^{-3}	111×10^{-3}	544×10^{-3}
200×10^{-3}	$P_{Simulations}$	500×10^{-3}	480×10^{-3}	590×10^{-3}	540×10^{-3}	530×10^{-3}
400×10^{-3}	*M*	76.4×10^{-3}	-27.0×10^{-3}	72.2×10^{-3}	-36.4×10^{-3}	14.3×10^{-3}
400×10^{-3}	*SE*	7.68×10^{-3}	-2.71×10^{-3}	7.25×10^{-3}	-3.66×10^{-3}	1.44×10^{-3}
400×10^{-3}	$p_{Wilcoxon}$	22.7×10^{-3}	819×10^{-3}	6.92×10^{-3}	471×10^{-3}	681×10^{-3}
400×10^{-3}	$P_{Simulations}$	640×10^{-3}	570×10^{-3}	700×10^{-3}	650×10^{-3}	450×10^{-3}
600×10^{-3}	*M*	2.06×10^{-3}	21.8×10^{-3}	2.12×10^{-3}	21.9×10^{-3}	-60.7×10^{-6}
600×10^{-3}	*SE*	207×10^{-6}	2.20×10^{-3}	214×10^{-6}	2.20×10^{-3}	-6.10×10^{-6}
600×10^{-3}	$p_{Wilcoxon}$	296×10^{-12}	3.96×10^{-18}	5.91×10^{-12}	3.96×10^{-18}	654×10^{-3}
600×10^{-3}	$P_{Simulations}$	780×10^{-3}	1.00	820×10^{-3}	1.00	500×10^{-3}
800×10^{-3}	*M*	2.12×10^{-3}	5.21×10^{-3}	2.17×10^{-3}	5.26×10^{-3}	-50.4×10^{-6}
800×10^{-3}	*SE*	213×10^{-6}	524×10^{-6}	218×10^{-6}	529×10^{-6}	-5.07×10^{-6}
800×10^{-3}	$p_{Wilcoxon}$	3.96×10^{-18}	3.96×10^{-18}	3.96×10^{-18}	3.96×10^{-18}	382×10^{-3}
800×10^{-3}	$P_{Simulations}$	1.00	1.00	1.00	1.00	480×10^{-3}
1.00	*M*	17.8×10^{-6}	17.9×10^{-6}	17.5×10^{-6}	17.7×10^{-6}	233×10^{-9}
1.00	*SE*	1.79×10^{-6}	1.80×10^{-6}	1.76×10^{-6}	1.78×10^{-6}	23.4×10^{-9}
1.00	$p_{Wilcoxon}$	5.51×10^{-18}	172×10^{-18}	2.31×10^{-15}	846×10^{-18}	849×10^{-3}
1.00	$P_{Simulations}$	980×10^{-3}	920×10^{-3}	880×10^{-3}	870×10^{-3}	490×10^{-3}

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*, 321–324.

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*(2), 247–299.

Endress, A. D., & Langus, A. (2017). Transitional probabilities count more than frequency, but might not be used for memorization. *Cognitive Psychology*, *92*, 37–64. doi: 10.1016/j.cogpsych.2016.11.004

Endress, A. D., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language*, *60*(3), 351-367.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review*, *107*(2), 289-344.

Letzkus, J. J., Wolff, S. B. E., Meyer, E. M. M., Tovote, P., Courtin, J., Herry, C., & Lüthi, A. (2011). A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature*, *480*, 331–335. doi: 10.1038/nature10674