Citation: Endress, A. (2024). Hebbian learning can explain rhythmic neural entrainment to statistical regularities. Developmental Science, 27(4), e13487. doi: 10.1111/desc.13487

Received: 28 April 2023 | Revised: 26 December 2023 | Accepted: 29 January 2024

DOI: 10.1111/desc.13487

Developmental Science WILEY

SHORT REPORT

Hebbian learning can explain rhythmic neural entrainment to statistical regularities

Ansgar D. Endress

Department of Psychology, City, University of London, London, UK

Correspondence
Ansgar D. Endress, Department of Psychology, City, University of London, Northampton Square, London EC1V 0HB, UK.
Email: ansgar.endress@m4x.org

Abstract

In many domains, learners extract recurring units from continuous sequences. For example, in unknown languages, fluent speech is perceived as a continuous signal. Learners need to extract the underlying words from this continuous signal and then memorize them. One prominent candidate mechanism is statistical learning, whereby learners track how predictive syllables (or other items) are of one another. Syllables within the same word predict each other better than syllables straddling word boundaries. But does statistical learning lead to memories of the underlying words—or just to pairwise associations among syllables? Electrophysiological results provide the strongest evidence for the memory view. Electrophysiological responses can be time-locked to statistical word boundaries (e.g., N400s) and show rhythmic activity with a periodicity of word durations. Here, I reproduce such results with a simple Hebbian network. When exposed to statistically structured syllable sequences (and when the underlying words are not excessively long), the network activation is rhythmic with the periodicity of a word duration and activation maxima on word-final syllables. This is because word-final syllables receive more excitation from earlier syllables with which they are associated than less predictable syllables that occur earlier in words. The network is also sensitive to information whose electrophysiological correlates were used to support the encoding of ordinal positions within words. Hebbian learning can thus explain rhythmic neural activity in statistical learning tasks without any memory representations of words. Learners might thus need to rely on cues beyond statistical associations to learn the words of their native language.

KEYWORDS
implicit learning, N400, neural entrainment, neural networks, statistical learning, transitional probabilities

Research Highlights

- Statistical learning may be utilized to identify recurring units in continuous sequences (e.g., words in fluent speech) but may not generate explicit memory for words.

- Exposure to statistically structured sequences leads to rhythmic activity with a period of the duration of the underlying units (e.g., words).
- I show that a memory-less Hebbian network model can reproduce this rhythmic neural activity as well as putative encodings of ordinal positions observed in earlier research.
- Direct tests are needed to establish whether statistical learning leads to declarative memories for words.

1 | INTRODUCTION

Word learning is challenging even when the phonological form of words is known (Gillette et al., 1999; Medina et al., 2011). However, speech in unknown languages often appears as a continuous signal with few cues to word boundaries (but see e.g., Brentari et al., 2011; Christophe et al., 2001; Endress & Hauser, 2010; Johnson & Jusczyk, 2001; Johnson & Seidl, 2009; Pilon, 1981). As a result, learners first need to discover where words start and where they end before they can commit any phonological word form to memory (e.g., Aslin et al., 1998; Saffran, Aslin et al., 1996; Saffran, Newport et al., 1996), and hopefully link this word form to some meaning. This challenge is called the segmentation problem.

Learners might solve the segmentation problem using co-occurrence statistics tracking the predictability of syllables. For example, a syllable following “the” is much harder to predict than a syllable following “whis.” After all, “the” can precede any noun, but there are very few words starting with “whis” (e.g., whiskey, whisker,…).

The most prominent version of such co-occurrence statistics involves transitional probabilities (TPs), that is, the conditional probability of a syllable σ_2 following another syllable σ_1, $P(\sigma_2|\sigma_1)$. Infants, newborns and nonhuman animals are all sensitive to TPs in a variety of modalities, for stimuli that can be sequentially or spatially organized (e.g., Aslin et al., 1998; Batterink & Paller, 2017; Chen & Ten Cate, 2015; Conway & Christiansen, 2005; Creel et al., 2004; Endress, 2010; Endress & Wood, 2011; Fiser & Aslin, 2002a, 2005; Fló et al., 2022; Glicksohn & Cohen, 2011; Hauser et al., 2001; Kirkham et al., 2002; Saffran, Newport et al., 1996; Saffran, Aslin et al., 1996; Saffran et al., 1999; Saffran & Griepentrog, 2001; Sohail & Johnson, 2016; Tompson et al., 2019; Toro et al., 2005; Turk-Browne et al., 2005; Turk-Browne & Scholl, 2009). While the ubiquity of these statistical learning abilities probably suggests that organisms have independent statistical learning abilities in different domains and brain areas (see Frost et al., 2015 vs. Endress, 2019, for different versions of this idea), I focus on sequential verbal regularities.

Following Saffran and colleagues' seminal work (Aslin et al., 1998; Saffran, Aslin et al., 1996; Saffran, Newport et al., 1996), participants in a typical verbal statistical learning experiment are first familiarized with a statistically structured speech stream (or a sequence in another modality such as auditory tones or visual symbols). The speech stream is a random concatenation of triplets of nonsense syllables (hereafter “words”). Syllables within words are, thus, more predictive of one another than syllable across word-boundaries. For example, if *ABC*, *DEF*, *GHJ*, and *KLM* are “words” (where each letter represents a syllable), the *C* syllable can be followed by the word-initial syllables of any of the other words, while syllables within words predict each other with certainty.

A sensitivity to TPs is then tested by measuring a preference between high-TP items (i.e., words) and low-TP items created by taking either the final syllable of one word and the first two syllables from another word (e.g., *CDE*) or by taking the last two syllables of one word and the first syllable of the next word (e.g., *BCD*); the low-TP items are called part-words. Participants (adults, infants, or other animals) usually discriminate between words and part-words, suggesting that they are sensitive to TPs.

1.1 | Does statistical learning help learners memorizing words?

While many authors propose that tracking TPs leads to the addition of words to the mental lexicon (and thus to storage of word candidates in declarative long-term memory, LTM; e.g., Erickson et al., 2014; Graf-Estes et al., 2007; Hay et al., 2011; Isbilen et al., 2020; Karaman & Hay, 2018; Perruchet, 2019; Shoaib et al., 2018), this issue is controversial, and the results supporting such views often have alternative explanations that do not involve declarative LTM (see Endress & de Seyssel, under review; Endress et al., 2020, for critical reviews). For example, while high-TP items are sometimes easier to memorize than low-TP items (e.g., Graf-Estes et al., 2007; Hay et al., 2011; Isbilen et al., 2020; Karaman & Hay, 2018), it is unclear if any LTM representation have been formed during statistical learning, or whether statistical associations simply facilitate subsequent associations. Likewise, while incomplete high-TP items are sometimes harder to recognize than entire items (e.g., an *AB* subunit from an *ABC* unit is harder to recognize that an entire *DE* unit; e.g., Fiser & Aslin, 2005; Giroux & Rey, 2009; Orbán et al., 2008; Slone & Johnson, 2018), such results can be explained by memory-less Hebbian learning mechanisms and other attentional accounts (Endress & de Seyssel, under review).

Critically, to the extent that a sensitivity to TPs relies on implicit learning mechanisms (e.g., Christiansen, 2018; Perruchet & Pacton, 2006), statistical learning might be dissociable from explicit, declarative memory (e.g., Cohen & Squire, 1980; Finn et al., 2016; Graf & Mandler, 1984; Knowlton et al., 1996; Poldrack et al., 2001; Sherman & Turk-Browne, 2020; Squire, 1992). In line with this view, tasks that are typically used in the declarative memory literature (e.g., familiarity ratings, confidence ratings, or choice tasks) are *not* diagnostic of the addition of statistically defined items to the mental lexicon (see e.g., Batterink, 2020; Batterink & Paller, 2017; Batterink et al., 2015, for experiments contrasting implicit and explicit measures of statistical learning). For example, observers sometimes prefer high-TP items to low-TP items when they have never encountered either of them (when the items are played backwards compared to the familiarization stream; Endress & Wood, 2011; Jones & Pashler, 2007; Turk-Browne & Scholl, 2009), and sometimes prefer high-TP items they have never encountered over low-TP items they have heard or seen (Endress & Langus, 2017; Endress & Mehler, 2009b). In such cases, a preference for high-TP items does not indicate that the high-TP items are stored in the mental lexicon, simply because learners have never encountered these items.[1]

The reason for which, in the case of statistical learning, traditional declarative LTM tasks might be misleading is a mismatch between the memory tasks and the learning problem. After all, learners do not only need to remember which item occur with which other items (which can be assessed with explicit memory tasks); they also need to create novel memory representations of those co-occurring items, and these memory representations need to be stored as integrated units (e.g., Batterink & Paller, 2017; Endress & Langus, 2017; Fló et al., 2022; Slone & Johnson, 2018). For example, after exposure to a Bugs Bunny cartoon, viewers are presumably highly confident that Bugs Bunny co-occurred more frequently with a carrot than with a gun (potentially by using episodic memory). However, such associations do not imply that the Bugs Bunny–carrot combination is stored as an integrated chunk in declarative LTM.

In a direct test of the contents of LTM after a statistical learning task, Endress and de Seyssel (under review) asked adults to repeat back the items they had encountered during a familiarization stream with as few as four items. While few participants produced words or part-words, two thirds of those who did produce such items produced exclusively *part-words* rather than words (see also Batterink, 2020, who reported in their methods section that most participants failed to recall any words in a free recall task). The reason for this counterintuitive behavior is that learners successfully track TPs in the speech stream, but do not know where words start and where they end. They thus initiate their production with a random syllable and continue with a high probability continuation. Given that, in these speech streams, two thirds of the syllables are word-medial or word-final, but only one third is word-initial, two thirds of the productions should be part-words, which is just what Endress and de Seyssel (under review) found. Given that participants do not know where words start and where they end, they cannot have declarative memories of words either.

These results do not imply that different memory mechanisms cannot interact. Procedural and declarative memory mechanisms can interact during consolidation (Robertson, 2022), and statistical information might facilitate subsequent declarative memory formation. However, there is no behavioral evidence that statistical learning per se leads to LTM of the underlying units.

Be that as it may, there is a straightforward explanation of such results that does not involve declarative LTM: a sensitivity to TPs might reflect Hebbian learning (Endress, 2010; Endress & Johnson, 2021). After all, the representations of syllables (or other elements in a stream) presumably does not cease to be active as soon as the syllable ends. As a result, multiple syllables are active together and can thus form Hebbian associations. Endress and Johnson (2021) showed that such a network can account for a number of behavioral statistical learning results.

However, there is another class of studies that seems to be inconsistent with a mere Hebbian interpretation of statistical learning results, and that seems to strongly support the possibility that statistical learning leads to the extraction of coherent units: electrophysiological responses to statistically structured sequences. I will now turn to this literature.

1.2 | Electrophysiological correlates of statistical learning

In one of the earliest electrophysiological studies of statistical learning, Sanders et al. (2002) first presented participants with a speech stream composed of non-sense words. Following this, they presented these words in isolation, and finally another speech stream with the same words. When they compared electrical brain responses to the second presentation of the stream and to its first presentation, they observed increased N100 and N400 responses. That is, they showed increased negativities around 100 and 400 ms after word onsets (see also Abla et al., 2008 for similar study with tones as stimuli). Cunillera et al. (2006) showed that N400 effects can also be obtained without explicitly training participants on the words, and even newborns show time-locked responses to statistically structured sequences (Kudo et al., 2011; Teinonen et al., 2009).

Following Buiatti et al. (2009), electrophysiological investigations of statistical learning focused on rhythmic entrainment to the speech streams rather than event-related responses such as the N400s. Specifically, if listeners learn the statistical structure of the speech stream, they should perceive the speech stream as a sequence of trisyllabic units (given that most statistical learning experiments tend to use tri-syllabic units or their equivalents in other domains, but see Benjamin, Fló, Palu et al., 2023; Johnson & Tyler, 2010), and thus perceive a rhythm with a periodicity of three syllable durations. If so, they should also show a *neural* rhythm with the same periodicity. While Buiatti et al. (2009) detected such a rhythm only when words were separated by brief silences, later investigations found such rhythms in continuous sequences in adults (Batterink, 2020; Batterink & Paller, 2017; see also

Moser et al., 2021 for a magneto-encephalography study and Henin et al., 2021; Sherman et al., 2023 for intracranial recordings), children (Moreau et al., 2022), infants (Choi et al., 2020; Kabdebon et al., 2015), and even newborns (Fló et al., 2022).

Such results seem to strongly suggest that statistical learning creates integrated units that can be stored in memory, though different authors stressed that rhythmic entrainment might also reflect processes that are separable from integrated memory encodings of words (e.g., Batterink & Paller, 2017; Fló et al., 2022). However, there is a simple associative explanation for this rhythmic activity. After the associations in a word *ABC* are learned (where each letter stands for a syllable), each syllable predicts subsequent syllables. Hence, the *C* syllable does not only receive (external) bottom-up excitation when it is heard, but receives additional associative excitation from the preceding *A* and *B* syllables (that predict the *C* syllable). As a result, one would expect a neural rhythm with a period of three syllable durations, and a maximum following the onset of the *word-final* syllable even if no word has been stored in memory.[2]

This account is also consistent with the original interpretation of the N400 component as reflecting (semantically) surprising and thus unpredictable stimuli (Kutas & Federmeier, 2000). In verbal statistical learning tasks, word onsets are always unpredictable, given that words are randomly concatenated. In contrast, the last syllable of each word is predictable based on the statistical structure of the streams, but only after learning. As a result, electrophysiological responses such as N400s might not so much index word onsets as reflect the increased predictability of word-final syllables (or the decreased relative predictability of word-initial syllables). As regularly occurring N400s (or other regular time-locked responses) create rhythmic activity, rhythmic entrainment to the statistical structure of a speech stream might similarly reflect rhythmic changes in the relative predictability of items rather than memorization of stimuli.

2 | THE CURRENT STUDY

Here, I provide computational support for this idea. While different processing models (e.g., Batchelder, 2002; Christiansen et al., 1998; Perruchet & Vinter, 1998) and ideal observer models (e.g., Benjamin, Fló, Al Roumi et al., 2023; Brent & Cartwright, 1996; Frank et al., 2010; Lynn et al., 2020; Orbán et al., 2008) have been dedicated to explain statistical learning results, here I attempt to show that very simple psychological mechanisms based on Hebbian learning can explain electrophysiological results on top of a variety of other behavioral results. I also show that the mechanisms provide an alternative to some more abstract coding schemes proposed by Henin et al. (2021).

The network used here is a fairly generic saliency map (e.g., Bays et al., 2010; Endress & Szabó, 2020; Gottlieb, 2007; Roggeman et al., 2010; Sengupta et al., 2014) augmented by a Hebbian learning component. The network comprises units representing populations of neurons encoding syllables (or other items). All units are fully connected with both excitatory and inhibitory connections (see Figure 1). Excitatory connections change according to a Hebbian learning rule, while inhibitory connections do not undergo learning. Additionally, activation decays exponentially in all units.[3] Further details of the model can be found in Supplementary Information A.

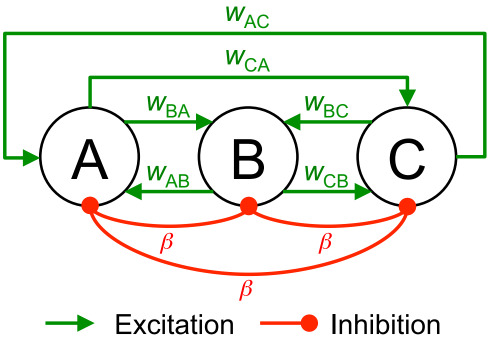

FIGURE 1 Illustration of the network architecture with only three units *A*, *B*, and *C*. These units encode syllables. All units mutually excite and inhibit one another. Excitatory connections undergo Hebbian learning. For example, unit *A* excites unit *B* with a tunable weight of w_{BA} as well as unit *C* with a weight of w_{CA}. In contrast, inhibitory weights do not undergo learning. In addition to excitation and inhibition, all units undergo forgetting.

Such an architecture can explain statistical learning results in a relatively intuitive way. If each syllable is represented by some population of neurons, and learners listen to some sequence *ABCD*…, associations should form between adjacent and nonadjacent syllables depending on the decay rate. If activation decay is slower than a syllable duration, the representations of two adjacent syllables will be active at the same time, and thus form an association. For example, if a neuron representing *A* is still active while *B* is presented, these neurons can form an association. Similarly, if a neuron representing *A* is still active when *C* is presented, an association between these neurons will ensue although the corresponding syllables are not adjacent.

Endress and Johnson (2021) showed that such a model can account for a number of statistical learning results (as long as the decay rate was set to a reasonable level)—in the absence of a dedicated memory store. Hence, statistical learning results can be explained even when participants do not create lexical entries for high-TP items.

However, the neural entrainment results above seem to suggest that learners go beyond mere associations among syllables, and extract statistically coherent units. Here, I argue that this simple Hebbian network can also account for the periodic activity found in electrophysiological recordings. Intuitively, if a high-TP item such as *ABC* is presented, *A* mostly receives external stimulation, but *B* receives external stimulation—as well as excitatory input from *A*. Likewise, *C* receives external stimulation as well as excitatory input from both *A* and *B*. As a result, the network activation should increase towards the end of a word, with a maximum on the third syllable, leading to periodic activity with a period of a word duration (though the presence of inhibitory connections might make the exact results more complex).

I tested this idea in Endress and Johnson's (2021) model. I exposed the network to a continuous sequence inspired by Saffran, Aslin et al.'s (1996) Experiment 2. The sequence consisted of four distinct words of three syllables each. The familiarization sequence was a random concatenation of these words, with each word occurring 100 times. During the test phase, I recorded the total network activation as each of the test-items (see below) was presented, and assumed that this activation reflects the network's familiarity with the words. I simulated 100 participants by repeating the familiarization and test cycle 100 times.

The test items follow by Saffran, Aslin et al. (1996) and Saffran, Newport et al. (1996), among many others. After exposure to the familiarization sequence, activation is recorded in response to words such as *ABC* and "part-words." As mentioned above, part-words comprise either the last two syllables from one word and the first syllable from the next word (e.g., *BC:D*, where the colon indicates the former word boundary that is not present in the stimuli) or the last syllable from one word and the first two syllables from the next word (e.g., *C:DE*). Part-words are thus attested in the familiarization sequence, but straddle a word boundary. Hence, they have weaker TPs than words. Accordingly, the network should be more familiar with words than with part-words. To assess whether the network can also account for results presented by Fló et al. (2022) (see below), I recorded activation after presenting the first two syllables of a word (e.g., *AB*) or the last two syllables (e.g., *BC*). I also show that the network has problems entraining to a rhythm when longer words are used (Benjamin, Fló, Palu et al., 2023), and that "positional" codes that were assumed to go beyond mere associations among syllables (Henin et al., 2021) can be side effects of associative processing.

During the simulations, the network parameters for self-excitation and mutual inhibition are kept constant (α and β in Supplementary Material A). However, in line with Endress and Johnson (2021), I used different forgetting rates (λ_{act} in Supplementary Material A) between 0.1 and 0.9. With exponential forgetting, a forgetting rate of 1 means that the activation completely disappears on the next time step (in the absence of excitatory input), a forgetting rate of zero means no forgetting at all, while a forgetting rate of 0.5 implies that the activation is halved on the next time step.[4]

3 | RESULTS

3.1 | Words versus part-words

To establish the forgetting rates at which discrimination between words and part-words (and thus learning) can be observed, I first replicate some of Endress and Johnson's (2021) results. I calculated normalized difference scores of activations for words and part-words, $d = \frac{\text{Word}-\text{Part-Word}}{\text{Word}+\text{Part-Word}}$, and evaluated these difference scores in two ways. First, I compared them to the chance level of zero using Wilcoxon tests. Second, I counted the number of simulations (representing different participants) preferring words to part-words, and evaluated this count using a binomial test. With 100 simulations per parameter set, performance is significantly different from the chance level of 50% if at least 61% of the simulations show a preference for the target items.

The results are shown in Figure 2 and tab. B1. Except for low forgetting rates of up to 0.4, the network prefers words over part-words, with somewhat better performance for words against *C:DE* part-words, as has been observed in human participants with syllables (Saffran, Newport et al., 1996), tone sequences (Saffran et al., 1999), and visual shapes (Fiser & Aslin, 2002b). In the following, I will thus use forgetting rates between 0.4 and 0.9 to model the electrophysiological results.

3.2 | Activation differences within words

I next asked whether a basic Hebbian learning model can explain periodic neural activity (e.g., Batterink, 2020; Batterink & Paller, 2017; Benjamin, Fló, Palu et al., 2023; Buiatti et al., 2009; Choi et al., 2020; Fló et al., 2022; Henin et al., 2021; Kabdebon et al., 2015; Moreau et al., 2022; Moser et al., 2021; Sherman et al., 2023), at least for the forgetting rates for which the network preferred words to part-words. In a first analysis, I recorded the total network activation after each syllable in a word had been presented. These activations were averaged for each syllable position (word-initial, word-medial and word-final) and for each participant after removing the first 200 words from the familiarization stream (during which the network was meant to learn).

As shown in Figure 3 and Table B2, activation was highest after word-final syllables (though not for very low forgetting rates for which no learning occurred either). As a result, a simple Hebbian learning model can account for rhythmic activity in electrophysiological recordings with a period equivalent to the word duration. Critically, and as mentioned above, while previous electrophysiological responses to statistical structured streams were interpreted in terms of a response to word onsets (e.g., Abla et al., 2008; Cunillera et al., 2006; Kudo et al., 2011; Sanders et al., 2002; Teinonen et al., 2009), the current results suggest an alternative interpretation of such effects. Rather than signaling the beginnings and ends of words, an activation maximum after the third syllable of each word might reflect the predictability of the third syllable, while a sudden drop in activation after the first syllable might indicate the lack of predictability. Importantly, such activation maxima can arise even if no word is stored in memory.[5]

3.3 | Does the network track specific frequencies?

I next analyzed the frequency response of the network. Specifically, I estimated the spectral density of the time series corresponding to the total network activation after each time step (again after a burn-in of 200 words), separately for each decay rate and simulation. I then extracted the frequency with the maximal density. As shown in Figure 4a, the modal frequency for decay rates of at least 0.4 was 1/3, corresponding to a period of three syllables. These results thus suggest again that a simple Hebbian learning mechanism can entrain to statistical rhythms in the absence of memory for words.

The analyses of the network activations suggest that activations are strongest for word-final syllables, and that the network entrains to a periodicity of three syllables. However, the traditional interpretation of electrophysiological responses to statistical learning is that neural responses index word-initial syllables. To address this issue

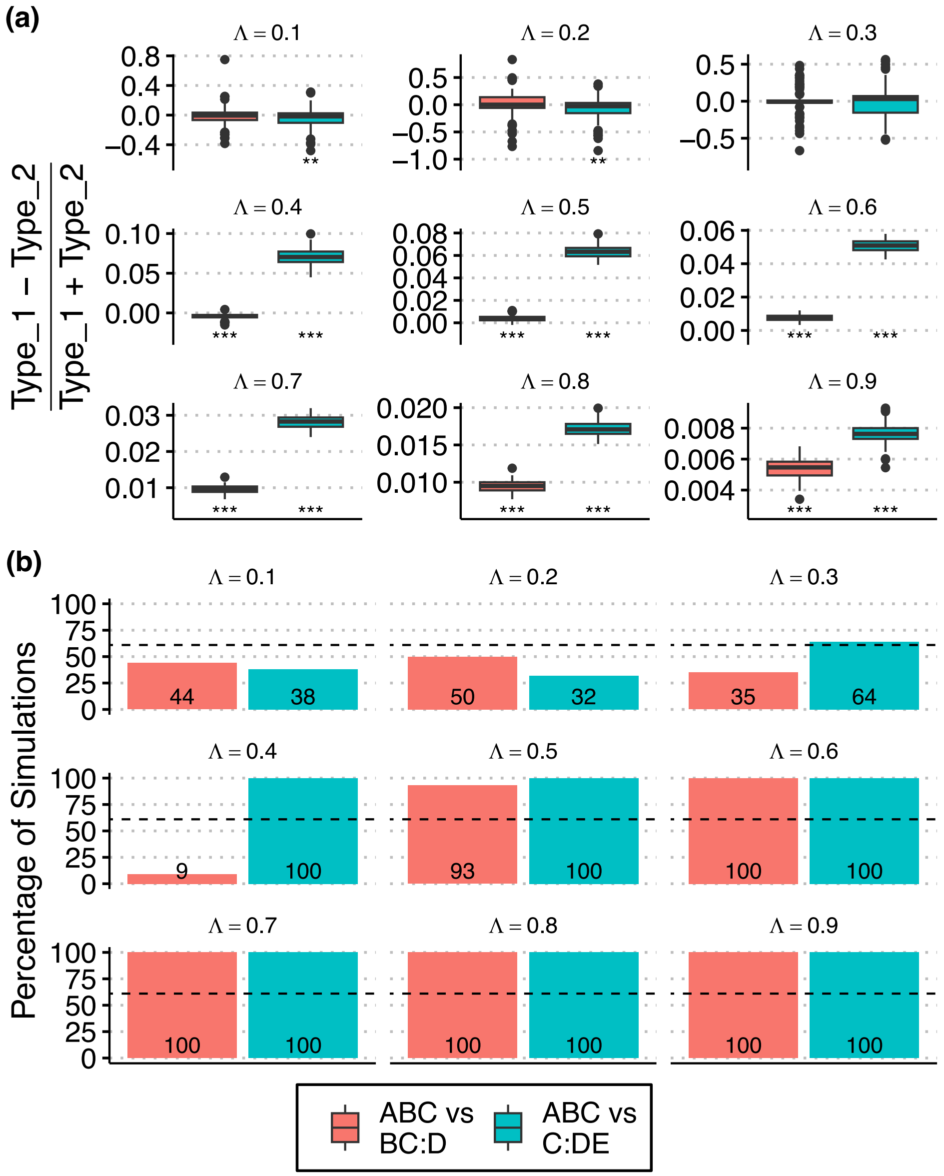

FIGURE 2 Results based on the global activation as a measure of the network's familiarity with words (*ABC* items) and part-words (*BC:D* and *C:DE* items), using forgetting rates (Λ) between 0.1 and 0.9. (a) Difference scores between words and part-words. Significance is assessed using Wilcoxon tests against the chance level of zero. (b) Percentage of simulations with a preference for words over part-words. The dashed line shows the minimum percentage of simulations that is significant according to a binomial test. Both measures show significant preferences for words over part-words for forgetting rates of at least 0.5.

more directly, I calculated the phase of the network activation relative to wave forms with maxima on word-initial, word-medial and word-final syllables, respectively. Specifically, I calculated the cross-spectrum phase at the winning frequency between the total network activation and (1) three cosine reference waves with maxima on the first, second or third syllable of a word, respectively, as well as (2) a saw-tooth function with its maximum on the third syllable. As shown in Figure 4b and Table B4, the activation had a small relative phase relative to the cosine with the maximum on the third syllable or the saw tooth function. In contrast, the phase relative to the cosine with the word-initial maximum was around 120ž, while that relative to the cosine with the maximum on the second syllable was around -120ž. These spectral analyses thus confirm that, at least for reasonable decay rates, the activation increases towards the end of a word, and that the network activation is roughly in phase with a function with a maximum on the third syllable.

3.4 | Memory for word onsets versus offsets (Fló et al., 2022)

The results so far suggest that a simple Hebbian network can reproduce rhythmic activity in the absence of memory for words. However, Fló et al. (2022) provided electrophysiological data suggesting that

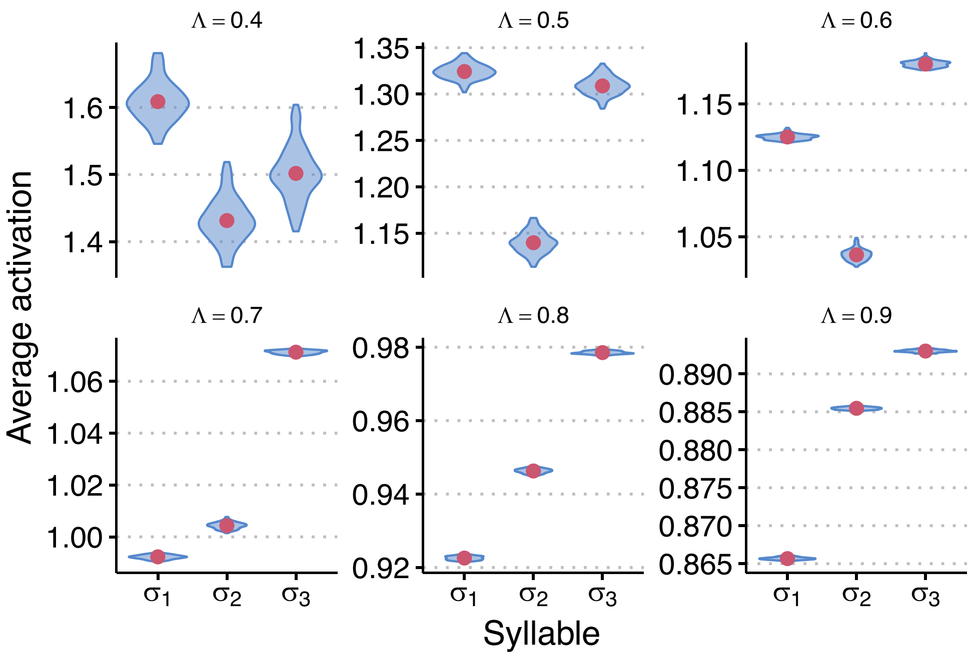

FIGURE 3 Average total network activation for different syllable positions ($\sigma_1, \sigma_2, \sigma_3$) during the familiarization with a stream following Saffran, Aslin et al. (1996). The facets show different forgetting rates. The results reflect the network behavior after the first 50 presentations of each word. In this and all other violin plots, red dots represent sample averages, the error bar the standard error from the mean, and the violin the distribution of the sample. Except for low forgetting rates, the network activation is maximal during the presentation of the word-final syllable.

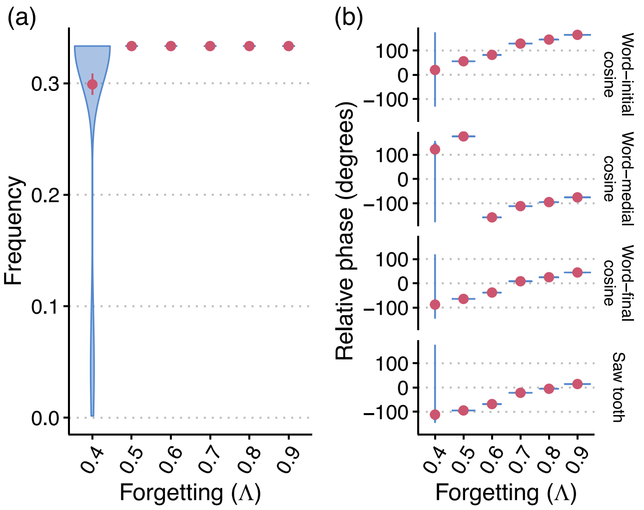

FIGURE 4 Spectral analysis of the total network activation during the familiarization with a stream following Saffran, Aslin et al. (1996). The results reflect the network behavior after the first 50 presentations of each word. (a) Maximal frequency as a function of the forgetting rate. For forgetting rates where learning takes place, the dominant frequency is 1/3, and thus corresponds to the word-length. (b) Relative phase (in degrees) at the maximal frequency of the total network activation relative to (from top to bottom) a cosine function with its maximum at word-initial syllables, word-medial syllables and word-final syllables, respectively, as well as a saw tooth function with the maximum on the word-final syllable. For forgetting rates where learning takes place, the total activation is in phase with a cosine with its maximum on the word-final syllable as well as with the corresponding saw tooth function.

neonates retain at least the first syllable of statistical defined words, if not the entire words. Specifically, after exposure to a speech stream, they presented newborns with items starting with two syllables that occurred word-initially (*AB*...), and with items starting with a word-medial syllable (*BC*...) and observed early ERP differences between these items.

To reproduce these results, I measured the activation of the network in response to isolated, bisyllabic *AB* and *BC* items, respectively. As shown in Figure 5a and Table B5, the network activation was always greater in response to *BC* items than to *AB* items except for the largest decay rates. The reason is presumably that *BC* associations are somewhat stronger than *AB* associations (see Figure 5b), thus leading to more spreading activation in *BC* items than in *AB* items. Be that as it might, these analyses show that a memoryless system can reproduce differential responses to *AB* and *BC* items.

3.5 | Effects of word-length

I next asked whether the network can entrain to statistical regularities when the familiarization streams are composed of words of arbitrary length. Intuitively, given that the periodicity reported here arises due to the increasing cumulative excitatory input towards the ends of words, one would expect the network to be unable to track statistical periodicities for excessively long words. After all, for sufficiently long words, the activation from earlier syllables will have disappeared once the input reaches the end of a word.

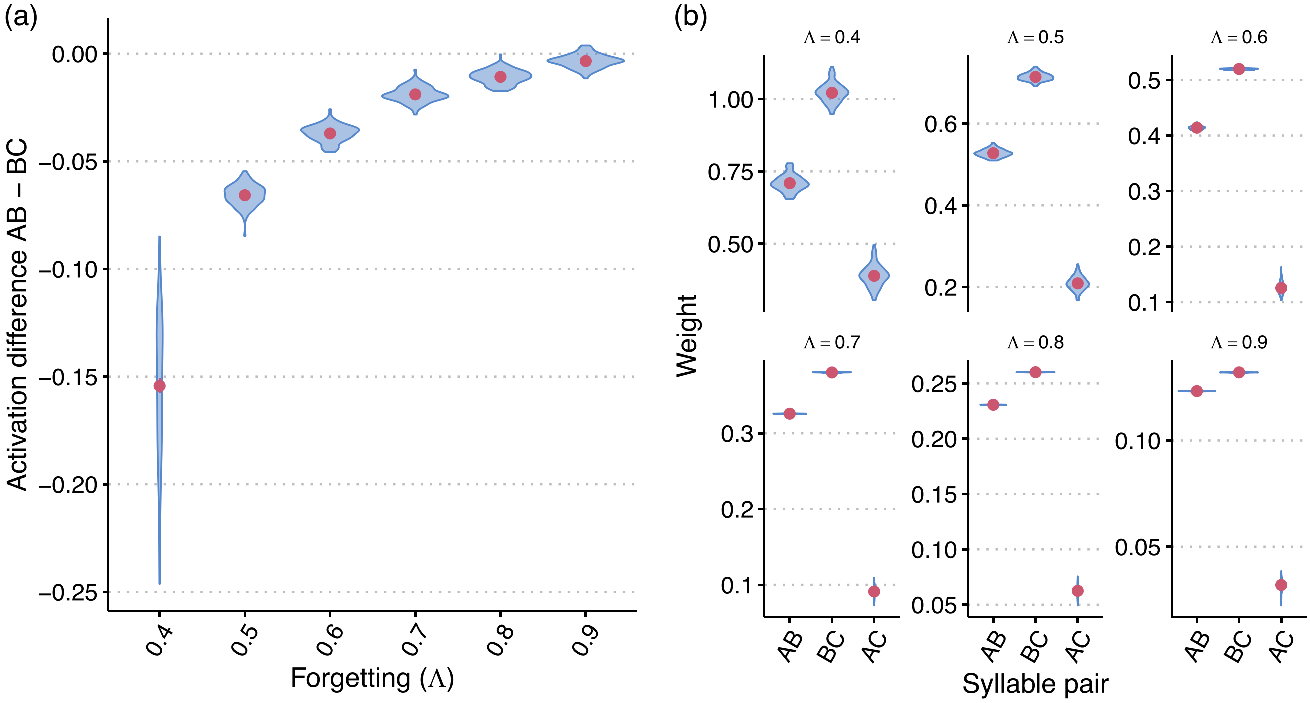

FIGURE 5 (a) Average difference in the total network activation for the first two syllables of a word (*AB*) and the first to syllables of a part-word (*BC*) after familiarization with a stream following Saffran, Aslin et al. (1996). The results reflect the network behavior after the first 50 presentations of each word. Positive values indicate greater activation for *AB* items than for *BC* items. (b) Weights between pairs of syllable representations for different forgetting rates (facets). *BC* connections are somewhat stronger than *AB* connections.

There is some evidence supporting this idea. For example, Benjamin, Fló, Palu et al. (2023) did not find neural entrainment to four-syllable words in newborns (though a failure to detect entrainment might have other reasons than the word-length). Computationally, however, it is also conceivable that networks can deal with longer words, if spreading activation due to higher order associations are sufficient to increase the network activation towards the end of a word.

To examine this issue, I repeated the simulations above, but with word-lengths between 3 and 18 syllables, again for the same forgetting rates as in the simulations above and 100 simulated participants. I estimated the spectral density of the time series corresponding to the total network activation after each time step (again after a burnin of 200 words), separately for each decay rate and simulated participant. I then extracted the frequency with the maximal density, and averaged these frequencies across participants.

As shown in Figure 6, the network successfully tracked the periodicity for word-lengths of up to and including eight syllables. For eight-syllable words, the winning frequency was either $\frac{1}{8}$ or $\frac{1}{4}$, depending on the forgetting rate. In other words, the network sometimes extracted a periodicity whose period was a fraction of the actual word-length. For longer words, the winning frequency was generally a fraction of the word-length, with multipliers of 2 or 3.

As a result, there seems to be a limit to how long words can be so that the network can entrain to a statistically induced rhythm. Here, the limit seems to be a word-length of eight syllables, but the specific limit likely depends on the interplay between the forgetting, excitation, and inhibition parameters.[6]

3.6 | Representational similarity analysis for TPs, ordinal positions, and word identity (Henin et al., 2021)

Other investigators used electrophysiological recordings to probe how more abstract information is extracted from statistically structured sequences. For example, Henin et al. (2021) used intracranial recordings to reveal the kinds of representations that emerge during exposure to statistically structured auditory and visual sequences.

In both modalities, activity in lower-order brain areas (such as the superior temporal gyrus in the auditory modality and occipital and parietal cortex in the visual modality) showed rhythmic activity with periodicities matching both the syllable durations and the word durations (or their visual equivalents). In contrast, higher-order brain areas (such as the inferior frontal gyrus and anterior temporal lobe in auditory modality and the frontal, parietal, and temporal cortex in the visual modality) entrained only at the word-level (or its visual equivalent). Such results might suggest that higher-order brain areas specifically engage in processing of entire statistically defined units. However, given that the current simulations suggest that the word-level rhythmicity might reflect an activation maximum for the most predictable item, it is also possible that this maximal activation is more easily transmitted to subsequent processing stages. Given that the current model just comprises a single level of processing, it thus cannot reproduce such regional processing differences.

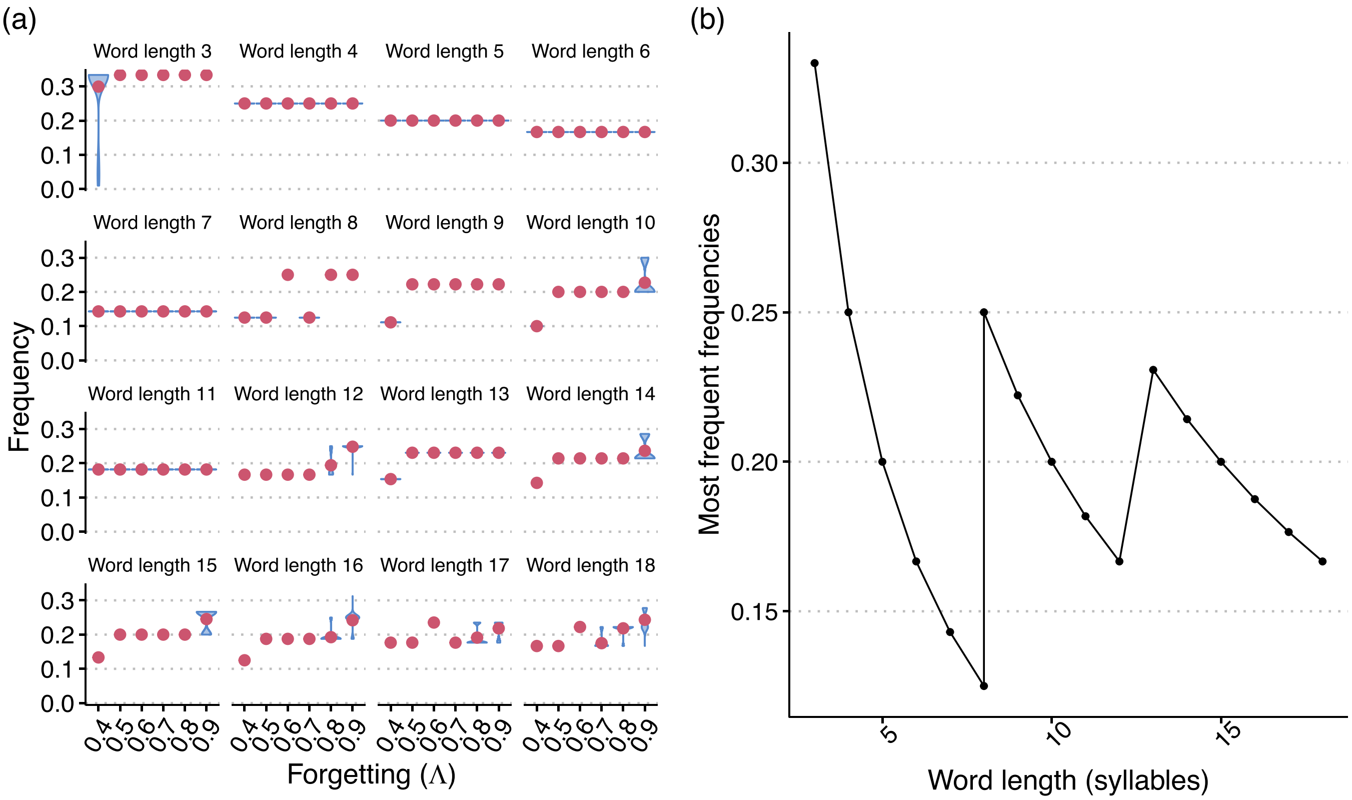

FIGURE 6 Entrainment as a function of word-length. (a) Dominant frequency as a function of the forgetting rate (*x*-axis, Λ) and the word-length (3–18 syllables; facets). (b) Most frequent dominant frequencies across forgetting rates as a function of the word-length. For words of up to and including eight syllables, the network entrains to a frequency equivalent to the word-length. For words of eight or more syllable, the network entrains to a multiple of that frequency.

Henin et al. (2021) also asked how neural populations encode TPs, ordinal positions within words (i.e., word-initial vs. word-medial vs. word-final syllables or their visual analogs) as well as word identity. Using representational similarity analyses, they found that all three types of information are encoded, albeit in different brain areas: TPs were encoded by electrodes entraining to both the syllable- and word-level rhythms, while ordinal positions were encoded by electrodes entraining only to the word-level rhythm. Finally, electrodes entraining to the word-level rhythm as well as electrodes located in the hippocampus encoded word identity. More specifically, Henin et al. (2021) found that the representations of pairs of syllables matching in one of these representational features (i.e., low incoming TPs; same ordinal position; same word) were more similar to one another than those of pairs of syllables mismatching in these features (i.e., high incoming TPs; different ordinal positions; different words).

However, behavioral data raises doubts about the functional role of these representational encodings, and, in fact, decodable information does not necessarily have a functional role.[7] For example, Henin et al. (2021) operationalize low TP syllables as syllables with low incoming TPs (i.e., word-initial syllables), and high TP syllables as syllables with high incoming TPs (i.e., word-medial and word-final syllables). However, this contrast is unlikely to reflect the computational properties of statistical learning, given that TPs do not appear to be directional (e.g., Endress & Wood, 2011; Jones & Pashler, 2007; Turk-Browne & Scholl, 2009). In fact, low incoming TP syllables such as word-initial syllables have high outgoing TPs (because they are highly predictable from word-medial syllables through backwards TPs), while high incoming TP syllables can have low outgoing TPs. Behaviorally, however, TPs do not seem to be directional.

Likewise, while Henin et al. (2021) propose that learners track ordinal positions of syllables within words, substantial behavioral evidence suggests that such positional codes are not available after familiarizations with continuous sequences (e.g., Endress & Bonatti, 2007, 2016; Endress & Mehler, 2009a; Marchetto & Bonatti, 2013; Peña et al., 2002). (While some evidence seems to support the existence such positional codes (e.g., Fló, 2021; Frost & Monaghan, 2016), Frost and Monaghan's (2016) results could not be replicated in Spanish/Catalan speakers (Canudas Grabolosa & Bonatti, unpublished data) or in lab-based or online English speakers (Endress, unpublished data). Further, as shown in Supplementary Material B.6, both sets of studies have alternative interpretations.)

Here, I, thus, propose an alternative interpretation of Henin et al.'s (2021) representational similarity analyses, and suggest that basic associative processes such as those implemented in the current model are sensitive to the same kinds of information on which those analyses rely. This is because syllables in the same ordinal position share contexts, and might be co-activated together with those contexts. For example, syllables in word-initial positions all follow the same set of word-final syllables. As a result, if a syllable and its context activate each other, one would expect a certain degree of representational overlap of syllables sharing a sequential position, even in the absence of any positional codes.

Likewise, representations of syllables from the same word are likely more similar to one another than syllables from different words, again simply because they are likely co-activated with similar contexts (i.e., the rest of the words).

To evaluate these ideas, I calculated the representations of syllables in the following way. First, I extracted the activation in all neurons in all time steps (after burn-in). I then calculated, for each forgetting rate, each simulated participant and each currently presented syllable, the average activation vector (across different presentations of the same syllable) and considered this average the "representation" of a syllable. To simulate Henin et al.'s (2021) representational similarity analyses, I calculated the cosine similarity (i.e., the normalized dot product) across the representations of the critical syllable pairs described below. Finally, I calculated an average similarity for each forgetting rate, simulated participant and syllable pair category.

As mentioned above, Henin et al. (2021) analyzed the similarity of representations based on (1) shared TPs, (2) shared ordinal positions, and (3) shared words. Their operationalization of TP encoding relied on the contrast between syllables with low incoming TPs (i.e., word-initial syllables from different words) and syllables with high incoming TPs (i.e., word-medial and word-final syllables). In line with their analyses, I, thus, asked whether the representations of pairs of syllables with low incoming TPs (i.e., pairs of word-initial syllables) would be more similar to each other than pairs of syllables with high incoming TPs (i.e., pairs of word-medial syllables and pairs of word-medial and word-final syllables; Henin et al. (2021) did not test pairs of word-final syllables). I then averaged these similarity scores for each forgetting rate, simulated participant and pair category (low vs. high incoming TPs), and compared these similarity scores in the difference score $\frac{\text{Low incoming TPs} - \text{High incoming TPs}}{\text{Low incoming TPs} + \text{High incoming TPs}}$ against the chance level of zero, using a Wilcoxon test.

In line with Henin et al.'s (2021) operationalization of the encoding of ordinal positions, I asked whether the representations of pairs of syllables sharing an ordinal positions were more similar to one another than pairs of syllables from different ordinal positions. Again, I averaged these similarity scores for each forgetting rate, simulated participant, and pair category (matching vs. mismatching positions), and compared the difference score $\frac{\text{Match} - \text{Non-match}}{\text{Match} + \text{Non-match}}$ against the chance level of zero, using a Wilcoxon test.

Finally, in line with Henin et al.'s (2021) operationalization of word encoding, I asked whether the representations of pairs of syllables from the same word were more similar to each other than pairs of syllables from different words. Again, I averaged these similarity scores for each forgetting rate, simulated participant and pair category (matching vs. mismatching word), and compared the difference score $\frac{\text{Match} - \text{Non-match}}{\text{Match} + \text{Non-match}}$ against the chance level of zero, using a Wilcoxon test.

As shown in Figure 7a and Table B6, the representations of low incoming TP syllables were *less* similar to each other than the representations of high incoming TP syllables (i.e., syllables in noninitial positions). Likewise, and as shown in Figure 7b and Table B6, pairs of syllables matching in their ordinal positions were *less* similar to each other than pairs of syllables not matching in their ordinal positions. The current model is, thus, sensitive to the same encodings as Henin et al.'s (2021), though the sign of the similarity difference is inverted. I will discuss possible reasons for this inversion below.[8] Finally, and as shown in Figure 7c and Table B6, pairs of syllables from the same word were more similar to one another than syllables from different words.

As mentioned above, these results likely reflect the types of syllables that are co-activated. To illustrate this idea, I summed the activations of all other syllables while a focal syllable was presented, separately for each sequential position of the co-activated syllables. For example, while the word-initial syllable *A* was presented, I separately summed the activations of all other word-initial syllables (excluding *A*), all word-medial syllables, and all word-final syllables, and averaged these sums for all (word-initial, word-medial, or word-final) focal syllable types. As shown in Figure 7d, and unsurprisingly given the localist coding scheme used here, there was little co-activation between syllables of the same (positional) type. In contrast, during presentation of word-initial syllables, word-final syllables were still relatively active. Likewise, during presentation of word-medial syllables, word-initial syllables were relatively active, and word-final syllables were co-activated with word-medial syllables.

These co-activations explain why the network can reproduce Henin et al.'s (2021) operationalization of TP encoding (albeit with an inverted sign). There is little co-activation among word-initial syllables (that is, Henin et al.'s (2021) operationalization of low TP syllables), though there is some overlap in their representations because they are co-active with word-medial and word-final syllables. In contrast, word-medial and word-final syllables (that is, Henin et al.'s (2021) operationalization of high TP syllables) are co-activated, and thus show greater representational overlap.

These co-activations also explain why the network can reproduce Henin et al.'s (2021) encoding of ordinal positions (again with an inverted sign). As mentioned above, to the extent that syllables in the same position have overlapping representations across words, it is via the lingering activation of syllables in *different* positions rather than in the same position.

The reason for the inversion of the sign of the difference for TP- and position-based similarity measures with respect to Henin et al. (2021) is presumably the different time course of the activations in the current model compared to actual biological tissue. In the current simulations, a syllable duration is a discrete time-step. The activations reported here are, thus, snapshots of more continuously evolving activations. As a result, the representations of, say, word-initial and word-final syllables overlap, given that these syllables excite each other in the same time step and still have lingering representations from previous time steps. In contrast, given the localist coding scheme used here, there is no overlap in the representations of syllables occupying the same sequential position.

In contrast, with more realistic activation time courses, the time-resolved similarity measures used by Henin et al. (2021) can capture the actual time courses of the associative activations. For example,

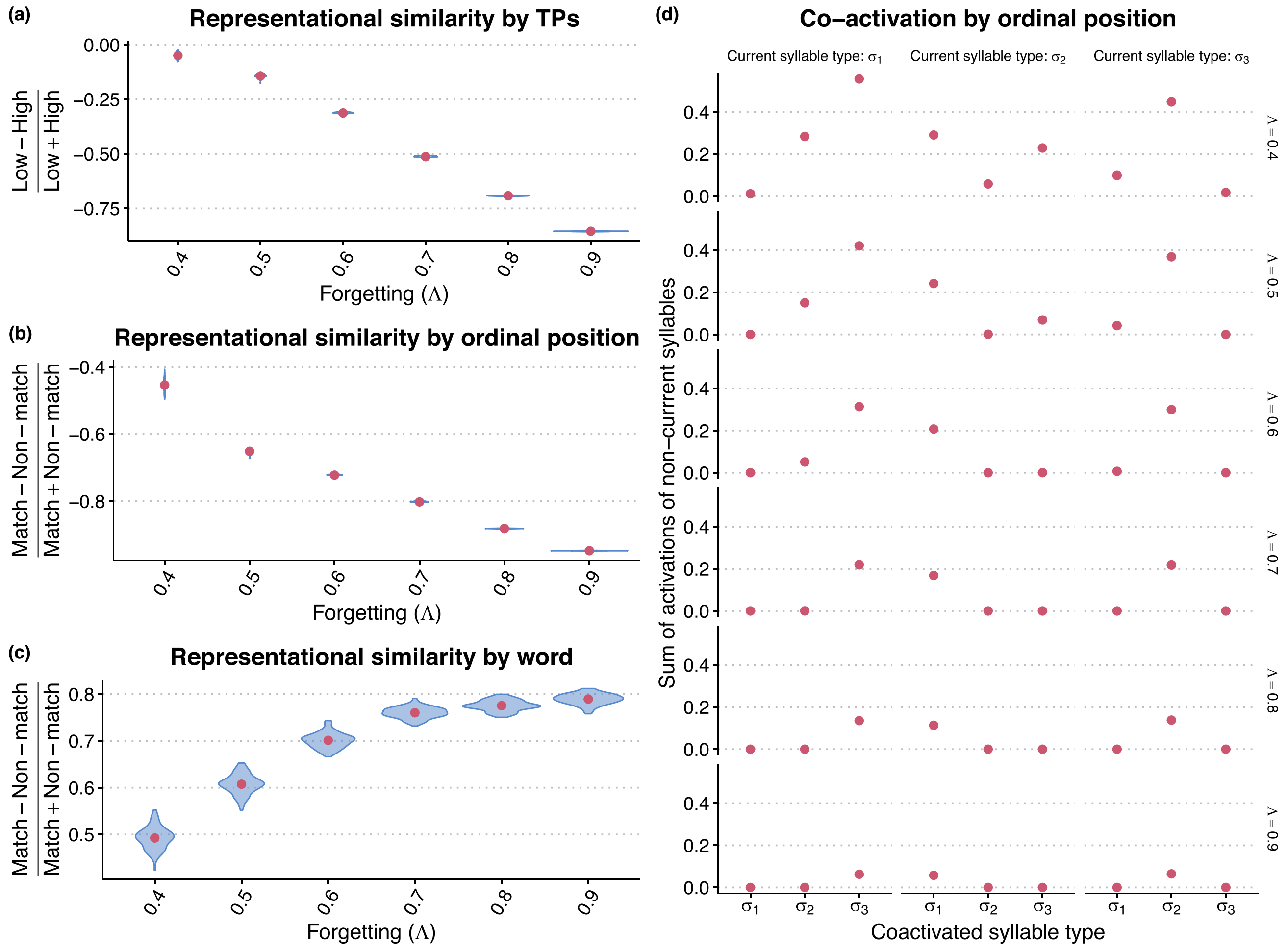

FIGURE 7 (a–c) Simulations of the representational similarity analyses from Henin et al. (2021). The similarity was calculated as the cosine similarity of the average activation vectors elicited by a given syllables. (a) *TP encoding*. Representations of syllables with weak incoming forward TPs (i.e., word-initial syllables) are *less* similar to each other than representations of syllables with strong incoming forward TPs (i.e., word-medial and word-final syllables). (b) *Ordinal position encoding*. Representations of syllables sharing an ordinal position (i.e., word-initial, word-medial, or word-final) are *less* similar to each other than representations of syllables *not* sharing an ordinal position. (c) *Word identity encoding*. Representations of syllables belonging to the same word but *not* sharing an ordinal position are more similar to each other than syllables from different words also not sharing an ordinal position. (d) *Pattern of co-activations*. Sum of co-activations of syllables in different sequential positions (x-axis) during presentations of word-initial, -medial and -final syllables (columns), for different forgetting rates (rows). Most concurrently activated syllables occupy a different ordinal position than the currently presented syllable.

upon presentation of a word-initial syllable, such measures can capture any lingering activation of the representations of the preceding word-final syllable as well as its reactivation through excitation from the word-initial syllable. Likewise, upon presentation of a word-final syllable, such measures can capture the time-courses of activations of the upcoming word-initial syllables. In contrast, such measures are unlikely to make the representations of syllables more similar if they come from different ordinal positions. For example, word-medial and word-final syllables are presumably active simultaneously, but their time course is likely to be different: while word-medial syllables activate upcoming word-final syllables, there is lingering activation of word-medial syllables while word-final syllables are presented. I surmise that these time series make the same same-position representations more similar (at least for word-initial and word-final syllables), without increasing the representational similarity across ordinal positions. Be that as it might, the current model can differentiate between the strength of incoming TPs and between different sequential positions.

Taken together, these results suggest that basic associative mechanisms are sensitive to the same kinds of information on which Henin et al.'s (2021) representational similarity analyses are based. However, to decide between Henin et al.'s (2021) and the current interpretation of these results, a better understanding of the functional role of these representational encodings and of their physiological mechanisms is required.

4 | DISCUSSION

To acquire the words of their native language, learners need to extract them from fluent speech, and might use co-occurrence statistics such as TPs to do so. If so, high-TP items should be stored in memory for later use as words. Strong evidence in favor of this possibility comes from electrophysiology, where rhythmic activity has been observed in response to statistically structured sequences. In the time domain, different authors have observed amplitude peaks around the boundaries of statistically defined words (e.g., Abla et al., 2008; Cunillera et al., 2006; Kudo et al., 2011; Sanders et al., 2002; Teinonen et al., 2009); in the frequency domain, a frequency response with a period of the word duration emerges as participants learn the statistical structure of the speech stream (e.g., Batterink, 2020; Batterink & Paller, 2017; Benjamin, Fló, Palu et al., 2023; Buiatti et al., 2009; Choi et al., 2020; Fló et al., 2022; Henin et al., 2021; Kabdebon et al., 2015; Moreau et al., 2022; Moser et al., 2021; Sherman et al., 2023).

Here, I show that such results can be explained by a simple Hebbian learning model. When exposed to statistically structured sequences, the network activation increased towards the end of words due to increased excitatory input from second order associations. As a result, the network exhibits rhythmic activity with a period of a word duration, at least when the words were not excessively long. Critically, given that the network could reproduce these results in the absence of memory representations for words, earlier electrophysiological results might also index the statistical predictability of syllables rather than the acquisition of coherent units. For example, and as mentioned above, N400 and other time-locked effects observed in statistical learning tasks (e.g., Abla et al., 2008; Cunillera et al., 2006; Kudo et al., 2011; Sanders et al., 2002; Teinonen et al., 2009) might not index the onset of words, but rather the lack of predictability of word-initial syllables (or the increased predictability of word-final syllables). This would also be more consistent with the initial description of the N400 component as an ERP component that indexes *unpredictable* events (Kutas & Federmeier, 2000).[9]

I also show that this network can account for a number of other results. For example, while Henin et al. (2021) proposed that participants in a statistical learning task can develop specific codes for ordinal positions in words, I show that such codes might actually reflect the context in which different items occur, in line with the proposal that such codes are not available after continuous familiarizations (e.g., Endress & Bonatti, 2007, 2016; Endress & Mehler, 2009a; Marchetto & Bonatti, 2013; Peña et al., 2002; but see Fló, 2021; Frost & Monaghan, 2016 and Supplementary Material B.6). However, to decide between Henin et al.'s (2021) and the current interpretation of the positional encodings, a better understanding of the functional role of these encodings and of their physiological mechanisms is required.

As mentioned in the introduction, the view that statistical learning does not necessarily lead to storage in declarative memory is consistent with long-established dissociations between declarative memory and implicit learning (e.g., Cohen & Squire, 1980; Finn et al., 2016; Graf & Mandler, 1984; Knowlton et al., 1996; Poldrack et al., 2001; Squire, 1992). It is also consistent with a variety of behavioral results (see Endress & de Seyssel, under review; Endress et al., 2020, for critical reviews), including behavioral preferences for unattested high-TP items (e.g., Endress & Langus, 2017; Endress & Mehler, 2009b; Endress & Wood, 2011; Jones & Pashler, 2007; Turk-Browne & Scholl, 2009), and the inability of adult learners to repeat back words from familiarization streams with as few as four words Endress and de Seyssel (under review) and even when they entrain to a statistical rhythm (Batterink, 2020).

While different memory systems can interact during acquisition (Robertson, 2022) and statistical learning might thus facilitate subsequent word learning, learners might still need to rely on other cues to identify words and word boundaries in fluent speech. Such cues include using known words as cues to word boundaries for other words (e.g., Bortfeld et al., 2005; Brent & Siskind, 2001; Mersad & Nazzi, 2012), paying attention to beginnings and ends of utterances (e.g., Monaghan & Christiansen, 2010; Seidl & Johnson, 2008; Shukla et al., 2007), phonotactic regularities (e.g., McQueen, 1998), and universal aspects of prosody (e.g., Brentari et al., 2011; Christophe et al., 2001; Endress & Hauser, 2010; Pilon, 1981). Computational results suggest that such cues are promising, given that a computational model attending to utterance edges showed excellent segmentation and word-learning abilities (Monaghan & Christiansen, 2010).

In contrast, statistical learning might well be important for predicting events across time (e.g., Endress & de Seyssel, under review; Morgan et al., 2019; Sherman & Turk-Browne, 2020; Turk-Browne et al., 2010; Verosky & Morgan, 2021) and space (Theeuwes et al., 2022), an ability that is clearly critical for mature language processing (e.g., Levy, 2008; Trueswell et al., 1999) as well as many other processes (e.g., Clark, 2013; Friston, 2010; Keller & Mrsic-Flogel, 2018). This suggests that predictive processing might also be crucial for word learning, but it is an important topic for further research to find out how predictive processing is used during language acquisition and which mechanisms are used for word segmentation.

ACKNOWLEDGMENTS

An `R` implementation of the model is available at https://github.com/aendress/tp_model_entrainment and https://doi.org/10.25383/city.22679410. I am indebted to A. Fló (Università degli Studi di Padova) for helpful comments on an earlier version of this manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Figshare at https://doi.org/10.25383/city.22679410.

ORCID

Ansgar D. Endress https://orcid.org/0000-0003-4086-5167

ENDNOTE

[1] An alternative explanation of such results is that learners do have declarative memory representations of statistically defined units, but then mentally reverse them to recognize backward items, or recognize unheard

items by computing their similarity to attested items. To date, however, there is no evidence for this possibility.

[2] There is substantial evidence that neural activation can be *suppressed* for predictable stimuli (e.g., de Lange et al., 2018), which is likely important of adaptive behavior (e.g., for learning; Stahl & Feigenson, 2017). Here, I just seek to provide a proof-of-concept model suggesting that simple Hebbian learning mechanisms can explain rhythmic brain activity in response to statistical learning tasks. Detection of predictable or old stimuli might rely on subsequent processing and involve complex computational mechanisms beyond the scope of the current investigations (e.g., a comparator mechanisms for novelty detection in the hippocampus; e.g., Kumaran & Maguire, 2007).

[3] These choices are to some extent arbitrary. While, in the current implementation, only excitatory connections are tuned by learning, the same results can likely be reproduced by tuning inhibition, for example, through tunable disinhibitory interneurons (Letzkus et al., 2011). Likewise, decay and interference might also affect connection weights rather than only activations (see Tovar & Westermann, 2023 vs. Endress & Johnson, 2023). Here, I just seek to show that Hebbian learning can explain rhythmic neural activity in a statistical learning task with a fairly generic network model, with no claim to psychological or biological realism.

[4] While I use the label "decay," I do not claim that "decay" reflects a psychological processes. The current implementation uses decay as a mechanism to limit activations in time, but the same effect could likely be obtained through inhibitory interactions or other mechanisms.

[5] The reason for which lower forgetting rates do not necessarily lead to rhythmic activity is the interplay between decay and inhibition. To assess this possibility, I recorded the number of active neurons after a burn-in phase of 600 items. As shown in Table B3 and Figure B1, more neurons remain active when the decay rate is lower, and can thus inhibit other neurons. When decay limits the effect of residual inhibitory input from other neurons, the pattern of connections between neurons then enables the network to exhibit periodic activity as well as a preference for high-TP items over low-TP items. In fact, for decay rates below 0.4, the estimated frequencies were 1/600 (i.e., the reciprocal of the total number of syllables) or 1/300, suggesting that no meaningful periodic activity was detected.

[6] See Supplementary Material C for other quantitative results where the network makes incorrect predictions.

[7] To take a non-psychological example, audio recordings often contain noise from the electric grid from which spatial and temporal localization information can be decoded (e.g., for forensic purposes; e.g. Grigoras, 2005). While this information is clearly present in the recordings, it is not relevant for the primary means by which audio information is consumed (i.e., by listening to it). *Mutatis mutandis*, some information might be present in neural activity as a side effect of the mechanics of neural processing, but whether this information is behaviorally relevant is an independent and empirical question, similarly to how, over evolutionary times, phenotypic traits might not have evolved for specific purposes, but rather reflect "spandrels" that evolved as side effects of other evolutionary processes.

[8] The current model cannot reproduce Henin et al.'s (2021) result that TP and positional encoding predominantly occurs in different sets of electrodes, simply because the current model comprises only a single level of processing.

[9] Of course, the ability of a Hebbian network to reproduce electrophysiological results does not necessarily imply that statistically defined units are not *also* represented as integrated memory items. A possible prediction to separate the word-learning view from the mere association view relies on the fact that TPs are not directional (e.g., Endress & Wood, 2011; Jones & Pashler, 2007; Turk-Browne & Scholl, 2009). As a result, according to the mere-association view, after prefamiliarization with a speech stream composed of words such as *ABC*, *DEF*, and *GHI*, entrainment to a speech stream composed of reversed words such as *CBA*, *FED*, and *IHG*, should be immediate, and much faster than entrainment to a speech stream composed of scrambled words such as *ECG* and *FAH*. In contrast, the word-learning view would predict that words need to be relearned in both the reversed and the scrambled stream, with similar lead times until a rhythmic response emerges.

REFERENCES

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience*, *20*(6), 952–964. https://doi.org/10.1162/jocn.2008.20058

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.

Batchelder, E. O. (2002). Bootstrapping the lexicon: A computational model of infant speech segmentation. *Cognition*, *83*(2), 167–206.

Batterink, L. J. (2020). Syllables in sync form a link: Neural phase-locking reflects word knowledge during language learning. *Journal of Cognitive Neuroscience*, *32*, 1735–1748. https://doi.org/10.1162/jocn_a_01581

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62–78. https://doi.org/10.1016/j.jml.2015.04.004

Bays, P. M., Singh-Curry, V., Gorgoraptis, N., Driver, J., & Husain, M. (2010). Integration of goal- and stimulus-related visual signals revealed by damage to human parietal cortex. *Journal of Neuroscience*, *30*, 5968–5978. https://doi.org/10.1523/JNEUROSCI.0997-10.2010

Benjamin, L., Fló, A., Al Roumi, F., & Dehaene-Lambertz, G. (2023). Humans parsimoniously represent auditory sequences by pruning and completing the underlying network structure. *eLife*, *12*. https://doi.org/10.7554/eLife.86430

Benjamin, L., Fló, A., Palu, M., Naik, S., Melloni, L., & Dehaene-Lambertz, G. (2023). Tracking transitional probabilities and segmenting auditory sequences are dissociable processes in adults and neonates. *Developmental Science*, *26*, e13300. https://doi.org/10.1111/desc.13300

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*(4), 298–304. https://doi.org/10.1111/j.0956-7976.2005.01531.x

Brent, M., & Cartwright, T. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*(1-2), 93–125.

Brent, M., & Siskind, J. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33–B44.

Brentari, D., González, C., Seidl, A., & Wilbur, R. (2011). Sensitivity to visual prosodic cues in signers and nonsigners. *Language and Speech*, *54*(1), 49–72.

Buiatti, M., Peña, M., & Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage*, *44*(2), 509–519. https://doi.org/10.1016/j.neuroimage.2008.09.015

Chen, J., & Ten Cate, C. (2015). Zebra finches can use positional and transitional cues to distinguish vocal element strings. *Behavioural Processes*, *117*, 29–34. https://doi.org/10.1016/j.beproc.2014.09.004

Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., & Werker, J. F. (2020). Preverbal infants discover statistical word patterns at similar rates as adults: Evidence from neural entrainment. *Psychological Science*, *31*, 1161–1173. https://doi.org/10.1177/0956797620933237

Christiansen, M. H. (2018). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, *11*(3), 468–481. https://doi.org/10.1111/tops.12332

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, *13*(2–3), 221–268.

Christophe, A., Mehler, J., & Sebastian-Galles, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, *2*(3), 385–394.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/s0140525x12000477

Cohen, N., & Squire, L. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, *210*(4466), 207–210. https://doi.org/10.1126/science.7414331

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24–39. https://doi.org/10.1037/0278-7393.31.1.24

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1119–1130. https://doi.org/10.1037/0278-7393.30.5.1119

Cunillera, T., Toro, J. M., Sebastián-Gallés, N., & Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research*, *1123*, 168–178. https://doi.org/10.1016/j.brainres.2006.09.046

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*, 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Endress, A. D. (2010). Learning melodies from non-adjacent tones. *Acta Psychologica*, *135*(2), 182–190. https://doi.org/10.1016/j.actpsy.2010.06.005

Endress, A. D. (2019). Duplications and domain-generality. *Psychological Bulletin*, *145*(12), 1154–1175. https://doi.org/10.1037/bul0000213

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*(2), 247–299. https://doi.org/10.1016/j.cognition.2006.09.010

Endress, A. D., & Bonatti, L. L. (2016). Words, rules, and mechanisms of language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(1), 19–35. https://doi.org/10.1002/wcs.1376

Endress, A. D., & de Seyssel, M. (under review). The specificity of sequential statistical learning: Statistical learning accumulates predictive information from unstructured input but is dissociable from (declarative) memory. *Cognition*.

Endress, A. D., & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology*, *61*(2), 177–199. https://doi.org/10.1016/j.cogpsych.2010.05.001

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

Endress, A. D., & Johnson, S. P. (2023). Hebbian, correlational learning provides a memory-less mechanism for statistical learning irrespective of implementational choices. *Cognition*, *230*, 105290. https://doi.org/10.1016/j.cognition.2022.105290

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

Endress, A. D., & Mehler, J. (2009a). Primitive computations in speech processing. *Quarterly Journal of Experimental Psychology*, *62*(11), 2187–2209. https://doi.org/10.1080/17470210902783646

Endress, A. D., & Mehler, J. (2009b). 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. https://doi.org/10.1016/j.jml.2008.10.003

Endress, A. D., Slone, L. K., & Johnson, S. P. (2020). Statistical learning and memory. *Cognition*, *204*, 104346. https://doi.org/10.1016/j.cognition.2020.104346

Endress, A. D., & Szabó, S. (2020). Sequential presentation protects memory from catastrophic interference. *Cognitive Science*, *44*(5). https://doi.org/10.1111/cogs.12828

Endress, A. D., & Wood, J. N. (2011). From movements to actions: Two mechanisms for learning action sequences. *Cognitive Psychology*, *63*(3), 141–171. https://doi.org/10.1016/j.cogpsych.2011.07.001

Erickson, L. C., Thiessen, E. D., & Estes, K. G. (2014). Statistically coherent labels facilitate categorization in 8-month-olds. *Journal of Memory and Language*, *72*, 49–58. https://doi.org/10.1016/j.jml.2014.01.002

Finn, A. S., Kalra, P. B., Goetz, C., Leonard, J. A., Sheridan, M. A., & Gabrieli, J. D. (2016). Developmental dissociation between the maturation of procedural memory and declarative memory. *Journal of Experimental Child Psychology*, *142*, 212–220. https://doi.org/10.1016/j.jecp.2015.09.027

Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458–467.

Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15822–15826. https://doi.org/10.1073/pnas.232472899

Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, *134*(4), 521–537. https://doi.org/10.1037/0096-3445.134.4.521

Fló, A., Benjamin, L., Palu, M., & Dehaene-Lambertz, G. (2022). Sleeping neonates track transitional probabilities in speech but only retain the first syllable of words. *Scientific Reports*, *12*, 4391. https://doi.org/10.1038/s41598-022-08411-w

Fló, A. (2021). Evidence of ordinal position encoding of sequences extracted from continuous speech. *Cognition*, *213*, 104646. https://doi.org/10.1016/j.cognition.2021.104646

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, *117*(2), 107–125. https://doi.org/10.1016/j.cognition.2010.07.005

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010

Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, *147*, 70–74. https://doi.org/10.1016/j.cognition.2015.11.010

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

Gallistel, C. R., Mark, T. A., King, A. P., & Latham, P. E. (2001). The rat approximates an ideal detector of changes in rates of reward: Implications for the law of effect. *Journal of Experimental Psychology. Animal Behavior Processes*, *27*, 354–372.

Gillette, J., Gleitman, H., Gleitman, L. R., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176.

Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science*, *33*, 260–272. https://doi.org/10.1111/j.1551-6709.2009.01012.x

Glicksohn, A., & Cohen, A. (2011). The role of gestalt grouping principles in visual statistical learning. *Attention, Perception and Psychophysics*, *73*(3), 708–713. https://doi.org/10.3758/s13414-010-0084-4

Gottlieb, J. (2007). From thought to action: The parietal cortex as a bridge between perception, action, and cognition. *Neuron*, *53*, 9–16.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior*, *23*(5), 553–568. https://doi.org/10.1016/s0022-5371(84)90346-3

Graf-Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*(3), 254–260. https://doi.org/10.1111/j.1467-9280.2007.01885.x

Grigoras, C. (2005). Digital audio recording analysis: The electric network frequency (ENF) criterion. *International Journal of Speech, Language and the Law*, *12*(1), 63–76. https://doi.org/10.1558/sll.2005.12.1.63

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53–64.

Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology*, *63*(2), 93–106. https://doi.org/10.1016/j.cogpsych.2011.06.002

Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., Doyle, W., Devinsky, O., & Melloni, L. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Science Advances*, *7*. https://doi.org/10.1126/sciadv.abc4530

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically induced chunking recall: A memory-based approach to statistical learning. *Cognitive Science*, *44*, e12848. https://doi.org/10.1111/cogs.12848

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*(4), 548–567.

Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science*, *12*(1), 131–141. https://doi.org/10.1111/j.1467-7687.2008.00740.x

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339–345. https://doi.org/10.1111/j.1467-7687.2009.00886.x

Jones, J., & Pashler, H. (2007). Is the mind inherently forward looking? Comparing prediction and retrodiction. *Psychonomic Bulletin & Review*, *14*, 295–300. https://doi.org/10.3758/bf03194067

Kabdebon, C., Pena, M., Buiatti, M., & Dehaene-Lambertz, G. (2015). Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain and Language*, *148*, 25–36. https://doi.org/10.1016/j.bandl.2015.03.005

Karaman, F., & Hay, J. F. (2018). The longevity of statistical learning: When infant memory decays, isolated words come to the rescue. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(2), 221–232. https://doi.org/10.1037/xlm0000448

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive processing: A canonical cortical computation. *Neuron*, *100*(2), 424–435. https://doi.org/10.1016/j.neuron.2018.10.003

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. https://doi.org/10.1016/s0010-0277(02)00004-5

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 1399–1402.

Kudo, N., Nonaka, Y., Mizuno, N., Mizuno, K., & Okanoya, K. (2011). On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Developmental Science*, *14*(5), 1100–1106. https://doi.org/10.1111/j.1467-7687.2011.01056.x

Kumaran, D., & Maguire, E. A. (2007). Which computational mechanisms operate in the hippocampus during novelty detection? *Hippocampus*, *17*, 735–748. https://doi.org/10.1002/hipo.20326

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470.

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. https://doi.org/10.1038/nature10674

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Lynn, C. W., Kahn, A. E., Nyema, N., & Bassett, D. S. (2020). Abstract representations of events arise from mental errors in learning and memory. *Nature Communications*, *11*, 2313. https://doi.org/10.1038/s41467-020-15146-7

Marchetto, E., & Bonatti, L. L. (2013). Words and possible words in early language acquisition. *Cognitive Psychology*, *67*(3), 130–150. https://doi.org/10.1016/j.cogpsych.2013.08.001

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*(1), 21–46.

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(22), 9014–9019. https://doi.org/10.1073/pnas.1105040108

Mersad, K., & Nazzi, T. (2012). When mommy comes to the rescue of statistics: Infants combine top-down and bottom-up cues to segment speech. *Language Learning and Development*, *8*(3), 303–315. https://doi.org/10.1080/15475441.2011.609106

Monaghan, P., & Christiansen, M. H. (2010). Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. *Journal of Child Language*, *37*(3), 545–564. https://doi.org/10.1017/S0305000909990511

Moreau, C. N., Joanisse, M. F., Mulgrew, J., & Batterink, L. J. (2022). No statistical learning advantage in children over adults: Evidence from behaviour and neural entrainment. *Developmental Cognitive Neuroscience*, *57*, 101154. https://doi.org/10.1016/j.dcn.2022.101154

Morgan, E., Fogel, A., Nair, A., & Patel, A. D. (2019). Statistical learning and gestalt-like principles predict melodic expectations. *Cognition*, *189*, 23–34. https://doi.org/10.1016/j.cognition.2018.12.015

Moser, J., Batterink, L. J., Li Hegner, Y., Schleger, F., Braun, C., Paller, K. A., & Preissl, H. (2021). Dynamics of nonlinguistic statistical learning: From neural entrainment to the emergence of explicit knowledge. *NeuroImage*, *240*, 118378. https://doi.org/10.1016/j.neuroimage.2021.118378

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(7), 2745–2750. https://doi.org/10.1073/pnas.0708424105

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*(5593), 604–607. https://doi.org/10.1126/science.1072901

Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. *Topics in Cognitive Science*, *11*, 520–535. https://doi.org/10.1111/tops.12403

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*, 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*, 246–263.

Pilon, R. (1981). Segmentation of speech in a foreign language. *Journal of Psycholinguistic Research*, *10*(2), 113–122.

Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, *414*, 546–550. https://doi.org/10.1038/35107080

Robertson, E. M. (2022). Memory leaks: Information shared across memory systems. *Trends in Cognitive Sciences*, *26*, 544–554. https://doi.org/10.1016/j.tics.2022.03.010

Roggeman, C., Fias, W., & Verguts, T. (2010). Salience maps in parietal cortex: Imaging and computational modeling. *NeuroImage*, *52*, 1005–1014. https://doi.org/10.1016/j.neuroimage.2010.01.060

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, *37*(1), 74–85.

Saffran, J. R., Johnson, E., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*, *5*, 700–703. https://doi.org/10.1038/nn873

Seidl, A., & Johnson, E. K. (2008). Boundary alignment enables 11-month-olds to segment vowel initial words from speech. *Journal of Child Language*, *35*(1), 1–24.

Sengupta, R., Surampudi, B. R., & Melcher, D. (2014). A visual sense of number emerges from the dynamics of a recurrent on-center off-surround neural network. *Brain Research*, *1582*, 114–124. https://doi.org/10.1016/j.brainres.2014.03.014

Sherman, B. E., Aljishi, A., Graves, K. N., Quraishi, I. H., Sivaraju, A., Damisah, E. C., & Turk-Browne, N. B. (2023). Intracranial entrainment reveals statistical learning across levels of abstraction. *Journal of Cognitive Neuroscience*, 1–17. https://doi.org/10.1162/jocn_a_02012

Sherman, B. E., & Turk-Browne, N. B. (2020). Statistical prediction of the future impairs episodic encoding of the present. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 22760–22770. https://doi.org/10.1073/pnas.2013291117

Shoaib, A., Wang, T., Hay, J. F., & Lany, J. (2018). Do infants learn words from statistics? Evidence from English-learning infants hearing Italian. *Cognitive Science*, *42*(8), 3083–3099. https://doi.org/10.1111/cogs.12673

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*(1), 1–32. https://doi.org/10.1016/j.cogpsych.2006.04.002

Slone, L. K., & Johnson, S. P. (2018). When learning goes beyond statistics: Infants represent visual sequences in terms of chunks. *Cognition*, *178*, 92–102. https://doi.org/10.1016/j.cognition.2018.05.016

Sohail, J., & Johnson, E. K. (2016). How transitional probabilities and the edge effect contribute to listeners' phonological bootstrapping success. *Language Learning and Development*, 1–11. https://doi.org/10.1080/15475441.2015.1073153

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*(2), 195–231. https://doi.org/10.1037/0033-295x.99.2.195

Stahl, A. E., & Feigenson, L. (2017). Expectancy violations promote learning in young children. *Cognition*, *163*, 1–14. https://doi.org/10.1016/j.cognition.2017.02.008

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*, 21. https://doi.org/10.1186/1471-2202-10-21

Theeuwes, J., Bogaerts, L., & van Moorselaar, D. (2022). What to expect where and when: How statistical learning drives visual selection. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2022.06.001

Tompson, S. H., Kahn, A. E., Falk, E. B., Vettel, J. M., & Bassett, D. S. (2019). Individual differences in learning social and nonsocial network structures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 253–271. https://doi.org/10.1037/xlm0000580

Toro, J. M., Trobalon, J. B., & Sebastián-Gallés, N. (2005). Effects of backward speech and speaker variability in language discrimination by rats. *Journal of Experimental Psychology. Animal Behavior Processes*, *31*(1), 95–100. https://doi.org/10.1037/0097-7403.31.1.95

Tovar, A. E., & Westermann, G. (2023). No need to forget, just keep the balance: Hebbian neural networks for statistical learning. *Cognition*, *230*, 105176. https://doi.org/10.1016/j.cognition.2022.105176

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89–134.

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Turk-Browne, N. B., & Scholl, B. J. (2009). Flexible visual statistical learning: Transfer across space and time. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(1), 195–202.

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience*, *30*, 11177–11187. https://doi.org/10.1523/JNEUROSCI.0858-10.2010

Verosky, N. J., & Morgan, E. (2021). Pitches that wire together fire together: Scale degree associations across time predict melodic expectations. *Cognitive Science*, *45*, e13037. https://doi.org/10.1111/cogs.13037

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Endress, A. D. (2024). Hebbian learning can explain rhythmic neural entrainment to statistical regularities. *Developmental Science*, e13487. https://doi.org/10.1111/desc.13487