Cognition 261 (2025) 106130

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Full length article

# The specificity of sequential statistical learning: Statistical learning accumulates predictive information from unstructured input but is dissociable from (declarative) memory for words[☆]

Ansgar D. Endress[a,*], Maureen de Seyssel[b,c,1]

[a] Department of Psychology, City St George's, University of London, UK
[b] Laboratoire de Sciences Cognitives et de Psycholinguistique, Département d'Etudes Cognitives, ENS, EHESS, CNRS, PSL University, Paris, France
[c] Laboratoire de Linguistique Formelle, Université Paris Cité, CNRS, Paris, France

ARTICLE INFO

Dataset link: https://github.com/aendress/segmenation_recall, https://figshare.com/s/dc3bf0cd35fe47156e99

*Keywords:*
Statistical learning
Declarative memory
Predictive processing
Language acquisition
Hebbian learning

ABSTRACT

Learning statistical regularities from the environment is ubiquitous across domains and species. It might support the earliest stages of language acquisition, especially identifying and learning words from fluent speech (i.e., word-segmentation). But how do the statistical learning mechanisms involved in word-segmentation interact with the memory mechanisms needed to remember words — and with the learning situations where words need to be learned? Through computational modeling, we first show that earlier results purportedly supporting memory-based theories of statistical learning can be reproduced by memory-less Hebbian learning mechanisms. We then show that, in a memory recall task after exposure to continuous, statistically structured speech sequences, participants track the statistical structure of the speech sequences and are thus sensitive to probable syllable transitions. However, they hardly remember any items at all, with 82% producing no high-probability items. Among the 30% of participants producing (correct) high- or (incorrect) low-probability items, half produced high-probability items and half low-probability items — even while preferring high-probability items in a recognition test. Only discrete familiarization sequences with isolated words yield memories of actual items. Turning to how specific learning situations affect statistical learning, we show that it predominantly operates in continuous speech sequences like those used in earlier experiments, but not in discrete chunk sequences likely more characteristic of early language acquisition. Taken together, these results suggest that statistical learning might be specialized to accumulate distributional information, but that it is dissociable from the (declarative) memory mechanisms needed to acquire words and does not allow learners to identify probable word boundaries.

## 1. Introduction

The ability to learn statistical regularities from the environment is remarkably widespread across species and domains (Aslin, Saffran, & Newport, 1998; Chen & Ten Cate, 2015; Hauser, Newport, & Aslin, 2001; Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996; Toro, Trobalon, & Sebastián-Gallés, 2005; Turk-Browne & Scholl, 2009), and might support a wide range of computations (e.g., Sherman, Graves, & Turk-Browne, 2020). Forms of statistical learning that allow learners to track sequential dependencies among sequence items might be especially important during language acquisition (Aslin & Newport, 2012; Saffran & Kirkham, 2018). However, their computational function is unclear. It is widely believed that such forms of statistical learning help learners acquire words from fluent speech (e.g., Aslin et al., 1998; Saffran, Aslin, & Newport, 1996), and thus (presumably) store word candidates in (declarative) memory (Graf-Estes, Evans, Alibali, & Saffran, 2007; Isbilen, McCauley, Kidd, & Christiansen, 2020). However, other authors suggest that statistical learning is important for predicting events (Sherman & Turk-Browne, 2020; Turk-Browne, Scholl, Johnson, & Chun, 2010). Here, we suggest that

☆ We are grateful to E. Dupoux, L. Leisten, and K. Hitczenko for helpful discussions about earlier versions of this manuscript. ADE was supported by a grant from Spanish Ministerio de Economía y Competitividad and Marie Curie Incoming Fellowship 303163-COMINTENT. The authors declare no conflict of interest. ADE performed research for Experiment 1. Both authors performed research for Experiment 2 and wrote the paper. Experiments, data, and analysis code are available at https://github.com/aendress/segmentation_recall and https://doi.org/10.25383/city.28592924. Experiment 1 is available as a template in the Testable Library (https://www.testable.org/library#496496; Statistical learning and memory recall).

* Correspondence to: Department of Psychology, City St George's, University of London, Northampton Square, London EC1V 0HB, UK.
*E-mail address:* ansgar.endress.1@citystgeorges.ac.uk (A.D. Endress).
[1] Currently at Apple.

https://doi.org/10.1016/j.cognition.2025.106130
Received 14 March 2025; Accepted 24 March 2025
Available online 17 April 2025

statistical learning is critical for predicting speech material and operates predominantly under conditions where prediction is possible. However, we also suggest that statistical learning does not lead to declarative memories of words, and that separate mechanisms are required to form these memories.

We note that the label "statistical learning" has also been used for a variety of other computations, including discovering phonemic and allophonic categories (e.g., Maye, Werker, & Gerken, 2002), learning relevant locations in visual search (e.g., van Moorselaar & Slagter, 2019), compressing redundant information in visual working memory (e.g., Brady, Konkle, & Alvarez, 2009), among others (see Sherman et al., 2020, for a review). Here, we focus on forms of statistical learning that allow learners to track sequential dependencies among items in continuous sequences (and possibly also to associate simultaneously presented items in vision). We surmise that other computations referred to as "statistical learning" likely rely on different mechanisms and might well have different properties.

### 1.1. Statistical learning vs. declarative memory of words in fluent speech

Speech is often thought to be a continuous signal (and often perceived as such in unknown languages, but see below), and before learners can commit any words to memory, they need to learn where words start and where they end. To do so, they might rely on Transitional Probabilities (TPs) among syllables, that is, the conditional probability of a syllable $\sigma_{i+1}$ given a preceding syllable $\sigma_i$, $P(\sigma_i\sigma_{i+1})/P(\sigma_i)$. Relatively predictable transitions are likely located inside words, while unpredictable ones straddle word boundaries. Early on, Shannon (1951) showed that human adults are sensitive to such distributional information. Subsequent work demonstrated that infants and non-human animals share this ability (Chen & Ten Cate, 2015; Hauser et al., 2001; Kirkham et al., 2002; Saffran, Aslin, & Newport, 1996; Toro, Trobalon, & Sebastián-Gallés, 2005; Turk-Browne & Scholl, 2009).

Statistical learning therefore supports predictive processing (Sherman & Turk-Browne, 2020; Turk-Browne et al., 2010), that is, the ability to anticipate stimuli and events based on current and past experience. This ability is critical for language (Levy, 2008; Trueswell, Sekerina, Hill, & Logrip, 1999) and other cognitive processes (Clark, 2013; Friston, 2010; Keller & Mrsic-Flogel, 2018). However, while words are clearly stored in declarative Long-Term Memory (after all, the point of knowing words is to "declare" them), statistical knowledge does not imply the formation of such memory representations. In fact, after exposure to sequences where some transitions are more likely than others, observers report greater familiarity with high-TP items than with low-TP items, even when they have never encountered either of these items and thus could not have memorized them (because the items are played backwards with respect to the familiarization sequence; Endress & Wood, 2011; Jones & Pashler, 2007; Turk-Browne & Scholl, 2009). Sometimes, observers even report greater familiarity with high-TP items they have *never* encountered than with low-TP items they have heard or seen (Endress, 2024b; Endress & Langus, 2017; Endress & Mehler, 2009b), suggesting that a preference for high-TP items over low-TP items does not necessarily imply that the high-TP items are encoded in declarative LTM. Further, and in line with this view, statistical learning abilities might reflect simple associative mechanisms such as Hebbian learning (Endress, 2010a, 2024; Endress & Johnson, 2021): If the representation of a syllable is still active while the next one is presented, the two syllable representations are active together and can thus form an association. These Hebbian associations will thus reflect the TPs among syllables.

While the question of whether statistical learning leads to memory for items (or chunks) is controversial (see e.g. Perruchet, 2019 vs. Endress, Slone, & Johnson, 2020 and below), statistical learning has been linked to implicit learning (e.g., Christiansen, 2018; Perruchet & Pacton, 2006; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), and is available to arguably implicit learners such as sleeping newborns, (Fló, Benjamin, Palu, & Dehaene-Lambertz, 2022). Dissociations between implicit learning and declarative memory have long been documented behaviorally (Graf & Mandler, 1984), developmentally (Finn, Kalra, Goetz, Leonard, Sheridan, & Gabrieli, 2016), and neuropsychologically (Cohen & Squire, 1980; Knowlton, Mangels, & Squire, 1996; Poldrack et al., 2001; Squire, 1992), to the extent that statistical predictions can *impair* declarative memory encoding in healthy adults (Sherman & Turk-Browne, 2020). If statistical learning operates similarly in a word-segmentation context as in other implicit learning situations, one would expect it to be dissociable from declarative Long-Term Memory.

That said, different memory systems can certainly interfere with each other during consolidation or support each others when the memories share a structure (see Robertson, 2022, for a review). However, given that the format of the representations created by statistical learning might differ from that used for linguistic stimuli (Endress & Langus, 2017; Fischer-Baum, Charny, & McCloskey, 2011; Miozzo, Petrova, Fischer-Baum, & Peressotti, 2016), it is at least an open question to what extent statistical learning supports declarative memories for words. In the General Discussion, we will discuss ways in which statistical learning might be useful for word learning even if it is dissociable from declarative memory.

In addition to possible dissociations between statistical learning and declarative memory, it is also unclear how continuous fluent speech really is. In fact, due to its prosodic organization, speech does not come as a continuous signal but rather as a sequence of smaller units (Cutler, Oahan, & van Donselaar, 1997; Nespor & Vogel, 1986; Shattuck-Hufnagel & Turk, 1996). This prosodic organization is perceived in unfamiliar languages (Brentari, González, Seidl, & Wilbur, 2011; Endress & Hauser, 2010; Pilon, 1981) and even by newborns (Christophe, Mehler, & Sebastian-Galles, 2001). It might affect the usefulness of statistical learning, because such speech cues tend to override statistical cues (Johnson & Jusczyk, 2001; Johnson & Seidl, 2009), and because statistical learning primarily operates *within* rather than across major prosodic boundaries (Shukla, Nespor, & Mehler, 2007; Shukla, White, & Aslin, 2011). As a result, the learner's segmentation task is not so much to integrate distributional information over long stretches of continuous speech, but rather to decide whether the correct grouping in prosodic groups such as "*thebaby*" is "*theba* + *by*" or "*the* + *baby*" (though prosodic groups are often longer than just three syllables; Nespor & Vogel, 1986).

### 1.2. Does statistical learning lead to declarative memories after all?

Contrary to our arguments so far, many authors suggest that statistical learning leads to declarative memories for chunks after all (Graf-Estes et al., 2007; Hay, Pelucchi, Graf Estes, & Saffran, 2011; Isbilen et al., 2020). We will fully discuss this evidence in the General Discussion, and focus here only on a particularly strong source of evidence for such views that we will refute below through neural network simulations.

Specifically, in some studies, after exposure to a statistical learning task, recognition performance is better for (statistically defined) units compared to (statistically defined) sub-units (e.g., Fiser & Aslin, 2005; Giroux & Rey, 2009; Orbán, Fiser, Aslin, & Lengyel, 2008; Slone & Johnson, 2018). In a word recognition analogy, hearing the word *hamster* makes it difficult to recognize that the first syllable of *hamster* is a word on its own (i.e., *ham*), though, in word recognition, the reduced availability of sub-units is at least partially driven by phonetic differences between syllables that are parts of words and syllables that are words on their own (e.g., Salverda, Dahan, & McQueen, 2003; Shatzman & McQueen, 2006a, 2006b; van Alphen & van Berkum, 2010).

Below, we will show that such results can be explained by a simple, memory-less Hebbian learning model, suggesting that such results do not provide any evidence for the idea that the output of statistical learning is stored in declarative long-term memory.

### 1.3. Statistical learning in continuous sequences and discrete chunks

If statistical learning mainly supports predictive processing rather than declarative memory, it might also operate predominantly under conditions that are conducive for prediction. Consequently, associations among syllables might form more easily when the syllables are part of a continuous sequence compared to when they are packaged into discrete items (e.g., through prosodic phrasing). After all, longer, continuous sequences provide more information on which predictions can be based than shorter chunks.

These are just the conditions under which statistical learning is usually explored. Specifically, participants are familiarized with *continuous* speech sequences consisting of random concatenations of non-sense "words" (or equivalent units in other modalities). As a result, syllables within words are more predictive of one another (and have higher TPs) than syllable combinations that straddle word boundaries. Following such a familiarization, (adult) participants typically complete a two-alternative forced-choice recognition task, where they have to choose between the words from speech stream and part-words. Part-words are tri-syllabic items that straddle a word boundary. For example, if *ABC* and *DEF* are two consecutive words, *BCD* and *CDE* are the corresponding part-words. Participants tend to choose words over part-words, suggesting that they are sensitive to the greater predictiveness (and TPs) of syllables within words. However, such results still leave open the question of whether participants can use this sensitivity to memorize words from fluent speech, and whether this sensitivity would be present in discrete sequences, or only in continuous sequences, especially given that continuous speech sequences are processed differently from discrete ones (e.g., Endress & Bonatti, 2016; Marchetto & Bonatti, 2015; Peña, Bonatti, Nespor, & Mehler, 2002).[2]

If statistical learning predominantly supports predictive processing, it might operate predominantly in continuous rather than discrete sequences. Conversely, discrete chunks might be more conducive for the formation of declarative memories, because such chunks have clear onsets and offsets, which appears to be a key requirement of the memory representations of linguistic stimuli (Endress & Langus, 2017; Fischer-Baum et al., 2011; Miozzo et al., 2016). The importance of discrete chunks for word learning is supported by the finding that a word-segmentation model relying just on information at the edges of discrete chunks (in the form of utterance boundaries) performed better than most other word-segmentation models (Monaghan & Christiansen, 2010), and that statistical information does not always lead to better performance when boundary information is provided (Sohail & Johnson, 2016).

Be that as it might, if statistical learning preferentially operates in continuous sequences, this would be one of numerous examples where statistical learning works better over some stimulus classes than others. The classic example is taste aversion, where rats readily associate tastes with sickness and external stimuli with pain but cannot associate taste with pain or external stimuli with sickness (Alberts & Gubernick, 1984; Garcia, Hankins, & Rusiniak, 1974; Martin & Alberts, 1979). Other examples include associations of objects with landmarks vs. boundaries (Doeller & Burgess, 2008), associations among social vs. non-social objects (Tompson, Kahn, Falk, Vettel, & Bassett, 2019), and associations among consonants vs. vowels (Bonatti, Peña, Nespor, & Mehler, 2005; Toro, Bonatti, Nespor, & Mehler, 2008).

### 1.4. The current experiments

Here, we explore the computational function of statistical learning in word-segmentation.

We first reconsider the strongest evidence purportedly supporting memory-based accounts of statistical learning, namely better recognition of entire units compared to sub-units (e.g., Fiser & Aslin, 2005; Giroux & Rey, 2009; Orbán et al., 2008; Slone & Johnson, 2018). However, we report neural network simulations showing that this evidence can be explained by memory-less Hebbian learning mechanisms. As a result, such results do not provide evidence for memory-based accounts of statistical learning.

In Experiment 1, we then ask if statistical learning leads to declarative memory of words. We exposed (adult) participants to the speech stream from Saffran, Aslin, and Newport's (1996) classic word-segmen tation experiment. The speech stream consists of four non-sense words randomly concatenated into a continuous speech sequence. As a result, TPs among syllables are higher within words than across word-boundaries. We presented the stream either as a continuous sequence (as in Saffran, Aslin, and Newport's (1996) experiments), or as a pre-segmented sequence of words, with brief silences across word boundaries. As mentioned above, these continuous vs. pre-segmented presentation modes trigger different sets of memory processes (Endress & Bonatti, 2016; Marchetto & Bonatti, 2015; Peña et al., 2002), but it is unknown if either of these processes involves declarative memory. Following this familiarization, we simply asked participants to recall what they remembered from the speech stream. In light of the finding that participants in statistical learning tasks sometimes endorse items they have never encountered (e.g., Endress & Wood, 2011; Jones & Pashler, 2007; Turk-Browne & Scholl, 2009) and can endorse them over items they *have* encountered (Endress, 2024b; Endress & Langus, 2017; Endress & Mehler, 2009b), we expected that participants would form declarative memories only after a pre-segmented familiarization.

To foreshadow our results, participants were able to recall items after a pre-segmented familiarization. However, after a continuous familiarization, 70% of the participants did not recall any (high- or low probability) items at all. Among those who did recall some items, half produced correct, high-probability items and half incorrect, low-probability items. We then verified that a prominent statistical learning model based on memories for chunks (Perruchet & Vinter, 1998) cannot explain these data.

Finally, in Experiment 2, we asked whether statistical learning operates in smaller chunks like those that might be encountered due to the prosodic organization of language, or only in longer stretches of continuous speech typical of statistical learning tasks. Participants listened to a speech sequence of tri-syllabic non-sense words. As in Experiment 1, the words were either *pre-segmented* (i.e., with a silence after each word) or continuously concatenated.

For half of the participants, both the TPs and the chunk frequency were higher between the first two syllables of the word than between the last two syllables (TPs of 1.0 vs. .33). A statistical learner should thus split triplets like *ABC* into an initial *AB* chunk followed by a singleton *C* syllable (hereafter *AB+C* pattern). For the remaining participants, both the TPs and the chunk frequency favored an *A+BC* pattern. To make the learning task as simple as possible, the statistical pattern of the words was thus consistent for each participant. Following this familiarization, participants heard pairs of *AB* and *BC* items, and had to indicate which item was more like the familiarization items. If statistical learning predominantly operates in continuous rather than pre-segmented sequences, participants should split the triplets into

[2] For example, Peña et al. (2002) familiarized participants with continuous speech streams as well as with discrete, "pre-segmented" speech streams, in which each word was followed by a brief silence. The brief silences triggered additional processes such as rule-like generalizations that were unavailable after continuous familiarizations. Critically, the rule-like generalizations observed after pre-segmented familiarizations might reflect memory processes. Endress and Mehler (2009a) suggested that the role of the silences was to act as Gestalt-like grouping cues that provided learners with the location of the word edges (i.e., onsets and offsets), and thus enabled generalizations based on those word-edges (see also Glicksohn & Cohen, 2011; Morgan, Fogel, Nair, & Patel, 2019 for other perceptual grouping effects in statistical learning). Given that the grouping cues resulted in a sequence of discrete chunks, the grouping cues might also support declarative memory processing.

their underlying chunks only after continuous but not pre-segmented familiarizations.

To preview our results, while Experiment 1 revealed that participants remember words only after listening to pre-segmented speech sequences, in Experiment 2, participants predominantly tracked TPs in continuous speech sequences, but less so in pre-segmented sequences.

## 2. Simulation 1: Does Hebbian learning provide an alternative to memory-based theories of statistical learning?

There is considerable debate about whether statistical learning leads to memory for recurring chunks (e.g., Endress et al., 2020; Goodsitt, Morgan, & Kuhl, 1993; Perruchet, 2019; Swingley, 2005; Thiessen, 2017), and some empirical results seem to support this idea.

While most of these results have alternative interpretations (see General Discussion), there is one research tradition that appears to provide strong evidence in favor of a memory-based theory of statistical learning.

Specifically, in some studies, recognition performance is better for (statistically defined) units compared to (statistically defined) sub-units (e.g., Fiser & Aslin, 2005; Giroux & Rey, 2009; Orbán et al., 2008; Slone & Johnson, 2018). In a word recognition analogy, hearing the word *hamster* makes it difficult to recognize that the first syllable of *hamster* is a word on its own (i.e., *ham*), though, in word recognition, the reduced availability of sub-units is at least partially driven by phonetic differences between syllables that are parts of words and syllables that are words on their own (e.g., Salverda et al., 2003; Shatzman & McQueen, 2006a, 2006b; van Alphen & van Berkum, 2010).

Similar effects are observed in statistical learning in both vision and audition. For example, the *AB* part of an *ABC* unit is harder to recognize than a complete *CD* unit, which would suggest that the entire units are stored in memory. We now provide simulation results suggesting that such results are compatible with a memory-less Hebbian learning mechanism, but discuss this issue separately for sequential, auditory sequences and simultaneously presented visual shapes as the arguments are somewhat different.

### 2.1. Units vs. sub-units in audition

As mentioned above, most statistical learning results can be explained by simple Hebbian learning: If the representation of a syllable is still active while the next one is presented, the two syllable representations are active together can thus form an association. An implementation of this idea is provided in models such as Endress and Johnson's (2021). In their neural network model, neurons are connected through both excitatory and inhibitory connections, where only the excitatory connections undergo Hebbian learning. After learning, when *B* (from *ABC*) is activated, it will excite (and inhibit) both *A* and *C* in turn. Critically, the excitatory connections between *A* and *C* are weaker than those between *A* and *B* and those between *B* and *C* (since there is less temporal overlap between their activations, and thus less Hebbian learning). This idea is illustrated in Fig. 1. After an (external) activation of neuron *A* (top), excitatory connections as well as external input to *B* will activate both *B* and *C* (bottom). Depending on the balance of excitation and inhibition between *A* and *C*, the net input from *C* to *A* might thus be inhibitory on the next time step. In contrast, in complete two-item units, there is no extra item like *C* that could reduce the activation within the unit due to inhibition.

We now illustrate this point, using Endress and Johnson's (2021) model to simulate one of the first experiments showing better recognition of units compared to units (Giroux & Rey, 2009). In their experiment, participants were presented with streams consisting of two three-syllable words and four two-syllable words. After such a familiarization, Giroux and Rey (2009) found better recognition for sub-units (i.e., two syllables from a three-syllable word) than for units (i.e., entire two-syllable words).

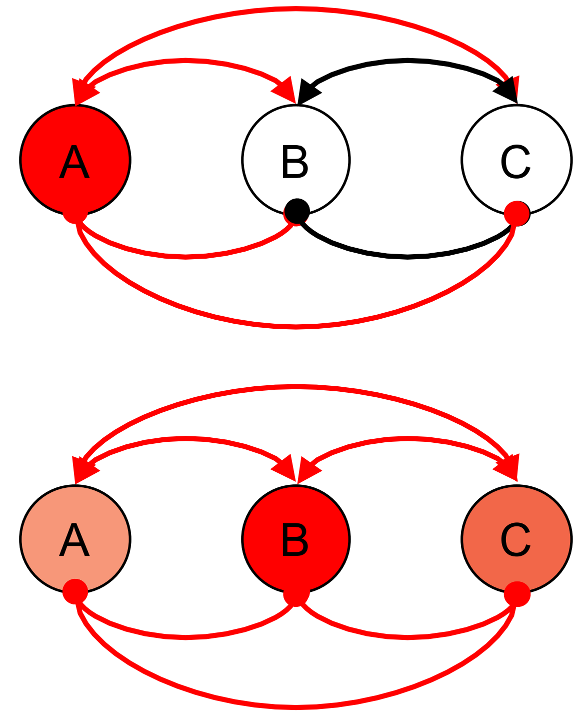


**Fig. 1.** After an (external) activation of the neuron *A* (top), excitatory connections as well as external input to *B* will activate both *B* and *C* (bottom). Depending on the balance of excitation and inhibition between *A* and *C*, the net input on from *C* to *A* might thus be inhibitory on the next time step.

The model is a fully connected network where all neurons send both excitatory and inhibitory input to all other units. Their activations also decays over time. Critically, excitatory connections are turned using a Hebbian learning rule.

In our simulations, we randomly concatenated these words into familiarization streams with 143 occurrences of each word (matching Giroux and Rey's (2009) familiarization). We then presented the network with test items (see below) and recorded the total network activation while each item was presented, using the total activation as a measure of the network's familiarity with the test item. We tested the network for different decay rates ($\Lambda$ in Endress & Johnson, 2021) and interference rates (*B* in Endress & Johnson, 2021). The cycle of familiarization and test was repeated 100 times for each parameter set, representing 100 simulated participants.

To compare the network's familiarity with two-syllable units and two-syllable sub-units, we created normalized difference scores $d = \frac{\text{Unit}-\text{Sub-unit}}{\text{Unit}+\text{Sub-unit}}$. We evaluated these difference scores against the chance level of zero using Wilcoxon tests.

As shown in Fig. 2, when averaging across trials comparing two-syllable units to *AB* and *BC* sub-units, there was a significant preference for units for most parameter sets (except for some simulations with low inhibition rates). A simple Hebbian network can thus account for better recognition of units compared to sub-units.

However, as shown in Fig. 3, while units were systematically preferred over *AB* sub-units for most parameter values, *BC* sub-units were sometimes preferred for very low or very high interference rates. Be that as it might, the current results clearly demonstrate that a simple Hebbian network can account for the preference for units over sub-units, though the level of inhibition might need to be adequate.

To support our contention that the preference for units over sub-units might arise from the interplay between learning (and thus excitation) and inhibition, Fig. 4 shows the weights between different pairs of neurons after learning. As suggested above, the connection between *A* and *C* in a three-syllable *ABC* unit is generally weaker than the

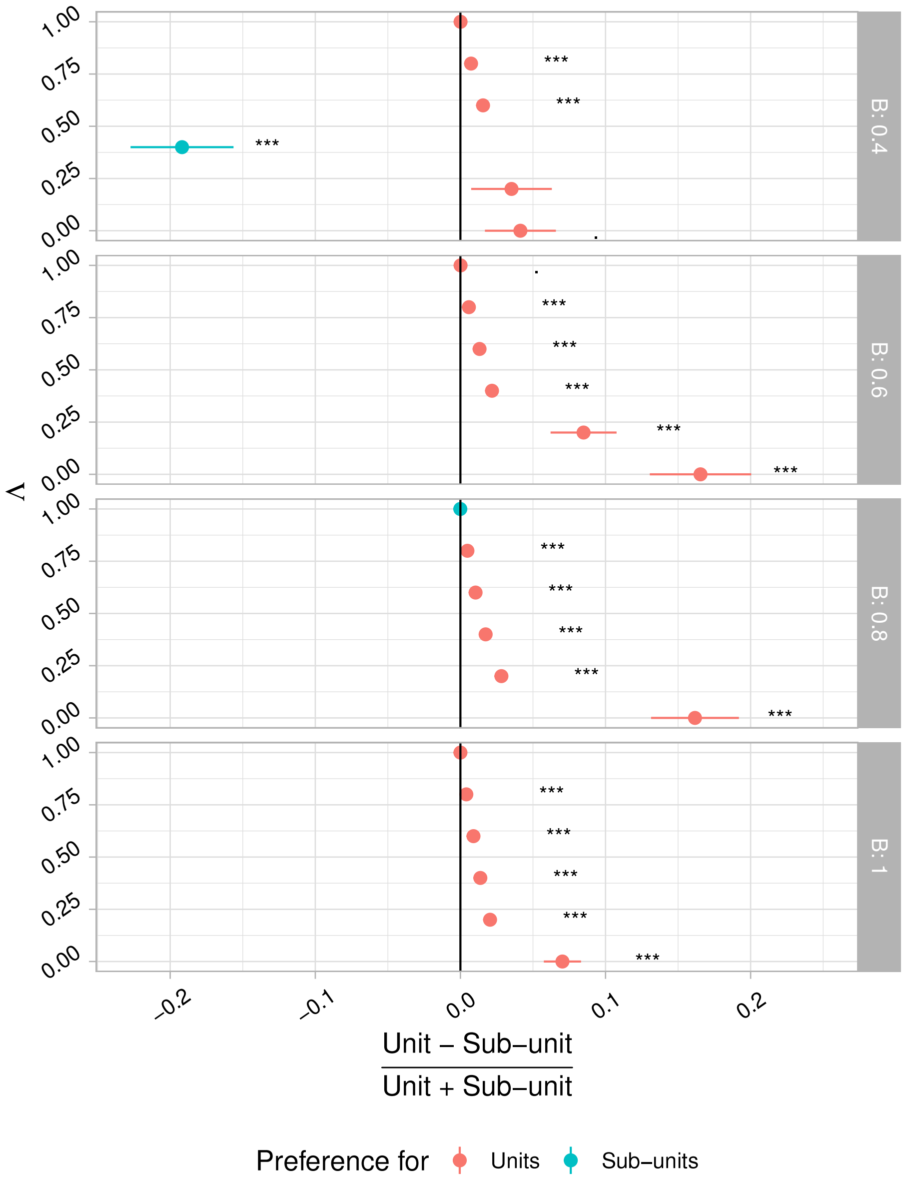


**Fig. 2.** Normalized average difference scores of network activations after presentation of entire two-syllable units and different types of two-syllable units (i.e., AB and BC from ABC units), as a function of the forgetting rate (y axis) and the interference rate (facets in rows). As in Giroux and Rey (2009), we do not separate *AB* and *BC* sub-units. Positive values indicate stronger activations for units. Significance stars reflect a Wilcoxon test against the chance level of zero. Units generally elicit greater activation compared to the average of AB and BC sub-units. Significance labels: ***: ≤ 0.001; **: ≤ 0.01; *: ≤ 0.05; .: ≤ 0.1.

other connections, and often substantially smaller than the interference rate. Depending on the parameter values, (second order) activation of *C* might thus partially suppress activation in *AB* sub-units, and activation of *A* might suppress activation in *BC* sub-units. However, the exact computational mechanisms, as well as the differences in behavior between *AB* and *BC* sub-units deserve further investigation. For the current purposes, we just conclude that the fact that a simple Hebbian learning model can account for a preference for units over sub-units demonstrates that such results do not provide evidence that units have been placed in memory, and thus do not license the conclusion that the units are stored as chunks in memory.

### 2.2. Units vs. sub-units in vision

The simulations reported above suggest that a simple Hebbian network can account for the preference for units over sub-units when items are presented sequentially (though the level of inhibition might need to be adequate). As a result, such results do not provide evidence that statistical learning leads to memory for chunks.

There is also evidence that units are easier to recognize than sub-units for simultaneously presented shapes in vision (e.g., Fiser & Aslin, 2005; Orbán et al., 2008). In such experiments, shape combinations are presented simultaneously, leading to patterns of spatial statistical regularities.

However, it is unclear how reliable such effects are. For example, Fiser and Aslin (2005) observed better recognition of units in their Experiments 1 and 4, but not in their Experiment 5. Further, when presenting shapes in a sequence rather than simultaneously, Slone and Johnson (2015) also failed to find evidence for better recognition of units in their Experiment 2, where they directly contrasted the strength of representation of units vs. sub-units.

To the extent that such findings are reliable, they are consistent with a similar explanation as the sequential case above. Presumably, the strength of associations among shapes depends on their spatial distance. Further, given the ubiquity of lateral inhibitory processes in

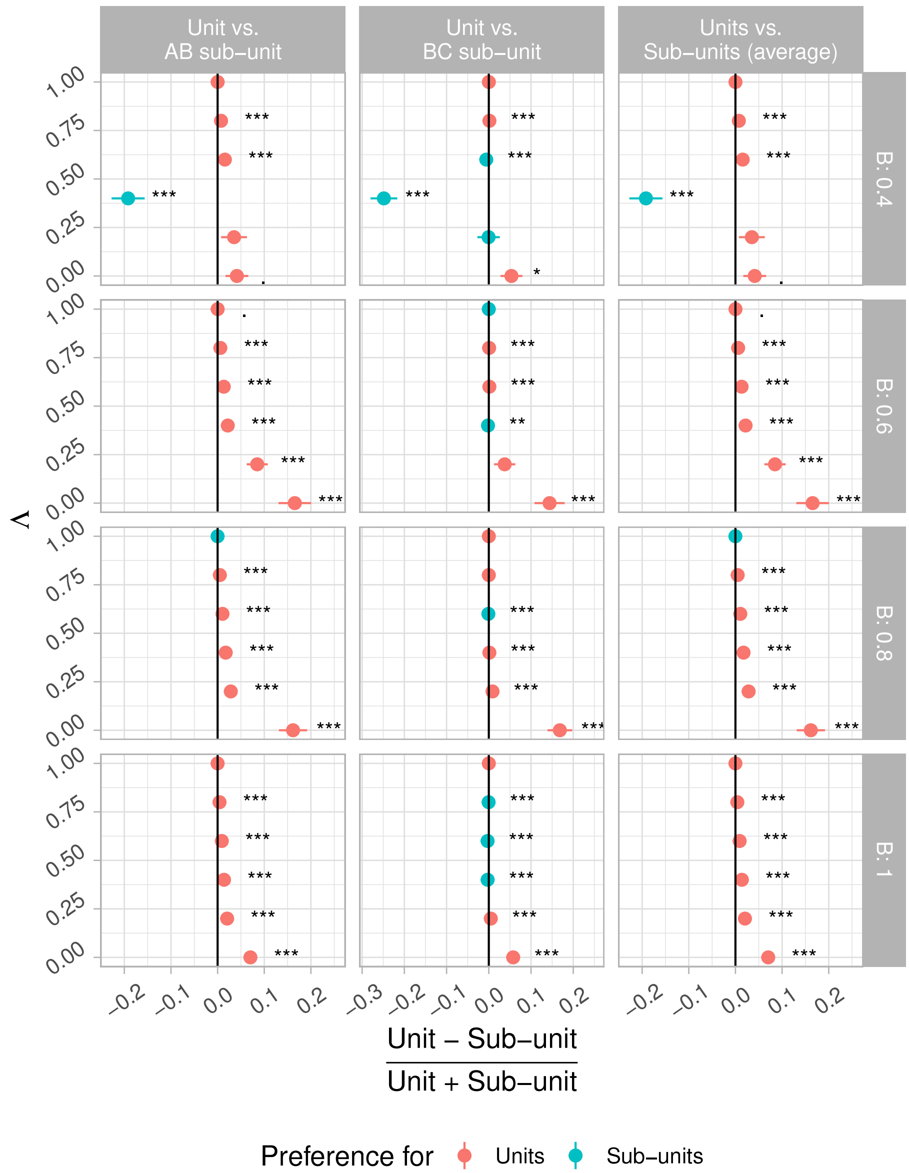


**Fig. 3.** Normalized difference scores of network activations after presentation of entire two-syllable units and different types of two-syllable units (i.e., AB and BC from ABC units), as a function of the forgetting rate (y axis) and the interference rate (facets in rows). The rightmost column is the average of the other columns reported by Giroux and Rey (2009). Positive values indicate stronger activations for units. Significance stars reflect a Wilcoxon test against the chance level of zero. Units generally elicit greater activation compared to AB sub-units and compared to the average; when compared to BC units, the sign of the difference score depends on the parameters. Significance labels: ***: ≤ 0.001; **: ≤ 0.01; *: ≤ 0.05; .: ≤ 0.1.

vision (Desimone & Duncan, 1995; Hampshire & Sharp, 2015; Kiyonaga & Egner, 2016), one would expect spatial inhibitory processes to take place in statistical learning tasks as well. As a result, one would expect a Hebbian-like model like the one above to reproduce better recognition of visually presented units compared to sub-units, though the temporal organization in the model above would need to be replaced with some spatial organization.

Better recognition for units compared to sub-units can thus be explained by simple Hebbian processes in the absence of the creation of memories for these units. However, we will now suggest further alternative interpretations of a preference for units over sub-units.

### 2.3. Further alternative explanations of a preference for units over sub-units

In the case of *sequential* statistical learning tasks, results that units are easier to recognize than sub-units have another mutually non-exclusive alternative explanation on top of the Hebbian explanation above. This explanation is based on predictive processing. If *C* is strongly associated with *AB*, hearing an *AB* fragment *during test* might lead to a prediction error because participants expect to hear *C* (or *A* for backward predictions after hearing *BC*) even when they have no memory representation of the entire *ABC* chunk. In contrast, in entire units, there is no such prediction error. This interpretation is in line with the classic finding that tasks such as stem completion do not require declarative LTM (Graf & Mandler, 1984). *Mutatis mutandis*, participants might make predictions in test items, without any units having been placed in memory, and these predictions might affect their familiarity judgments.

In the case of *spatial* statistical learning, attentional processes provide a further alternative explanation in terms of the preference for units over sub-unit. This account relies on the spatial regions attended by participants. In unpublished results, we presented participants with simultaneously presented shape combinations, and then tested for recognition of entire units or of sub-units. We found better recognition of units than of sub-units, but only when these sub-units

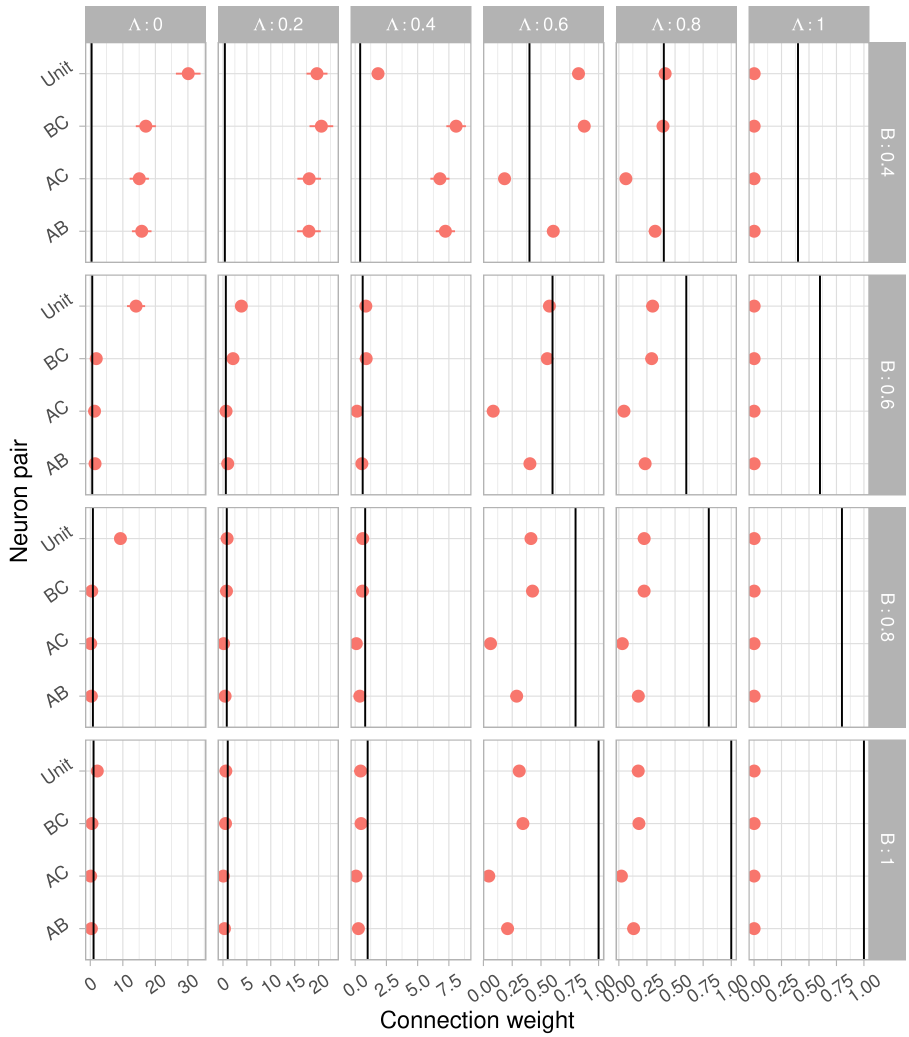


**Fig. 4.** Connection weights between different pairs of neurons as a function of the forgetting rate (columns) and the interference rate (rows). The figure shows connection weights within a trisyllabic unit (ABC) and a bisyllabic unit (Unit). The black line represents the interference rate. The A-C connection is generally smaller than the other connections, and often substantially smaller than the interference rate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

were located in parts of the display that do not attract attention. In contrast, when the sub-units came from salient parts of the units, recognition was as good as for units (Endress, in preparation).

Taken together, it seems reasonable to conclude that a preference for units over sub-units is not diagnostic of memory representations of the units. Rather, such results can be explained by simple and memory-less Hebbian learning mechanisms, or by the other explanations above. In Experiment 1, we thus ask directly if Statistical Learning leads to declarative memory representations.

## 3. Experiment 1: Do learners remember items in a statistical learning task?

In Experiment 1, we directly assessed whether Statistical Learning leads to declarative memories, asking if participants would remember items that occurred in a speech stream.

Adult participants listened to the artificial languages from Saffran, Aslin, and Newport's (1996) Experiment 2 with 8-months-old infants, except that, to increase the opportunity for learning the statistical structure of the speech stream, we doubled the exposure to 90 repetitions of each word.[3] The languages comprised four tri-syllabic words, with a TP of 1.0 within words and 0.33 across word boundaries. The words were presented in a continuous stream or as a pre-segmented word sequence. We ran a lab-based version of the experiment (Experiment 1a) and an online replication with a larger sample size (Experiment 1b). As the results of both experiments were similar, we present them jointly.

Following a retention interval, participants had to repeat back the words they remembered from the speech stream.[4] Lab-based participants responded vocally, while online participants typed their answers into a comment field. Finally, participants completed a recognition test during which we pitted words against part-words. Part-words are tri-syllabic items that straddle a word-boundary. For example, if *ABC* and *DEF* are two consecutive words, *BCD* and *CDE* are the corresponding part-words. If participants reliably choose words over part-words, they must be sensitive to TPs (even though such a sensitivity might arise from different mechanisms).

[3] We doubled the exposure with respect to Saffran, Aslin, and Newport's (1996) infant studies to maximize the chance of observing successful learning, given that even the experimenters found the learning task challenging with the stimuli from Saffran, Newport, and Aslin's (1996) (adult) experiment.

[4] Given that the focus of our experiments is the potential usefulness of statistical learning for placing items into declarative memory, we introduced a brief retention interval to mimic slightly longer-term retention than in typical statistical learning studies (but see e.g. Karaman & Hay, 2018; Vlach & DeBrock, 2019).

**Table 1**
Demographics of the final sample in Experiments 1 and 2. In Experiment 1a, the (lab-based) participants completed both segmentation conditions. In Experiment 2b, we conducted two independent replications with the same American English voice due to unexpected results with the British English voice in Experiment 2a.

| Sequence type | Voice | N | Females | Male | Age ($M$) | Age (range) |
|---|---|---|---|---|---|---|
| **Experiment 1a: Lab-based recall experiment** | | | | | | |
| continuous | us3 | 13 | 13 | 0 | 19.2 | 18–22 |
| pre-segmented | us3 | 13 | 13 | 0 | 19.2 | 18–22 |
| **Experiment 1b: Online recall experiment** | | | | | | |
| continuous | us3 | 56 | 18 | 38 | 30.6 | 19–71 |
| pre-segmented | us3 | 56 | 12 | 44 | 30.0 | 18–62 |
| **Experiment 2a – Lab-based segmentation experiment (British English voice)** | | | | | | |
| pre-segmented | en1 | 30 | 22 | 8 | 25 | 18–42 |
| continuous | en1 | 30 | 20 | 10 | 23.9 | 18–45 |
| **Experiment 2b – Lab-based segmentation experiment (American English voice)** | | | | | | |
| pre-segmented | us3 | 30 | 18 | 12 | 26.3 | 18–43 |
| continuous | us3 (1) | 32 | 26 | 6 | 20.1 | 18–44 |
| continuous | us3 (2) | 30 | 20 | 10 | 23.2 | 18–36 |

**Table 2**
Languages used Experiment 1. The words are the same as in Experiment 2 in Saffran, Aslin, and Newport (1996).

| L1 | L2 |
|---|---|
| pabiku | bikuti |
| tibudo | pigola |
| daropi | tudaro |
| golatu | budopa |

We also asked if a prominent chunking model of word segmentation (Perruchet & Vinter, 1998) can account for the results presented here.

### 3.1. Materials and methods

#### 3.1.1. Participants

As we had no prior expectation about the effect size, we targeted a sample of at least 30 participants for each of the conditions (i.e., continuous vs. pre-segmented × Language 1 vs. Language 2, see below) in the (laboratory-based) Experiment 1a. This number was chosen because it is realistic in the time-frame available for a third-year honors project. In the (online) Experiment 1b, we tested 50 participants per language and segmentation condition. Participants reported to be native speakers of English, but we did not further assess their English proficiency. At least in Experiment 1a, participants were most likely exposed to English from childhood, as the experiment took place in London, UK, and the experimenters did not notice any clear non-native accents.

To reduce performance differences between the pre-segmented and the continuous familiarization conditions, participants were excluded from analysis if their accuracy in the recognition test was below 50% ($N = 8$ in Experiment 1a; $N = 40$ in Experiment 1b). Given that our aim was to assess the role of statistical learning in the formation of declarative LTM representations of words, we restricted our analysis to participants who were most likely to have engaged in the statistical learning task.

Another 12 participants were excluded from Experiment 1b because parsing their productions took an excessive amount of computing time, though their productions did not seem to resemble the familiarization items in the first place.[5] In Experiment 1b, once the final sample of participants in the continuous condition was established, we randomly removed participants from the pre-segmented condition to equate the number of participants across the conditions. As a result, any differences between the continuous and the pre-segmented conditions were not just a consequence of differences in statistical power. (This was not necessary in the within-participant design of Experiment 1a.) The final sample included 26 participants in the lab-based version (Experiment 1a), and 112 in the online version (Experiment 1b). Demographic information is given in Table 1. Except for the exclusions due to excessive computing time (which we did not anticipate), the exclusion criteria were set forth prior to analysis.

#### 3.1.2. Materials

We re-synthesized the languages used in Saffran, Aslin, and Newport's (1996) Experiment 2. The four words in each language are given in Table 2. Each word was composed of three syllables, which were composed of two segments in turn. Stimuli were synthesized using the us3 (male American English) voice[6] of the mbrola synthesizer (Dutoit, Pagel, Pierret, Bataille, & van der Vreken, 1996), at a constant $F_0$ of 120 Hz and at a rate of 216 ms per syllable (108 ms per phoneme). This syllable duration is comparable to that in Saffran, Aslin, and Newport (1996) (222 ms per syllable).

During familiarization, words were presented 45 times each. We generated random concatenations of 45 repetitions of the 4 words, with the constraint that words could not occur in immediate repetition. For continuous streams, each randomization was then synthesized into a continuous speech stream (with no silences between words) using mbrola (Dutoit et al., 1996) and then converted to mp3 using ffmpeg (https://ffmpeg.org/). For pre-segmented streams, words were synthesized in isolation. Each randomization was then used to concatenate the words into a pre-segmented stream, with silences of 222 ms between words, which was then converted to mp3. Streams were faded in and out for 5 s using sox (http://sox.sourceforge.net/). For continuous streams, this yielded a stream duration of 1 min 57 s; for segmented streams, the duration was 2 min 37. Syllable transitions had TPs of 1.0 within words and $^1/_3$ across word boundaries. We created 20 versions of each stream with different random orders of words.

As the role of the silences in the pre-segmented stream was to create clearly identifiable chunks, the silence duration was chosen to result in clearly perceptible syllable groups (according to the experimenters' perception). Other investigations with pre-segmented material used

[5] When participants produce excessively long items (e.g., *takahsakakakaratatataikokokokotatakatakatakatakatakatakataka*, *matikulatatitulapapitularimatitulaatitula*), it can take our recursive parsing algorithm (see below) a substantial amount of computing time to generate all possible matches to the speech stream. When the analysis of a single participant exceeded several days of calculations, we decided to remove this participant from analysis. Critically, and as mentioned above, the productions for which this occurred did not resemble the statistically defined words in the first place.

[6] Experiment 1 was chronologically carried out after Experiment 2, but we changed the order for readability. We chose the us3 voice because the alternative en1 (British English) voice introduced artifacts in Experiment 2a.

shorter silences (e.g., Peña et al., 2002), longer ones (e.g., Endress & Mehler, 2009a; Sohail & Johnson, 2016) or natural prosodic phrasing (Seidl & Johnson, 2008; Shukla et al., 2007). Relatedly, other experiments mimicking the prosodic organization of speech used natural prosodic phrasing (Seidl & Johnson, 2008; Shukla et al., 2007) or grouped several "words" together using silences (Sohail & Johnson, 2016). In the light of Experiment 2, where we ask if statistical learning can be used to break up small prosodic groups such as "thebaby" into their underlying words (i.e., "the+baby"), we follow Peña et al. (2002) and present silences after each word instead of inducing longer groupings.

For the online Experiment 1b, the speech streams were combined with a silent video with no clear objects to increase attention to the stimuli. We used a panning of the Carina nebula, obtained from https://esahubble.org/videos/heic0707g/. The video was combined with the speech streams using the muxmovie utility.

#### *3.1.3. Apparatus*

The lab-based Experiment 1a was run using Psyscope X (http://psy.ck.sissa.it) in a quiet room. The online Experiment 1b was run on https://testable.org.

#### *3.1.4. Procedure*

*3.1.4.1. Familiarization.* Participants were informed that they would be listening to an unknown language and that they should try to learn the words from that language. The familiarization stream was presented twice, leading to a total familiarization duration of 3 min 53 for the continuous streams and 5 min 13 for the segmented streams. Participants could proceed to the next presentation of the stream by pressing a button.

In the online Experiment 1b, participants watched a video with no clear objects during the familiarization.

Following the familiarization, there was a 30 s retention interval. In both Experiment 1a and 1b, participants were instructed to count backwards from 99 in time with a metronome beat at 3s per beat. Performance was not monitored. Given that our objective was to investigate the role of statistical learning in the formation of declarative LTM representations of words, we attempted to make our memory tests at least somewhat long-term by introducing this filled retention interval.

*3.1.4.2. Recall test.* Following the retention interval, participants completed the recall test. In Experiment 1a, participants had 45 s to repeat back the words they remembered; their vocalizations were recorded using ffmpeg and saved in mp3 format. In Experiment 1b, participants had 60 s to type their answer into a comment field, during which they viewed a progress bar.

*3.1.4.3. Recognition test.* Following the recall test, participants completed a recognition test during which we pitted words against part-words. The (correct) test words for Language 1 (and part-words for Language 2) were /pAbiku/ and /tibudO/; the (correct) test words for Language 2 (and part-words for Language 1) were /tudArO/ and /pigOlA/. These items were combined into 4 test pairs.

#### *3.1.5. Analysis strategy*

As we used performance in the recognition test to restrict the analysis to those participants most likely to have engaged in statistical learning, performance in the recognition test in the final sample is not representative of the whole sample, and is thus not compared to a chance level. Therefore, we focus on the participants' recall responses.

It turned out that the written recall responses required substantial pre-processing because participants transcribed syllables using different orthographies and misperceived some phonemes, among other inconsistencies. The detailed analysis procedure is described in Supplementary Material SM1. All analytic choices were made to maximize the correspondence between the participants' responses and the syllable sequences attested in the speech stream.

In brief, the responses were first transformed using a set of substitutions rules to allow for misperceptions (e.g., confusion between /b/ and /p/) or orthographic variability (e.g., *ea* and *ee* both reflect the sound /i/).

Second, the responses were segmented into their underlying units. This was necessary because some participants separated only words by spaces, while others separated syllables by spaces, and groups of syllables (e.g., words) by other characters (e.g., commas). For example, responses such as *bee coo tee,two da ra,bout too pa* likely reflected the words *bikuti*, *tudaro* and *budopa*.

Third, we applied another set of substitution rules to allow for other misperceptions.

Finally, we selected the best matches to the familiarization stimuli. We selected these matches by (1) maximizing the length of the match and (2) minimizing the number of substitutions with respect to the original responses.

Based on these matches, we calculate a various properties of these matches (see Table S2). For readability, we will introduce these measures in the Results section. Exclusion criteria for responses with unattested syllables are given in Supplementary Material SM1.4.

In Experiment 1a, the (lab-based) participants' verbal responses were recorded and transcribed by two independent observers. Disagreements were resolved by discussion.[7] Online participants typed their responses directly into a comment box. Analysis of these responses was fully automatic.

We use likelihood ratios to provide evidence for the various null hypotheses. Following Glover and Dixon (2004), we fit the participant averages to (i) a linear model comprising only an intercept and (ii) the null model fixing the intercept to the appropriate baseline level, and evaluated the likelihood of these models after correcting for the difference in the number of parameters using the Bayesian Information Criterion.

### *3.2. Results*

#### *3.2.1. Analysis of the participants' productions*

We present the results in three steps. First, we report some general measures of the recall items to show that participants engage in the task and track TPs in both the continuous and the pre-segmented condition. Second, we ask whether participants are more likely produce words than part-words. Third, we ask whether participants know where words start and where they end.

Descriptives, comparisons to chance levels as well as comparisons between the continuous and the pre-segmented conditions are given in Table 3.

*3.2.1.1. General measures: Do participants engage in the task?* As shown in Table 3 and Figs. 5a and b, participants produced about 4 items. Neither the number of items produced nor their lengths differed across the segmentation conditions. Critically, and as shown in Table 3 and Figs. 6a and b, forward and backward TPs in the participants' responses were significantly greater than the chance level of 0.083 in both segmentation conditions. These TPs were greater in the pre-segmented condition. These TPs likely underestimate the participants' actual performance, as we included responses with unattested syllables that might reflect misperceptions (and thus lower TPs); after removing such responses, TPs in the participants' responses were about twice as large. Participants were thus clearly sensitive to the TPs in the speech stream.

We next examined the production of two-syllable chunks. Such chunks can be either high-TP chunks (if they are part of a word) or low-TP chunks (if they straddle a word boundary). For example, with two consecutive words *ABC* and *DEF*, the high-TP chunks are *AB*, *BC*, . . . , while the low-TP chunk is *CD*. As a result, two-syllable

[7] The number of disagreements can no longer be recovered.

**Table 3**

Main analyses pertaining to the productions as well as test against their chances levels in the recall phase of Experiments 1a and 1b. The $p$ value in the rightmost column reflects a Wilcoxon test comparing the continuous and the pre-segmented conditions. $N$'s refer to a total sample of 13 participants (lab-based) and 56 participants (online), respectively. Further results are given in Table S3.

| | Continuous | Pre-segmented | $p$(continuous vs. pre-segmented) |
|---|---|---|---|
| **Recognition accuracy** | | | |
| lab-based (Exp. 1a) | N = 13, $M$ = 0.615, $SE$ = 0.0476, $p$ = N/A | N = 13, $M$ = 0.923, $SE$ = 0.0455, $p$ = N/A | N/A |
| online (Exp. 1b) | N = 56, $M$ = 0.737, $SE$ = 0.029, $p$ = N/A | N = 56, $M$ = 0.946, $SE$ = 0.014, $p$ = N/A | N/A |
| **Number of items** | | | |
| lab-based (Exp. 1a) | N = 13, $M$ = 4.23, $SE$ = 0.756, $p$ = 0.0016 | N = 13, $M$ = 4.23, $SE$ = 0.818, $p$ = 0.0015 | 0.812 |
| online (Exp. 1b) | N = 56, $M$ = 3.8, $SE$ = 0.332, $p$ = 6.83e−11 | N = 56, $M$ = 3.16, $SE$ = 0.235, $p$ = 6.11e−11 | 0.226 |
| **Number of syllables/item** | | | |
| lab-based (Exp. 1a) | N = 13, $M$ = 3.78, $SE$ = 0.421, $p$ = 0.0016 | N = 13, $M$ = 2.97, $SE$ = 0.02, $p$ = 0.0007 | 0.026 |
| online (Exp. 1b) | N = 56, $M$ = 2.65, $SE$ = 0.103, $p$ = 5.61e−11 | N = 56, $M$ = 2.95, $SE$ = 0.04, $p$ = 3.41e−12 | < 0.001 |
| **Forward TPs** | | | |
| lab-based (Exp. 1a) | N = 13, $M$ = 0.301, $SE$ = 0.07, $p$ = 0.011 | N = 13, $M$ = 0.634, $SE$ = 0.092, $p$ = 0.00159 | 0.006 |
| online (Exp. 1b) | N = 56, $M$ = 0.383, $SE$ = 0.0385, $p$ = 1.42e−08 | N = 56, $M$ = 0.576, $SE$ = 0.0472, $p$ = 6.82e−10 | 0.003 |
| **Backward TPs** | | | |
| lab-based (Exp. 1a) | N = 13, $M$ = 0.301, $SE$ = 0.0702, $p$ = 0.0107 | N = 13, $M$ = 0.634, $SE$ = 0.092, $p$ = 0.00159 | 0.006 |
| online (Exp. 1b) | N = 56, $M$ = 0.383, $SE$ = 0.0385, $p$ = 1.42e−08 | N = 56, $M$ = 0.576, $SE$ = 0.0472, $p$ = 6.82e−10 | 0.003 |
| **Proportion of High-TP chunks among High- and Low-TP chunks** | | | |
| lab-based (Exp. 1a) | N = 4, $M$ = 0.75, $SE$ = 0.289, $p$ = 0.424 (vs. 0.5); 0.85 (vs. $^2/_3$) | N = 12, $M$ = 1, $SE$ = 0, $p$ = 0.000627 (vs. 0.5); 0.000627 (vs. $^2/_3$) | 1.000 |
| online (Exp. 1b) | N = 38, $M$ = 0.752, $SE$ = 0.0575, $p$ = 0.000246 (vs. 0.5); 0.0163 (vs. $^2/_3$) | N = 45, $M$ = 0.967, $SE$ = 0.0249, $p$ = 2.53e−10 (vs. 0.5); 2.16e−09 (vs. $^2/_3$) | < 0.001 |
| **Proportion of words among words and part-words (or concatenations thereof)** | | | |
| lab-based (Exp. 1a) | N = 7, $M$ = 0.321, $SE$ = 0.153, $p$ = 0.322 (vs. 0.5); 0.798 (vs. $^1/_3$) | N = 12, $M$ = 1, $SE$ = 0, $p$ = 0.000627 (vs. 0.5); 0.000627 (vs. $^1/_3$) | 0.034 |
| online (Exp. 1b) | N = 17, $M$ = 0.588, $SE$ = 0.127, $p$ = 0.484 (vs. 0.5); 0.019 (vs. $^1/_3$) | N = 39, $M$ = 1, $SE$ = 0, $p$ = 4.46e−10 (vs. 0.5); 4.46e−10 (vs. $^1/_3$) | < 0.001 |
| **Proportion of items with correct initial syllables** | | | |
| lab-based (Exp. 1a) | N = 13, $M$ = 0.333, $SE$ = 0.105, $p$ = 0.856 (vs. $^1/_3$); 0.481 (vs. 0.375) | N = 13, $M$ = 0.809, $SE$ = 0.0694, $p$ = 0.00186 (vs. $^1/_3$); 0.00209 (vs. 0.375) | 0.016 |
| online (Exp. 1b) | N = 56, $M$ = 0.446, $SE$ = 0.0472, $p$ = 0.0521 (vs. $^1/_3$); 0.25 (vs. 0.375) | N = 56, $M$ = 0.727, $SE$ = 0.045, $p$ = 9.41e−09 (vs. $^1/_3$); 5.1e−08 (vs. 0.375) | < 0.001 |
| **Proportion of items with correct final syllables** | | | |
| lab-based (Exp. 1a) | N = 13, $M$ = 0.456, $SE$ = 0.125, $p$ = 0.5 (vs. $^1/_3$); 0.525 (vs. 0.375) | N = 13, $M$ = 0.818, $SE$ = 0.0829, $p$ = 0.00222 (vs. $^1/_3$); 0.00278 (vs. 0.375) | 0.025 |
| online (Exp. 1b) | N = 56, $M$ = 0.403, $SE$ = 0.0514, $p$ = 0.38 (vs. $^1/_3$); 0.815 (vs. 0.375) | N = 56, $M$ = 0.721, $SE$ = 0.0532, $p$ = 4.13e−08 (vs. $^1/_3$); 1.8e−07 (vs. 0.375) | < 0.001 |

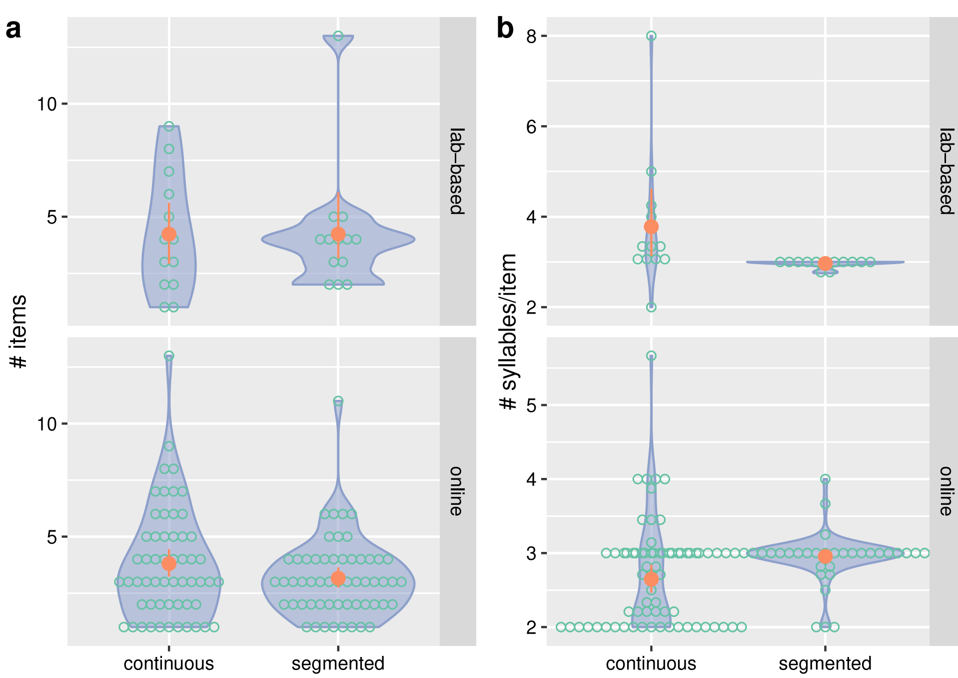


**Fig. 5.** Number of items produced and number of syllables per item in the recall phase of Experiments 1a (top) and 1b (bottom).

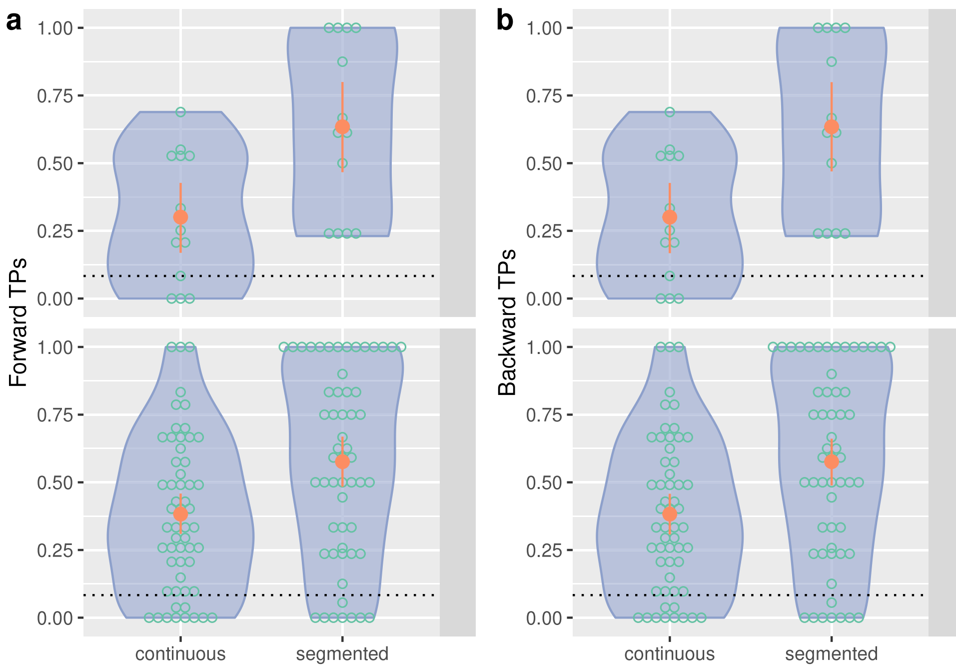


**Fig. 6.** Forward and backward TPs in the participants' productions in the recall phase of Experiments 1a (top) and 1b (bottom). The dotted line represents the chance level for a randomly ordered syllable sequence.

items have a $^2/_3$ probability of being a high-TP chunk. As shown in Fig. 7b, the proportion of high-TP among chunks high- and low-TP chunks exceeded chance in both the pre-segmented condition and the continuous condition in Experiment 1b (though not in the continuous condition of Experiment 1a), with a significantly higher proportion in the pre-segmented versions. These results thus confirm that participants are sensitive to TPs or high frequency chunks (which are confounded in the current design).

*3.2.1.2. Are participants more likely to produce words rather than part-words?* We now turn to the question of whether a sensitivity to TPs implies memory for words. We address this issue in two ways, by using the traditional contrast between words and part-words and by turning to the question at the heart of word segmentation — do participants know where words start and where they end?

The traditional analysis of word segmentation experiments relies on the contrast between words and part-words. As mentioned above, part-words are tri-syllabic items that straddle a word-boundary. We thus calculated the proportion of words among words and part-words recalled by the participants.

As shown in Table 3 and in Fig. 7a, the proportion of words among words and part-words was close to 100% in the pre-segmented conditions, but did not differ from 50% in the continuous conditions. This difference between the segmentation conditions was statistically significant. Likelihood ratio analysis suggests that, in the continuous condition of Experiment 1b, participants were 3.2 times more likely to perform at 50% than to perform at a level different from chance; in Experiment 1a, the likelihood ratio was 1.2. An alternative chance level is $^1/_3$. In fact, if participants faithfully produce trisyllabic sequences from the stream, they can start the sequences on the first, second or third syllable of a word, but only the first possibility yields a word rather than a part-word. As a result, if participants initiate their productions with a random syllable, a third of their productions should be words. Using this chance level, Table 3 and Fig. 7a show that, in the continuous conditions, the proportion of words among words and part-words differs from chance in Experiment 1b, but not in Experiment 1a.

However, either chance level drastically overestimates the participants' performance. As shown in Fig. 8 and Table S3, 70% of the participants in the continuous condition of Experiment 1b (46% in Experiment 1a) produced neither words nor part-words. Further, among those who produced words or part-words, half produced words and half produced part-words. As a result, it seems reasonable conclude that a familiarization with a continuous speech stream does not lead to declarative memories in the overwhelming majority of the participants.

This conclusion is also supported by noting that the distributions in the continuous conditions are bimodal, with some participants producing only words, and others producing only part-words (see Fig. 7a). Such a behavior can arise if participants pick a syllable as their starting-point, and segment the rest of the stream accordingly. If they happen to pick a word-initial syllable, they will produce only words; if they pick the second or the third syllable of a word, all subsequent items will be part-words.

Assuming that the number of participants producing words vs. part-words is binomially distributed, we calculated the likelihood ratio of a model where learners identify word boundaries (and should produce words with probability 1), and a model where they track TPs and initiate productions at random positions (and should produce words with a probability of $^1/_3$). As shown in SM4, the likelihood ratio in favor of the first model is $3^{N_W}$ if participants produce no part-words (i.e., after a pre-segmented familiarization), where $N_W$ is the number of participants producing words; otherwise, the likelihood ratio in favor of the second model is infinity. Given that the overwhelming majority of participants produce words only after a pre-segmented familiarizations, these results thus suggest that, despite their ability to track TPs, participants initiate productions at random positions in the sequence, and thus do not remember statistically defined words.

However, as shown in Fig. 8, even these results might be misleading because, in the continuous conditions, most participants produced neither words *nor* part-words. (In the pre-segmented condition, most participants produce at least one word, with an average of 1.26.) We thus turn to the question of whether participants know where words start and end, asking if participants produce correct initial and final syllables.

*3.2.1.3. Do participants know where words start and where they end?* If participants use statistical learning to remember words, they should

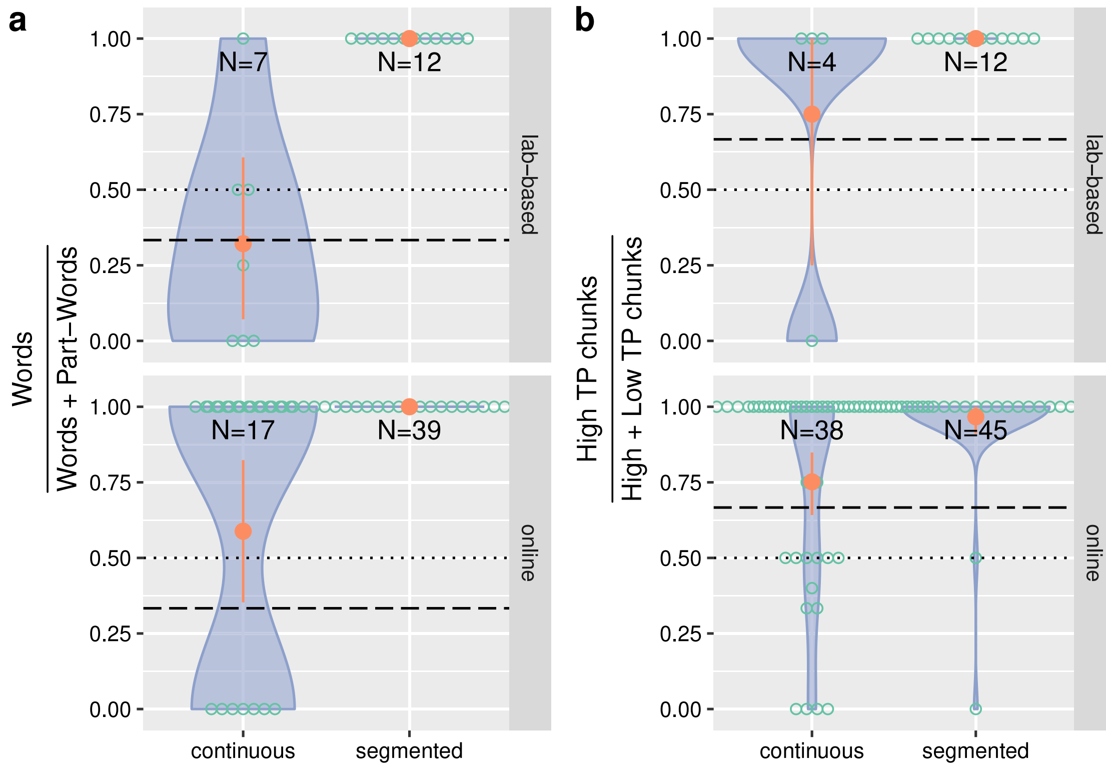


**Fig. 7.** Analyses of the participants' productions in the recall phase of Experiments 1a (top) and 1b (bottom). (a) Proportion of words among words and part-words. The dotted line represents the chance level of 50% in a two-alternative forced-choice task, while the dashed line represents an alternative chance level of 33%. (b) Proportion of high-TP chunks among high- and low-TP chunks. The dashed line represents the chance level of 66% that an attested 2 syllable-chunk is a high-TP rather than a low-TP chunk. N's reflect the numbers of participants producing words or part-words out of a total of 13 (Exp. 1a) and 56 (Exp. 1b), respectively.

know where words start and where they end. In contrast, if they just track TPs, they should initiate their responses with random syllables.

However, for this analysis, the applicable chance level is somewhat unclear. On the one hand, one can calculate the chance level by assuming that the participants' productions are not influenced by their knowledge of TPs. As there are four words with one correct initial and final syllable each, and 12 syllables in total, $^4/_{12} = {}^1/_3$ of the productions should have "correct" initial syllables, and, similarly, $^1/_3$ should have correct final syllables.

On the other hand, knowledge of TPs might change this chance level. Given that participants tend to produce high-TP two-syllable chunks (i.e., *AB* and *BC* rather than *CD* chunks), the actual baseline level should reflect this pattern. In fact, participants in the continuous condition produce about 75% high-TP chunks. If they initiate their productions with high-TP chunks, one would expect them to produce about 75%/2 = 37.5% of items with correct initial syllables. This chance level also applies to items with correct final syllables.

The results are shown in Table 3 and Fig. 9a and b. With the chance level of $^1/_3$, participants produced items with correct initial or final syllables at greater than chance level only in the pre-segmented conditions, but not in the continuous conditions (though the proportion of items with correct initial syllables was marginally greater than $^1/_3$ in Experiment 1b). When using the chancel level of 0.375, none of the continuous conditions yielded above chance performance. With this chance-level, in the continuous condition of Experiment 1b, the likelihood ratio in favor of the null hypothesis was 2.4 for initial syllables and 6.5 for final syllables; in Experiment 1b, the likelihood ratios were 3.3 and 2.9, respectively. Critically, only between 33% and 44% of the productions had a correct initial syllable, which is unexpected if participants knew where words start and where they end. Together with the finding that the overwhelming majority of participants produce no words at all, these results thus suggest that TPs do not allow learners to reliably detect onsets and offsets of words.

### 3.2.2. Can chunking models account for these results?

Taken together, the results of Experiment 1 suggest that participants can learn statistical information from fluent speech. However, the information they retain does not allow them to learn (statistically defined) chunks that might then be encoded as word candidates in declarative long-term memory. Rather, few participants produced any words or part-words at all, and among those participants who produced such items, less than half produced words. Further, only about a third of the participants produced items starting with word-initial syllables, while two-thirds produced items starting with word-medial or word-final syllables. Such results suggest that statistical learning does not support the very function for which it was motivated originally — to identify word boundaries in fluent speech, and thus to learn words from fluent speech.

Given the debate about whether statistical learning entails memories for chunks (see e.g. Perruchet, 2019 vs. Endress et al., 2020 and General Discussion), we illustrate the conclusion that chunking models will not produce part-words rather than words. Specifically, in SM6, we report simulations with PARSER (Perruchet & Vinter, 1998), a prominent chunking model of word segmentation, where we attempt to bias the model to prefer part-words over words (see also Endress & Langus, 2017, for related simulations). However, despite our attempt to bias the model, it never preferred part-words to words.

Given that, in our recall experiment, the majority of those participants who produced either words or part-words produced part-words, these results suggest that chunking models (or at least one rather prominent chunking model) either cannot account for the current results, or, to the extent that other chunking models might account for them, that these models learn information that does not allow them to recover word boundaries from fluent speech.

Critically, such models would also need to account for the fact that participants produce part-words even when they prefer words in a

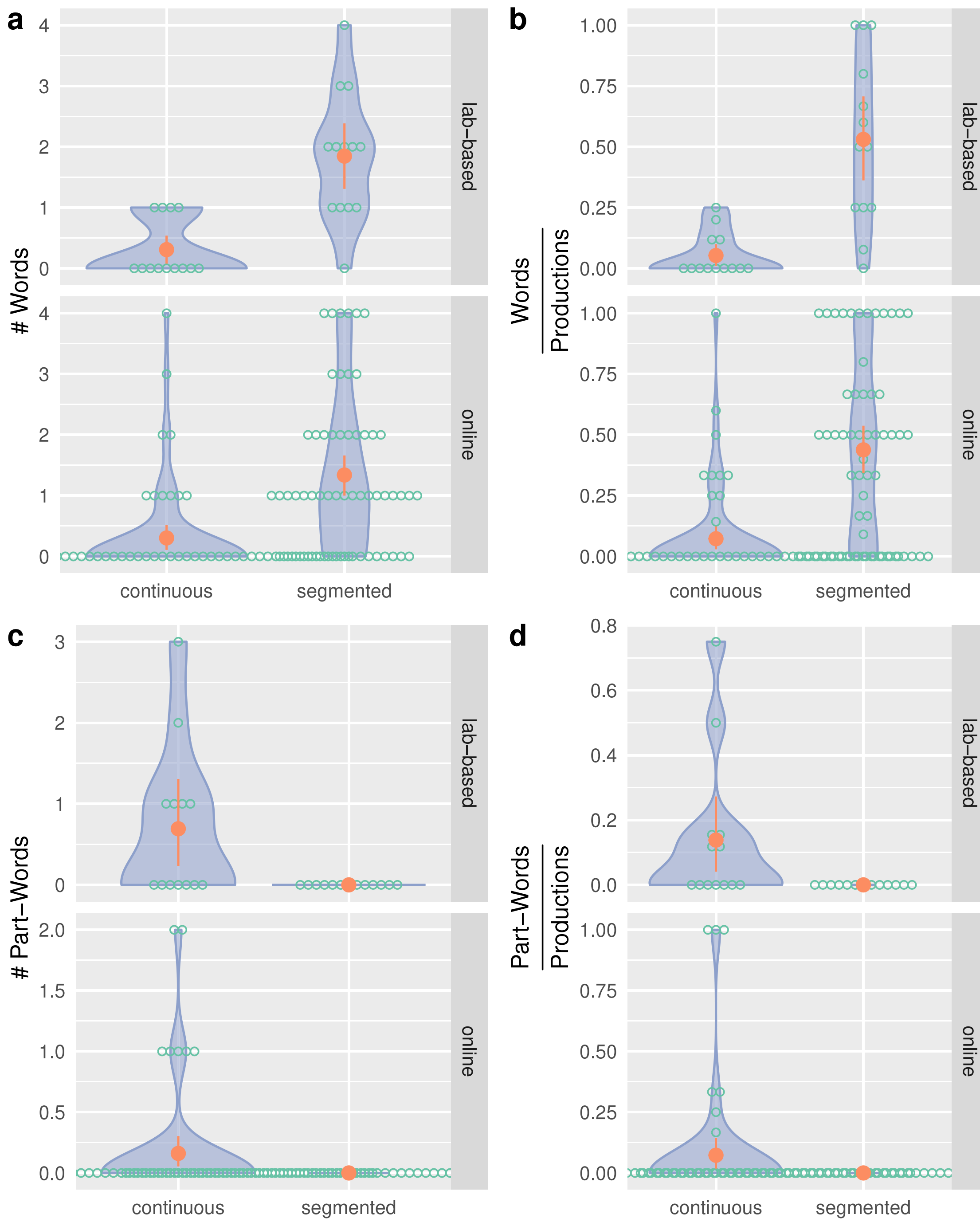


**Fig. 8.** Number and proportion (among vocalizations) of words and part-words in the recall phase of Experiments 1a (top) and 1b (bottom).

recognition test. As a result, while it might be possible to create chunking model that produce part-words (even though this would contradict their original purpose),[8] such models are unlikely to simultaneously prefer words in a recognition test. After all, the preferences of chunking models are driven by those chunks with the strongest memory representations. If these chunks happen to be words, the models will prefer words in both recognition and recall; if they are part-words, the models will prefer part-words, again in both recognition and recall. As a result, we believe that the current results are fundamentally incompatible with chunking models of statistical learning.

### 3.2.3. Relations between recall and recognition

The results so far suggest that the information extracted in statistical learning tasks does not allow participants to identify word boundaries. Further, the pattern of performance is unlikely to be explained by chunking models of word segmentation. As mentioned above, such models are driven by the memory strength of those chunks they happen to have memorized. As a result, even if it is possible to bias such models to prefer low-probability items, it is unclear how such models could prefer words over part-words in a recognition test (and thus have stronger memory traces of words), and simultaneously produce

[8] For example, it is possible to add an "attentional" component that forces the model to start chunks with word-medial syllables. We are grateful to a reviewer for pointing out this possibility.

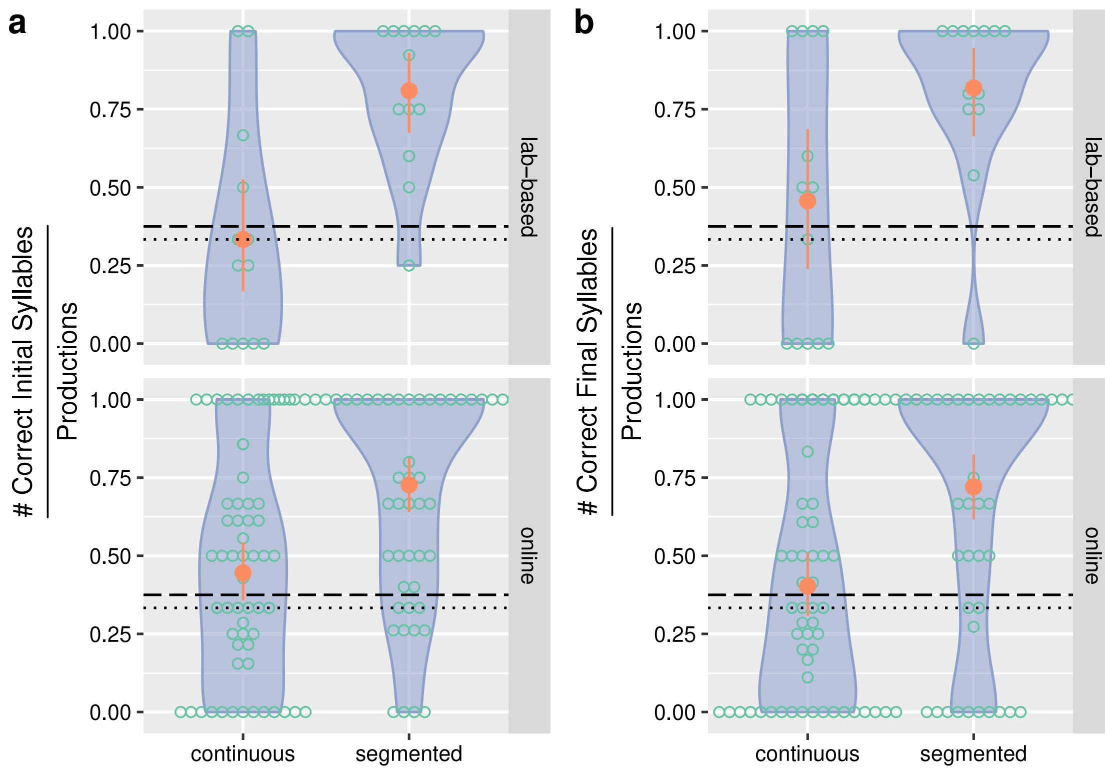


**Fig. 9.** Analyses of the participants' productions in the recall phase of Experiments 1a (top) and 1b (bottom). (a) Proportion of productions with correct initial syllables and (b) with correct final syllables. The dotted and the dashed lines represent alternative chance levels of 33% and 37.5%, respectively.

part-words rather than words in a recall test (and thus have stronger memory traces of part-words).

That being said, statistical learning performance (as measured in the recognition test) might still be related to memory for word candidates (as measured by the participants' productions), albeit indirectly. For example, and as mentioned above, participants might focus on particular individual syllables, and preferentially track statistics around those syllables they happen to focus on.

Given that attention affects statistical learning (e.g., Toro, Sinnett, & Soto-Faraco, 2005; Turk-Browne, Jungé, & Scholl, 2005), focusing on particular syllables might also direct the participants' attention and thus what they learn from the streams. For example, if participants happen to focus on word-medial or word-final syllables, they would also focus on statistically less cohesive syllable sequences as a result. Conversely, if participants happen to focus on word-initial syllables, they would also focus on statistically more cohesive syllables. This, in turn, might affect recognition performance: Those participants who produced part-words might have focused on those syllables at the beginning of part-words, and those who produced words might have focused on word-initial syllables. The syllables participants happen to focus on might be chosen randomly.

Critically, while our evidence does not allow us to decide whether participants focused on particular syllables, such views would imply that, in Experiment 1, many participants focused on other syllables than word-initial syllables, given that less than half of the participants who produced either words or part-words produced words, and that up to 82% of the participants produced no words at all — even when they preferred words to part-words in a recognition test. While we show in SM5 that recall performance is related to recognition performance, any memory-based views would thus still imply that statistical learning does not lead to memories of high probability sequences in most participants, which would make statistical learning unsuitable for word learning in turn.

### 3.3. Discussion

Experiment 1 provided the first direct test of the contents of the participants' (episodic or semantic) declarative memory after exposure to a statistical learning task. The results suggest that, even when participants successfully track statistical information, they remember familiarization items only when familiarized with a pre-segmented sequence. In contrast, when familiarized with a continuous sequence, their productions start with random syllables rather than actual word onsets. Given that the memory representations of linguistic items are based on their initial and final syllables (Endress & Langus, 2017; Fischer-Baum et al., 2011; Miozzo et al., 2016), these results thus suggest that statistical learning did not lead to the creation of declarative memory representations.

These results do not imply that statistical learning might not play a critical role in word segmentation. As mentioned above, speech is prosodically organized (Cutler et al., 1997; Nespor & Vogel, 1986; Shattuck-Hufnagel & Turk, 1996), and a learner's segmentation task is not so much to integrate distributional information over long stretches of continuous speech, but rather to decide whether the correct grouping in prosodic groups such as "*thebaby*" is "*theba* + *by*" or "*the* + *baby*". In principle, statistical learning might be well suited to this task. For example, implicit knowledge of statistical regularities might help learners acquire words more effectively once (prosodic) segmentation cues are given (but see e.g. Ngon, Martin, Dupoux, Cabrol, Dutat, & Peperkamp, 2013; Sohail & Johnson, 2016). In Experiment 2, we test this issue directly, asking whether statistical learning would help participants splitting up prosodically defined units.

## 4. Experiment 2: Is statistical learning available in both continuous and pre-segmented speech ?

Experiment 1 suggests that participants do not form declarative memory traces of words when the only available cues are statistical in nature. In contrast, they readily form declarative memories when items are pre-segmented. In Experiment 2, we ask if statistical learning allows learners to split prosodically defined units into their underlying words (though we use silences as a simplified form of prosody).

Participants listened to a speech sequence of tri-syllabic non-sense words. For half of the participants, both the TPs and the chunk frequency were higher between the first two syllables of the word than between the last two syllables. We thus expected learners to split a triplet like *ABC* into an *AB+C* pattern. For the remaining participants, both the TPs and the chunk frequency favored an *A+BC* pattern. In the *pre-segmented* condition, the words were presented separated from each other and with a silence after each word. In the *continuous* condition, they were continuously concatenated. Following this familiarization, participants heard pairs of *AB* and *BC* items and had to indicate which item was more like the familiarization items. In Experiment 2a, stimuli were synthesized with the en1 (British English male) voice, though this voice turned out to produce artifacts in the continuous stream. In Experiment 2b, stimuli were synthesized using the us3 (American English male) voice.

If, as we initially assumed, statistical learning allows learners to extract "correct" syllable groupings, they should recognize high-freq uency chunks after both continuous and pre-segmented familiarizations. In contrast, if statistical learning predominantly supports predictive processing (Sherman & Turk-Browne, 2020; Turk-Browne et al., 2010), participants should extract high frequency groupings predominantly after continuous familiarizations in the *continuous* condition.

## 4.1. Material and methods

We prepared two versions of Experiment 2, differing in the voice used to synthesize the stimuli. In Experiment 2a, we used a British English male (en1) voice. In Experiment 2b, we used an American English male (us3) voice. Both experiments were lab-based.

### 4.1.1. Participants

Participants were recruited from the City, University London participant pool and received course credit or monetary compensation for their time. We targeted 30 participants per experiment (15 per language). This number was chosen because it is realistic in the time-frame available for a third-year honors project. Participants reported to be native speakers of English, but we did not assess their English proficiency. However, participants were most likely exposed to English from childhood, as the experiment took place in London, UK, and the experimenters did not notice any clear non-native accents in most participants and excluded the few participants with non-native accents from analysis. The final demographic information is given in Table 1. In Experiment 2a, an additional 3 participants took part in the experiment but were not retained for analysis because they were much older than the rest of the sample ($N = 3$) or because they had a noticeable non-native accent $N = 1$. In Experiment 2b, an additional six participants were excluded from analysis because they had taken part in a prior version of this experiment ($N = 4$), were much older than the rest of our sample ($N = 2$), or used their phone during the experiment or were visibly inattentive ($N = 2$).

### 4.1.2. Design

Participants were familiarized with a sequence of tri-syllabic words. In Language 1, both the TPs and the chunk frequency were higher in the bigram formed by the first two syllables than in the bigram formed by the last two syllables. As a result, a statistical learner should split a triplet like *ABC* into an initial *AB* chunk followed by a singleton *C* syllable (hereafter *AB+C* pattern). In Language 2, both the TPs and the chunk frequency favored an *A+BC* pattern. The basic structure of the words is shown in Table 4.

As a result, in Language 1, the first bigram had a (forward and backward) TP of 1.0, while the second bigram had a (forward and backward) TP of .33. In contrast, in Language 2, the first bigram had a (forward and backward) TP of .33, while the second bigram had a (forward and backward) TP of 1.0. Likewise, the initial bigrams were three times as frequent as the final ones for Language 1, while the opposite holds for Language 2.

**Table 4**
Design of Experiment 2. (Left) Language structure. (Middle) Structure of test items. Correct items for Language 1 are foils for Language 2 and vice versa. (Right) Actual items in SAMPA format; dashes indicate syllable boundaries.

| Word structure for | | Test item structure for | | Actual words for | |
|---|---|---|---|---|---|
| Language 1 | Language 2 | Language 1 | Language 2 | Language 1 | Language 2 |
| ABC | ABC | AB | BC | w3:-le-gu: | w3:-le-gu: |
| ABD | FBC | FG | GD | w3:-le-vOI | faI-le-gu: |
| ABE | HBC | HJ | JE | w3:-le-nA: | rV-le-gu: |
| FGC | AGD | | | faI-zO:-gu: | w3:-zO:-vOI |
| FGD | FGD | | | faI-zO:-vOI | faI-zO:-vOI |
| FGE | HGD | | | faI-zO:-nA: | rV-zO:-vOI |
| HJC | AJE | | | rV-b{-gu: | w3:-b{-nA: |
| HJD | FJE | | | rV-b{-vOI | faI-b{-nA: |
| HJE | HJE | | | rV-b{-nA: | rV-b{-nA: |

We asked whether participants would extract initial bigrams or final bigrams. The test items are given in Table 4.

### 4.1.3. Stimuli

Stimuli in Experiment 2a were synthesized using the *en1* (British English male) voice from mbrola (Dutoit et al., 1996). However, as discussed below, it turned out to be of relatively low quality and introduced artifacts in the data. Stimuli in Experiment 2b were synthesized using the *us3* voice (American English male) voice from mbrola (Dutoit et al., 1996).

Segments had a constant duration of 60 ms (syllable duration 120 ms) with a constant $F_0$ of 120 Hz. These values were chosen to match recordings of natural speech that were intended to be used in investigations of prosodic cues to word segmentation.

For continuous streams, a single file with 45 repetitions of each word was synthesized for each language (2 min 26 s duration). It was faded in and out for 5 s using sox (http://sox.sourceforge.net/) and then compressed to an mp3 file using ffmpeg (https://ffmpeg.org/). The stream was then presented 3 times to a participant (total familiarization duration: 7 min 17 s). The random order of the words was different for each participant.

For segmented streams, words were individually synthesized using mbrola. We then used a custom-made Perl script to randomize the words for each participant and concatenate them into a familiarization file using sox. The order of words was then randomized for each participant and concatenated into a single aiff file using sox. The silence among words was 540 ms (1.5 word durations). The total stream duration was 6 min 12s. The stream was then presented 3 times to a participant (total familiarization: 18 min 14 s).

### 4.1.4. Apparatus

The experiment was run using Psyscope X (http://psy.ck.sissa.it). Stimuli were presented over headphones in a quiet room. Responses were collected from pre-marked keys on the keyboard.

### 4.1.5. Procedure

Participants were informed that they would listen to a monologue by a talkative Martian, and instructed to try to remember the Martian words. Following this, they listened to three repetitions of the familiarization stream described above, for a total familiarization duration of 7 min 17 s (continuous stream) or 18 min 14 s (segmented stream).

Following this familiarization, participants were presented with pairs of items with an inter-stimulus interval of 500 ms, and had to choose which items was more like what they heard during familiarization. One item comprised the first two syllables of a word, and was a correct choice for Language 1. The other item comprised the last two

**Table 5**
Performance differences across familiarization conditions in Experiment 2. The differences were assessed using a generalized linear model for the trial-by-trial data, using participants, correct items and foils as random factors. Random factors were removed from the model when they did not contribute to the model likelihood.

| Term | Voice | Log odds | | | Odds ratios | | | *t* | *p* |
|---|---|---|---|---|---|---|---|---|---|
| | | Estimate | *SE* | *CI* | Estimate | *SE* | *CI* | | |
| **Pre-segmented familiarization, British English voice (Exp. 2a)** | | | | | | | | | |
| language = L2 | en1 | −0.097 | 0.441 | [−0.96, 0.767] | 0.908 | 0.400 | [0.383, 2.15] | −0.22 | 0.826 |
| **Continuous familiarization, British English voice (Exp. 2a)** | | | | | | | | | |
| language = L2 | en1 | −1.024 | 0.410 | [−1.83, −0.22] | 0.359 | 0.147 | [0.161, 0.803] | −2.50 | 0.013 |
| **Pre-segmented vs. continuous familiarization, British English voice (Exp. 2a)** | | | | | | | | | |
| language = L2 | en1 | −1.061 | 0.382 | [−1.81, −0.313] | 0.346 | 0.132 | [0.164, 0.732] | −2.779 | 0.005 |
| stream type = segmented | en1 | −0.242 | 0.360 | [−0.949, 0.464] | 0.785 | 0.283 | [0.387, 1.59] | −0.673 | 0.501 |
| language = L2 ×stream type = segmented | en1 | 0.967 | 0.508 | [−0.0292, 1.96] | 2.631 | 1.338 | [0.971, 7.13] | 1.903 | 0.057 |
| **Pre-segmented familiarization, American English voice (Exp. 2b)** | | | | | | | | | |
| language = L2 | us3 | 0.114 | 0.673 | [−1.2, 1.43] | 1.121 | 0.754 | [0.3, 4.19] | 0.170 | 0.865 |
| **Continuous familiarization (replication 1), American English voice (Exp. 2b)** | | | | | | | | | |
| language = L2 | us3 | −0.184 | 0.480 | [−1.12, 0.757] | 0.832 | 0.400 | [0.325, 2.13] | −0.383 | 0.702 |
| **Continuous familiarization (replication 2), American English voice (Exp. 2b)** | | | | | | | | | |
| language = L2 | us3 | 0.317 | 0.786 | [−1.22, 1.86] | 1.372 | 1.079 | [0.294, 6.4] | 0.403 | 0.687 |
| **Pre-segmented vs. continuous familiarization, American English voice (Exp. 2b, replication 1)** | | | | | | | | | |
| language = L2 | us3 | −0.019 | 0.558 | [−1.11, 1.07] | 0.982 | 0.547 | [0.329, 2.93] | −0.033 | 0.973 |
| stream type = segmented | us3 | −0.328 | 0.188 | [−0.696, 0.0391] | 0.720 | 0.135 | [0.499, 1.04] | −1.752 | 0.080 |
| **Pre-segmented vs. continuous familiarization, American English voice (Exp. 2b, replication 2)** | | | | | | | | | |
| language = L2 | us3 | 0.215 | 0.657 | [−1.07, 1.5] | 1.240 | 0.814 | [0.342, 4.49] | 0.327 | 0.743 |
| stream type = segmented | us3 | −0.608 | 0.244 | [−1.09, −0.13] | 0.544 | 0.133 | [0.337, 0.878] | −2.493 | 0.013 |

syllables of a word, and was a correct choice for Language 2. There were three items of each kind. They were combined into 9 test pairs. The test pairs were presented twice, with different item orders, for a total of 18 test trials.

#### 4.1.6. Analysis strategy

Accuracy was averaged for each participant, and the scores were tested against the chance level of 50% using Wilcoxon tests. Performance differences across the languages (Language 1 vs. 2) and, when applicable, familiarization conditions (pre-segmented vs. continuous) were assessed using a generalized linear mixed model for the trial-by-trial data with the fixed factors language and, where applicable, familiarization condition, as well as random slopes for participants, correct items and foils. Following Baayen, Davidson, and Bates (2008), random factors were removed from the model when they did not contribute to the model likelihood.

We use likelihood ratios to provide evidence for the null hypothesis that performance did not differ from the chance level of 50%. Following Glover and Dixon (2004), we fit the participant averages to (i) a linear model comprising only an intercept and (ii) the null model fixing the intercept to the appropriate baseline level, and evaluated the likelihood of these models after correcting for the difference in the number of parameters using the Bayesian Information Criterion.

### 4.2. Results

#### 4.2.1. Experiment 2a (British English voice)

We first report the results from Experiment 2a, using a British English voice. When the familiarization stream was pre-segmented, participants failed to split smaller utterances into their underlying components. As shown in Fig. 10 (top), the average performance did not differ significantly from the chance level of 50% when the stream was synthesized with the *en1* voice ($M = 54.26$, $SD = 25.09$), Cohen's $d = 0.17$, $CI_{.95} = 44.89, 63.63$, ns. Likelihood ratio analysis favored the null hypothesis by a factor of 3.55 after correction with the Bayesian Information Criterion. Further, as shown in Table 5, performance did not depend on the language condition.

In contrast to the common finding that humans and other animals are sensitive to TPs, our participants failed to use TPs to split pre-segmented utterances into their underlying units. We thus asked if, in line with previous research, they can track TPs units are embedded into a *continuous* speech stream. That is, participants in the continuous condition listened to the very same artificial speech stream as in the pre-segmented condition, except that the stream was continuous and had no silences between words.

Participants also failed to use TPs to segment words when the speech stream was continuous. Specifically, and as shown in Fig. 10 (top), the average performance did not differ significantly from the chance level of 50%, ($M = 48.89$, $SD = 19.65$), $t(29) = -0.31$, $p = 0.759$, Cohen's $d = 0.057$, $CI_{.95} = 41.55, 56.23$, ns, $V = 166$, $p = 0.818$. Likelihood analyses revealed that the null hypothesis was 5.22 times more likely than the alternative hypothesis after a correction with the Bayesian Information Criterion. However, as shown in Table 5, performance was much better for Language 1 than for Language 2, presumably due to some click-like sounds the synthesizer produced for some stops and fricatives (notably /f/ and /g/). These sounds likely affected grouping, and prevented participants from using statistical learning. We thus decided to replicate Experiment 2a with a different, American English voice.

#### 4.2.2. Experiment 2b (American English voice)

When the familiarization stream was pre-segmented, participants failed to split smaller utterances into their underlying components. As shown in Fig. 10 (bottom), the average performance did not differ significantly from the chance level of 50% when the stream was synthesized with the *us3* voice ($M = 51.67$, $SD = 15.17$), $V = 216$, $p = 0.307$. Likelihood ratio analysis favored the null hypothesis by a factor of 4.57 after correction with the Bayesian Information Criterion. As shown in Table 5, performance did not depend on the language condition. However, Fig. 10 also shows a clearly defined outlier. In Supplementary Information SM7, we remove participants for Experiments 2a and 2b who differ by more than 2.5 standard deviations from the condition mean. This analysis yields similar results to the unfiltered analyses.

The failure to use statistical learning to split pre-segmented units was conceptually replicated in a pilot experiment with Spanish/Catalan speakers using chunk frequency and backwards TPs as the primary cues (SM8).

As in Experiment 2a, and in contrast to the common finding that humans and other animals are sensitive to TPs, our participants failed to use TPs to split pre-segmented utterances into their underlying units. We thus asked if they could track TPs units that are embedded into a *continuous* speech stream. As in Experiment 2a, participants in

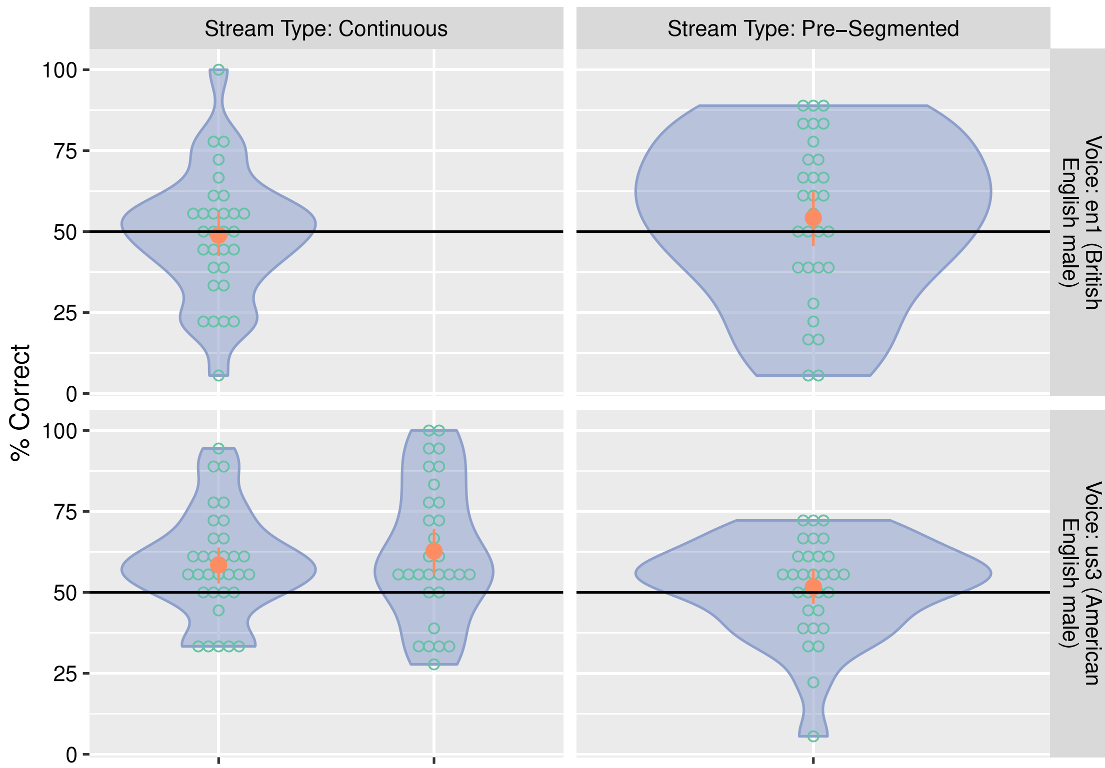


**Fig. 10.** Results of Experiment 2. Each dot represents a participant. The central red dot is the sample mean; error bars represent standard errors from the mean. The results show the percentage of correct choices in the recognition test after familiarization with (left) a continuous familiarization stream or (right) a pre-segmented familiarization stream, with a British English voice (en1, top) or an American English voice (us3, bottom). The two continuous conditions with the American English voice are replications of one another.

the continuous condition listened to the very same artificial speech stream as in the pre-segmented condition, except that the stream was continuous and had no silences between words.

As shown in Fig. 10 (bottom), when the speech stream was synthesized with the *us3* voice, the average performance differed significantly from the chance level of 50%, ($M = 58.51$, $SD = 16.21$), Cohen's $d = 0.52$, $CI_{.95} = 52.66, 64.35$, $V = 306.5$, $p = 0.02$. As shown in Table 5, performance did not depend on the language condition, and was marginally better than in the pre-segmented condition ($p = .08$).

Given the likely confound introduced by the voice used in Experiment 2a, we sought to ensure that the results of Experiment 2b would be reliable, and replicated the successful tracking of statistical information using a new sample of participants, still with the *us3* voice. As shown in Fig. 10 (bottom), the average performance differed significantly from the chance level of 50%, ($M = 62.78$, $SD = 21.35$), Cohen's $d = 0.6$, $CI_{.95} = 54.81, 70.75$, $V = 320$, $p = 0.008$. As shown in Table 5, performance did not depend on the language condition, and was significantly better than in the pre-segmented condition ($p = .013$).

Taken together, these results thus suggest that statistical learning mechanisms predominantly operate in continuous sequences, but less so in pre-segmented sequences (see also Shukla et al., 2007, 2011). Such a result is compatible with the view that statistical learning is important for predictive processing, given that continuous sequences are more conducive for prediction. In contrast, it raises doubts as to whether participants can use statistical learning mechanisms to memorize words, given that they do not seem to be able to do so in pre-segmented streams.

### *4.3. Discussion*

In Experiment 2, participants tracked statistical dependencies predominantly when they were embedded in a continuous speech stream, but not across pre-segmented chunk sequences. This finding does not contradict the results from the Experiment 1 above, where TPs were somewhat higher in the pre-segmented condition; after all, if participants faithfully recall familiarization items, the resulting TPs will be high as well.

This result is also consistent with earlier findings that statistical learning predominantly occurs within major prosodic groups, and, within these groups, predominantly at the edges of those groups (Seidl & Johnson, 2008; Shukla et al., 2007). We show that, with shorter and better separated groups, statistical learning can be weakened further, to the extent that it is no longer detectable (at least in the current experiment).

In line with results from conditioning experiments (Alberts & Gubernick, 1984; Garcia et al., 1974; Gubernick & Alberts, 1984; Martin & Alberts, 1979), statistical learning, and maybe associative learning in general, can thus be enhanced or suppressed depending on the learning situation. The enhanced statistical learning in continuous sequences is consistent with the view that statistical learning is important for predictive processing (Sherman & Turk-Browne, 2020; Turk-Browne et al., 2010), given that prediction is arguably more useful in lengthy chunks. It is also consistent with the view that statistical learning may be less important for memorizing words (or at least to break up utterances so that the underlying words can be memorized), especially given that, due to its prosodic organization, speech tends to be pre-segmented into smaller groups (Brentari et al., 2011; Christophe et al., 2001; Cutler et al., 1997; Endress & Hauser, 2010; Nespor & Vogel, 1986; Pilon, 1981; Shattuck-Hufnagel & Turk, 1996).

A possible alternative interpretation is that, in the continuous streams of Experiment 2, repeated bisyllabic items pop out (and are thus remembered), while, in the pre-segmented streams, chunking cues (in the form of silences) prevent sub-chunks from popping out. However, if repeated bisyllabic items pop out in Experiment 2's continuous streams, then repeated *trisyllabic* items (i.e., words) should pop out in Experiment 1 as well, and participants should be able to recall them as a result. As this prediction is falsified, a reasonable conclusion is that statistical learning does not make repeating elements pop out.

Conversely, the availability of chunks might make statistical learning of within-chunk regularities more difficult, especially if chunks are memorized as whole units. This possibility would also confirm that statistical learning is separable from the (declarative) mechanisms involved in memorizing chunks.

Further, while our trisyllabic items are relatively short, so are utterances in infant-directed speech. For example, infant-directed utterances have a typical duration of about 1 s (with some cross-language variability; see e.g., Fernald, Taeschner, Dunn, Papousek, de Boysson-Bardies, & Fukui, 1989; Grieser & Kuhl, 1988), with a mean utterance length of about 4 (e.g., Smolak & Weinraub, 1983; Snow, 1977; see also Martin, Igarashi, Jincho, & Mazuka, 2016). As a result, if statistical learning is difficult in shorter utterances, the utility of statistical learning for language acquisition might be reduced.

This is not to say that statistical learning can never occur in pre-segmented units. While the available statistical information does not always improve performance when chunking information is available (e.g., Sohail & Johnson, 2016), Shukla et al. (2007) showed that, when adults learners are exposed to 10-syllables chunks (defined by intonational contours), they have some sensitivity to statistical information within the chunks, though they might also use declarative memory mechanisms to remember sub-chunks (see also Endress & Bonatti, 2007; Endress & Mehler, 2009a; Endress & Wood, 2011 for additional results suggesting that statistical learning is possible within chunks, at least when the structure of the test items made the TP contrast rather salient). However, Shukla et al. (2007) also found that participants predominantly retain information at chunk edges rather than at chunk medial positions. At minimum, it is thus an empirical question to what extent statistical learning is useful for word segmentation in the short utterances infants are faced with.

## 5. General discussion

In the current experiments, we explored to what extent statistical learning can fulfill the function that is often attributed to it: Identifying word boundaries in fluent speech so that participants can learn words and, ultimately, commit them to declarative LTM.[9] In Experiment 1, we exposed (adult) participants to the speech streams from Saffran, Aslin, and Newport's (1996) classic word-segmentation experiment with infants, and asked whether they would be able to recall the words contained in these speech streams. When the speech streams were continuous, participants clearly tracked TPs in the speech streams, but we found no evidence that they had remembered any words at all. The overwhelming majority produced neither words nor part-words. Even among those who produced word or part-words, half produced words and half part-words. Further, less than half of the participants produced items starting with word-initial syllables, while the remainder produced items starting with word-medial or word-final syllables. Statistical learning thus does not appear to provide participants with the ability to identify word boundaries in fluent speech nor to remember the words to which they have been exposed. Through simulations with a prominent chunking model (Perruchet & Vinter, 1998), we confirmed that these results cannot be explained by chunking models of word segmentation. Further, and as mentioned above, the fact that participants produce part-words even when they prefer words in a recognition test is fundamentally incompatible with such models, given that the models' preferences are driven by those chunks with the strongest memory representations, in *both* recall and recognition. As a result, they should show the same preferences in both recall and recognition. In contrast, when brief silences were inserted at word boundaries, mimicking the prosodic organization of speech, participants reliably produced words.

In Experiment 2, we asked whether statistical learning operates in smaller chunks, such as those that might be encountered due to the prosodic organization of language, or only in longer stretches of continuous speech. Participants listened to a speech sequence of trisyllabic non-sense words. As in Experiment 1, the words were either *pre-segmented* (i.e., with a silence after each word) or continuously concatenated. We found that participants preferred high probability sequences only after exposure to continuous but not to pre-segmented streams, suggesting that statistical learning might be much less effective in the short and prosodically structured sequences that are typical of language acquisition (e.g., Fernald et al., 1989; Grieser & Kuhl, 1988; Martin et al., 2016; Smolak & Weinraub, 1983; Snow, 1977).[10]

Taken together, Experiments 1 and 2 suggest that statistical learning does not lead to declarative LTM representations of words, does not allow learners to identify word boundaries, and might not even operate under those conditions likely encountered during language acquisition. As a result, statistical learning and (declarative) memory might fulfill different computational functions in the process of word segmentation.

These results echo dissociations between associative learning and declarative memory (Cohen & Squire, 1980; Finn et al., 2016; Graf & Mandler, 1984; Knowlton et al., 1996; Poldrack et al., 2001; Squire, 1992), suggesting that the (cortical) declarative memory system might be independent of a (neostriatal) system for associative learning (Knowlton et al., 1996; Poldrack et al., 2001; Squire, 1992), though other authors propose that both types of memory involve the hippocampus (Ellis, Skalaban, Yates, Bejjanki, Córdova, & Turk-Browne, 2021; Schendan, Searl, Melrose, & Stern, 2003; Sherman & Turk-Browne, 2020) and different memory systems can interact during consolidation (Robertson, 2022). In line with earlier proposals (Sherman & Turk-Browne, 2020; Turk-Browne et al., 2010), we thus suggest that the computational function of statistical learning might be distinct from that of (declarative) memory encoding, and that statistical learning might be more important for predictive processing. The relative salience of these mechanisms might depend on how useful and adaptive they are for the learning problem at hand.

### 5.1. Can chunking models account for word-segmentation data?

As mentioned above, there is considerable debate about whether statistical learning leads to memory for recurring chunks (e.g., Endress et al., 2020; Goodsitt et al., 1993; Perruchet, 2019; Swingley, 2005; Thiessen, 2017). However, and as also mentioned above, there are a number of results that seem incompatible with a declarative memory theory of statistical learning.

For example, observers sometimes report greater familiarity with high-TP items than with low-TP items when they have never encountered either of them (because the items are played backwards with respect to the familiarization sequence; Endress & Wood, 2011; Jones & Pashler, 2007; Turk-Browne & Scholl, 2009). Further, observers sometimes report greater familiarity with high-TP items they have *never* encountered than with low-TP items they have heard or seen (Endress, 2024b; Endress & Langus, 2017; Endress & Mehler, 2009b), a result that has been indirectly replicated even in findings that purportedly challenge these conclusions (Perruchet & Poulin-Charronnat, 2012).[11] Such results clearly demonstrate that a sensitivity to statistical structure does not imply that the statistically favored items have been encoded in LTM. In line with this view, many statistical learning results can

[9] As mentioned above, we focus on forms of statistical learning that allow learners to track sequential dependencies among items in continuous sequences and possibly also to associate simultaneously presented items in vision. Other forms of statistical learning might well have different properties.

[10] As mentioned above, we do not propose that statistical learning is impossible within chunks, and there is evidence that statistical learning can occur within chunks under some conditions.

[11] In Perruchet and Poulin-Charronnat's (2012), as in Endress and Langus's (2017) and Endress and Mehler's (2009b) experiments, it was much harder to choose between words and unattested high-TP items than to choose between words and part-words, a result that is incompatible with current chunking models.

be explained by purely correlational, memory-less Hebbian learning mechanisms (e.g., Endress, 2024; Endress & Johnson, 2021, 2023; Verosky & Morgan, 2021).

In our view, the main evidence in favor of memory-based models of statistical learning comes in three flavors (see Endress et al., 2020, for a critical review of other evidence). First, different authors suggested that statistically favored items are preferentially encoded in memory (e.g., Graf-Estes et al., 2007; Hay et al., 2011; Isbilen et al., 2020). Such experiments generally proceed in two phases. During a statistical learning phase, participants are exposed to some statistically structured sequence. Then, they are exposed to items presented in isolation, and show some processing advantage for isolated high-probability items compared to isolated low-probability items. However, we proposed that such experiments have a two-step explanation that does not involve declarative memory (Endress & Langus, 2017). First, during the statistical learning phase, participants acquire statistical knowledge without remembering any specific items. When experimenters subsequently provide participants with *isolated* chunks, the accumulated statistical knowledge facilitates processing of the experimenter-provided chunks (e.g., due to predictive processing), without these chunks having been acquired before being supplied by the experimenter. In contrast to such indirect designs, we provide a direct measure of declarative knowledge of sequence items, and show that participants do not form declarative memories of sequence items unless the sequence is pre-segmented.

The second major source of evidence for a memory-based model for statistical learning is the observation that statistically structured sequences can elicit periodic electrophysiological activity with rhythms corresponding to word durations. For example, if words are three syllables long, a neural rhythm with a periodicity of three syllables can arise (e.g., Batterink & Paller, 2017; Buiatti, Peña, & Dehaene-Lambertz, 2009; Fló et al., 2022; Kabdebon, Pena, Buiatti, & Dehaene-Lambertz, 2015; Moser et al., 2021). At first sight, such results seem to suggest that participants must track (and thus remember) words, though not all of these authors espoused a memory-based perspective of statistical learning. However, it turns out that this periodic activity can also result from Hebbian learning mechanisms that do not place any items in memory (Endress, 2024). After all, in each word, the final syllable is maximally predictive, and thus receives more associative input from other syllables than word-initial and word-medial syllables. As a result, one would expect an activation peak on word-final syllables, and thus a rhythm with a periodicity of a word duration.

The third major source of evidence for a memory-based model of statistical learning comes from studies revealing better recognition of (statistically defined) units compared to (statistically defined) units (e.g., Fiser & Aslin, 2005; Giroux & Rey, 2009; Orbán et al., 2008; Slone & Johnson, 2018). In the word recognition analogy used above, hearing the word *hamster* makes it difficult to recognize that the first syllable of *hamster* is a word on its own (i.e., *ham*; leaving aside phonetic differences between syllables that are parts of words and syllables that are words on their own; e.g., Salverda et al., 2003; Shatzman & McQueen, 2006a, 2006b; van Alphen & van Berkum, 2010). In actual statistical learning tasks, the *AB* part of an *ABC* unit is harder to recognize than a complete *CD* unit, which would suggest that the entire units are stored in memory.

However, the simulations reported here suggest that such results are compatible with memory-less Hebbian learning mechanisms, due to the interplay between excitation and inhibition. We also provided additional alternative explanations, which suggest that the evidence for chunk-based memory due to statistical learning is much weaker than commonly believed.

Taken together, these results suggest there are several alternative explanations for better recognition of units than of sub-units that do not involve declarative memory representations of the units. Given that the relatively direct memory test presented here revealed no evidence that statistical learning leads to memory representation for recurring units, a plausible conclusion is that it does not. Potentially, statistical learning might reflect simple Hebbian learning as in Endress and Johnson's (2021) model.[12]

The conclusion that statistical learning does not lead to declarative memories of words does not imply that statistical learning has no role in word learning. For example, and as mentioned above, prior associations among syllables (or other phonological units) might facilitate the subsequent establishment of declarative memory representations for words once suitable cues become available. Pre-existing associations might be particularly useful for word learning if the initial (phonological) representations of word sounds are not yet integrated in the mental lexicon, and if this integration requires additional exposure to these words (e.g., Gaskell & Dumay, 2003; see also Viviani & Crepaldi, 2022, for evidence that lexica are acquired gradually in second language acquisition). However, most words are exceedingly rare (Yang, 2013), which, in turn, raises the question of whether sufficient exposure would be available to learners to acquire all but the most frequent words. Conversely, when potential meanings are available, people can learn words from just one or a few exposures (e.g., Aravind et al., 2018; Carey & Bartlett, 1978; Stevens, Gleitman, Trueswell, & Yang, 2017; Trueswell, Medina, Hafri, & Gleitman, 2013), suggesting that considerable exposure is not required for all forms of word learning.

Be that as it may, the current results also demonstrate that statistical learning does not allow learners to identify the beginnings and endings of words in the absence of other cues. While statistical learning might lead to helpful prior associations among syllables, other cues seem to be required to identify the (phonological) word forms that can later be consolidated.

### *5.2. Cues to word boundaries*

These current results have implications for how words can be learned from fluent speech. If learners cannot use statistical learning to encode word candidates in (declarative) memory, they need to use other cues. Possible cues include using known words as delimiters for other words (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Brent & Siskind, 2001; Mersad & Nazzi, 2012), attentional allocation to beginnings and ends of utterances (Monaghan & Christiansen, 2010; Seidl & Johnson, 2008; Shukla et al., 2007), legal sound sequences (McQueen, 1998) and universal aspects of prosody (Brentari et al., 2011; Christophe et al., 2001; Endress & Hauser, 2010; Pilon, 1981). Such cues might plausibly support declarative memories of words because they (but not transition-based associative information) are consistent with how linguistic sequences are encoded in declarative long-term memory: Linguistic sequences are encoded with reference to their first and their last element (Endress & Langus, 2017; Fischer-Baum et al., 2011; Miozzo et al., 2016). Moreover, even a fairly simple computational model attending to utterance edges yielded excellent segmentation and word-learning performance (Monaghan & Christiansen, 2010), suggesting that such cues might be useful for actual language learners as well.

[12] This conclusion does not imply that there are no explicit components to statistical learning. In fact, statistical learning is sensitive to attentional manipulations (Toro, Sinnett, & Soto-Faraco, 2005; Turk-Browne et al., 2005), and recognition performance in statistical learning tasks tends to be better when participants are more confident in their responses (e.g., Batterink, Reber, Neville, & Paller, 2015; Smalle, Daikoku, Szmalec, Duyck, & Möttönen, 2022). However, such results do not imply that statistical learning leads to declarative memory for words. For example, after familiarization with an episode of Looney Tunes, participants would presumably be highly confident in the association between Bugs Bunny and a carrot. However, this association does not imply that the Bugs Bunny–carrot combination is stored as a chunk in LTM.

### 5.3. Potential roles of statistical learning

This is no to say that statistical learning might play no implicit role in word learning even when it is not sufficient to produce memories that can be recalled. For example, and as mentioned above, associations among syllables might facilitate the establishment of declarative memories once suitable (and explicit) segmentation cues become available (Endress & Langus, 2017), and, once words are acquired, word processing is not immune to unconscious stimuli such as masked primes (e.g., Forster, 1998; Kouider & Dupoux, 2005). Statistical learning might also facilitate word learning indirectly, for example through the acquisition of phonotactic constraint that might affect word learning in turn (e.g., Friederici & Wessels, 1993; Mattys, Jusczyk, Luce, & Morgan, 1999; McQueen, 1998). However, the extent to which statistical learning supports such computations remains to be established. For example, the phonotactic regularities above can be learned by keeping track of material at utterance boundaries (Monaghan & Christiansen, 2010), and thus just using the type of cues we introduced in the presegmented conditions. However, given that the current results suggest that statistical learning and declarative memory might have separable functions, and that statistical learning does not lead to memory for words nor to knowledge of word boundaries, we believe that it is an important topic for further research to determine the role statistical learning plays in word acquisition.

## CRediT authorship contribution statement

**Ansgar D. Endress:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Maureen de Seyssel:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cognition.2025.106130.

## Data availability

Data and analysis scripts are available at https://github.com/aendress/segmenation_recall and https://figshare.com/s/dc3bf0cd35fe47156e99.

## References


Alberts, J. R., & Gubernick, D. J. (1984). Early learning as ontogenetic adaptation for ingestion by rats. *Learning and Motivation*, *15*(4), 334–359. http://dx.doi.org/10.1016/0023-9690(84)90002-X.

Aravind, A., de Villiers, J., Pace, A., Valentine, H., Golinkoff, R., Hirsh-Pasek, K., et al. (2018). Fast mapping word meanings across trials: Young children forget all but their first guess. *Cognition*, *177*, 177–188. http://dx.doi.org/10.1016/j.cognition.2018.04.008.

Aslin, R. N., & Newport, E. L. (2012). Statistical learning. *Current Directions in Psychological Science*, *21*(3), 170–176. http://dx.doi.org/10.1177/0963721412436806.

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. http://dx.doi.org/10.1111/1467-9280.00063.

Baayen, R. H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005.

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45. http://dx.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. http://dx.doi.org/10.1016/j.jml.2015.04.004.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processinghe role of consonants and vowels in continuous speech processing. *Psychological Science*, *16*(8), 451–459. http://dx.doi.org/10.1111/j.0956-7976.2005.01556.x.

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

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology. General*, *138*, 487–502. http://dx.doi.org/10.1037/a0016797.

Brent, M., & Siskind, J. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33–44. http://dx.doi.org/10.1016/s0010-0277(01)00122-6.

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. http://dx.doi.org/10.1177/0023830910388011.

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. http://dx.doi.org/10.1016/j.neuroimage.2008.09.015.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. In *Proceedings of the stanford child language conference, vol. 15* (pp. S 17–29).

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

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

Christophe, A., Mehler, J., & Sebastian-Galles, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, *2*(3), 385–394. http://dx.doi.org/10.1207/s15327078in0203_6.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. http://dx.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. http://dx.doi.org/10.1126/science.7414331.

Cutler, A., Oahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review literature review. *Language and Speech*, *40*(2), 141–201. http://dx.doi.org/10.1177/002383099704000203.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. http://dx.doi.org/10.1146/annurev.ne.18.030195.001205.

Doeller, C. F., & Burgess, N. (2008). Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(15), 5909–5914. http://dx.doi.org/10.1073/pnas.0711433105.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vreken, O. (1996). The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In *Proceedings of the fourth international conference on spoken language processingroceedings of the fourth international conference on spoken language processing, vol. 3* (pp. 1393–1396). Philadelphia.

Ellis, C. T., Skalaban, L. J., Yates, T. S., Bejjanki, V. R., Córdova, N. I., & Turk-Browne, N. B. (2021). Evidence of hippocampal learning in human infants. *Current Biology*, *31*, 3358–3364. http://dx.doi.org/10.1016/j.cub.2021.04.072, e4.

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

Endress, A. D. (2024). Hebbian learning can explain rhythmic neural entrainment to statistical regularities. *Developmental Science*, http://dx.doi.org/10.1111/desc.13487.

Endress, A. D. (2024b). Transitional probabilities outweigh frequency of occurrence in statistical learning of simultaneously presented visual shapes. *Memory and Cognition*, http://dx.doi.org/10.31234/osf.io/dbcm4.

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*(2), 247–299. http://dx.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. http://dx.doi.org/10.1002/wcs.1376.

Endress, A. D., & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology*, *61*(2), 177–199. http://dx.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*, Article 104621. http://dx.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*, Article 105290. http://dx.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. http://dx.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. http://dx.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. http://dx.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*, Article 104346. http://dx.doi.org/10.1016/j.cognition.2020.104346.

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

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501.

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. http://dx.doi.org/10.1016/j.jecp.2015.09.027.

Fischer-Baum, S., Charny, J., & McCloskey, M. (2011). Both-edges representation of letter position in reading. *Psychonomic Bulletin and Review*, *18*(6), 1083–1089. http://dx.doi.org/10.3758/s13423-011-0160-3.

Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: statistical learning of visual feature hierarchies. *Journal of Experimental Psychology. General*, *134*(4), 521–537. http://dx.doi.org/10.1038/s41598-022-08411-w.

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), http://dx.doi.org/10.25865749.

Forster, K. I. (1998). The pros and cons of masked priming. *Journal of Psycholinguistic Research*, *27*, 203–233. http://dx.doi.org/10.1023/a:1023202116609.

Friederici, A., & Wessels, J. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception and Psychophysics*, *54*(3), 287–295. http://dx.doi.org/10.3758/bf03205263.

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

Garcia, J., Hankins, W. G., & Rusiniak, K. W. (1974). Behavioral regulation of the milieu interne in man and rat. *Science*, *185*(4154), 824–831. http://dx.doi.org/10.1126/science.185.4154.824.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*, 105–132. http://dx.doi.org/10.1016/s0010-0277(03)00070-2.

Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science*, *33*, 260–272. http://dx.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. http://dx.doi.org/10.3758/s13414-010-0084-4.

Glover, S., & Dixon, P. (2004). Likelihood ratios: a simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin and Review*, *11*(5), 791–806. http://dx.doi.org/10.3758/bf03196706.

Goodsitt, J., Morgan, J. L., & Kuhl, P. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language*, *20*(2), 229–252. http://dx.doi.org/10.1017/s0305000900008266.

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. http://dx.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.tatistical segmentation and word learning. *Psychological Science*, *18*(3), 254–260. http://dx.doi.org/10.1111/j.1467-9280.2007.01885.x.

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*, *24*(1), 14–20. http://dx.doi.org/10.1037/0012-1649.24.1.14.

Gubernick, D. J., & Alberts, J. R. (1984). A specialization of taste aversion learning during suckling and its weaning-associated transformation. *Developmental Psychobiology*, *17*, 613–628. http://dx.doi.org/10.1002/dev.420170605.

Hampshire, A., & Sharp, D. J (2015). Contrasting network and modular perspectives on inhibitory control. *Trends in Cognitive Sciences*, *19*, 445–452. http://dx.doi.org/10.1016/j.tics.2015.06.006.

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.tatistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53–64. http://dx.doi.org/10.1016/s0010-0277(00)00132-3.

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. http://dx.doi.org/10.1016/j.cogpsych.2011.06.002.

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*, Article e12848. http://dx.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. http://dx.doi.org/10.1006/jmla.2000.2755.

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

Jones, J., & Pashler, H. (2007). Is the mind inherently forward looking? Comparing prediction and retrodiction. *Psychonomic Bulletin & Review*, *14*, 295–300. http://dx.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. http://dx.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. http://dx.doi.org/10.1037/xlm0000448.

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive processing: A canonical cortical computation. *Neuron*, *100*(2), 424–435. http://dx.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. http://dx.doi.org/10.1016/s0010-0277(02)00004-5.

Kiyonaga, A., & Egner, T. (2016). Center-surround inhibition in working memory. *Current Biology : CB*, *26*, 64–68. http://dx.doi.org/10.1016/j.cub.2015.11.013.

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

Kouider, S., & Dupoux, E. (2005). Subliminal speech priming. *Psychological Science*, *16*, 617–625. http://dx.doi.org/10.1111/j.1467-9280.2005.01584.x.

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

Marchetto, E., & Bonatti, L. L. (2015). Finding words and word structure in artificial speech: the development of infants' sensitivity to morphosyntactic regularities. *Journal of Child Language*, *42*(4), 873–902. http://dx.doi.org/10.1017/S0305000914000452.

Martin, L. T., & Alberts, J. R (1979). Taste aversions to mother's milk: the age-related role of nursing in acquisition and expression of a learned association. *Journal of Comparative and Physiological Psychology*, *93*, 430–445. http://dx.doi.org/10.1037/h0077568.

Martin, A., Igarashi, Y., Jincho, N., & Mazuka, R. (2016). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Cognition*, *156*, 52–59. http://dx.doi.org/10.1016/j.cognition.2016.07.015.

Mattys, S. L., Jusczyk, P. W., Luce, P., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465–494. http://dx.doi.org/10.1006/cogp.1999.0721.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–11. http://dx.doi.org/10.1016/s0010-0277(01)00157-3.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*(1), 21–46. http://dx.doi.org/10.1006/jmla.1998.2568.

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. http://dx.doi.org/10.1080/15475441.2011.609106.

Miozzo, M., Petrova, A., Fischer-Baum, S., & Peressotti, F. (2016). Serial position encoding of signs. *Cognition*, *154*, 69–80. http://dx.doi.org/10.1016/j.cognition.2016.05.008.

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. http://dx.doi.org/10.1017/S0305000909990511.

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

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

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., & Peperkamp, S. (2013). (Non)words, (non)words, (non)words: evidence for a protolexicon during the first year of life. *Developmental Science*, *16*(1), 24–34. http://dx.doi.org/10.1111/j.1467-7687.2012.01189.x.

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. http://dx.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. http://dx.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. http://dx.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. http://dx.doi.org/10.1016/j.tics.2006.03.006.

Perruchet, P., & Poulin-Charronnat, B. (2012). Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *Journal of Memory and Language*, *66*(4), 807–818. http://dx.doi.org/10.1016/j.jml.2012.02.010.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation: 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. http://dx.doi.org/10.1007/bf01068032.

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

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

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. http://dx.doi.org/10.1126/science.274.5294.1926.

Saffran, J. R., & Kirkham, N. Z (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181–203. http://dx.doi.org/10.1146/annurev-psych-122216-011805.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cueshe role of distributional cues. *Journal of Memory and Language*, *35*, 606–621. http://dx.doi.org/10.1006/jmla.1996.0032.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental LANGUAGE learning: Listening (and learning) out of the comer of your ear. *Psychological Science*, *8*(2), 101–105. http://dx.doi.org/10.1111/j.1467-9280.1997.tb00690.x.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*(1), 51–89. http://dx.doi.org/10.1016/s0010-0277(03)00139-2.

Schendan, H. E., Searl, M. M., Melrose, R. J., & Stern, C. E. (2003). An FMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron*, *37*, 1013–1025. http://dx.doi.org/10.1016/s0896-6273(03)00123-5.

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. http://dx.doi.org/10.1017/s0305000907008215.

Shannon, C. E. (1951). Prediction and entropy of printed english. *Bell System Technical Journal*, *30*(1), 50–64. http://dx.doi.org/10.1002/j.1538-7305.1951.tb01366.x.

Shattuck-Hufnagel, S., & Turk, A. E (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*(2), 193–247. http://dx.doi.org/10.1007/bf01708572.

Shatzman, K. B., & McQueen, J. M (2006a). Prosodic knowledge affects the recognition of newly acquired words. *Psychological Science*, *17*(5), 372–377. http://dx.doi.org/10.1111/j.1467-9280.2006.01714.x.

Shatzman, K. B., & McQueen, J. M (2006b). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception and Psychophysics*, *68*(1), 1–16. http://dx.doi.org/10.3758/bf03193651.

Sherman, B. E., Graves, K. N., & Turk-Browne, N. B. (2020). The prevalence and importance of statistical learning in human cognition and behavior. *Current Opinion in Behavioral Sciences*, *32*, 15–20. http://dx.doi.org/10.1016/j.cobeha.2020.01.015.

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. http://dx.doi.org/10.1073/pnas.2013291117.

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. http://dx.doi.org/10.1016/j.cogpsych.2006.04.002.

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6038–6043. http://dx.doi.org/10.1073/pnas.1017617108.

Slone, L. K., & Johnson, S. (2015). Statistical and chunking processes in adults' visual sequence learning. In D. C. Noelle, & et al. (Eds.), *Proceedings of the 37th annual conference of the cognitive science society* (pp. 2218–2223). Austin, TX: Cognitive Science Society, Paper presented at the annual meeting of the cognitive science society.

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

Smalle, E. H. M., Daikoku, T., Szmalec, A., Duyck, W., & Möttönen, R. (2022). Unlocking adults' implicit statistical learning by cognitive depletion. In *Proceedings of the national academy of sciences of the United states of america, 119*. http://dx.doi.org/10.1073/pnas.2026011119.

Smolak, L., & Weinraub, M. (1983). Maternal speech: strategy or response? *Journal of Child Language*, *10*(2), 369–380. http://dx.doi.org/10.1017/S0305000900007820.

Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language*, *4*, 1–22. http://dx.doi.org/10.1017/s0305000900000453.

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. http://dx.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. http://dx.doi.org/10.1037/0033-295x.99.2.195.

Stevens, J. S., Gleitman, L. R., Trueswell, J. C., & Yang, C. (2017). The pursuit of word meanings. *Cognitive Science*, *41*(S4), 638–676. http://dx.doi.org/10.1111/Cogs.12416.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*(1), 86–132. http://dx.doi.org/10.1016/j.cogpsych.2004.06.001.

Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 372. http://dx.doi.org/10.1098/rstb.2016.0056.

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*(2), 253–271. http://dx.doi.org/10.1037/xlm0000580.

Toro, J. M., Bonatti, L., Nespor, M., & Mehler, J. (2008). Finding words and rules in a speech stream: functional differences between vowels and consonants. *Psychological Science*, *19*, 137–144. http://dx.doi.org/10.1111/j.1467-9280.2008.02059.x.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25–34. http://dx.doi.org/10.1016/j.cognition.2005.01.006.

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. http://dx.doi.org/10.1037/0097-7403.31.1.95.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*, 126–156. http://dx.doi.org/10.1016/j.cogpsych.2012.10.001.

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. http://dx.doi.org/10.1016/s0010-0277(99)00032-3.

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. http://dx.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. http://dx.doi.org/10.1037/0096-1523.35.1.195.

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. http://dx.doi.org/10.1523/JNEUROSCI.0858-10.2010.

van Alphen, P. M., & van Berkum, J. J A (2010). Is there pain in champagne? Semantic involvement of words within words during sense-making.emantic involvement of words within words during sense-making. *Journal of Cognitive Neuroscience*, *22*, 2618–2626. http://dx.doi.org/10.1162/jocn.2009.21336.

van Moorselaar, D., & Slagter, H. A (2019). Learning what is irrelevant or relevant: Expectations facilitate distractor inhibition and target facilitation through distinct neural mechanisms. *Journal of Neuroscience*, *39*, 6953–6967. http://dx.doi.org/10.1523/JNEUROSCI.0593-19.2019.

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

Viviani, E., & Crepaldi, D. (2022). Masked morphological priming and sensitivity to the statistical structure of form-to-meaning mapping in L2. *Journal of Cognition*, *5*(30), http://dx.doi.org/10.5334/joc.221.

Vlach, H. A., & DeBrock, C. A. (2019). Statistics learned are statistics forgotten: Children's retention and retrieval of cross-situational word learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *45*(4), 700–711. http://dx.doi.org/10.1037/xlm0000611.

Yang, C. D. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6324–6327. http://dx.doi.org/10.1073/pnas.1216803110.