The specificity of sequential Statistical Learning

Ansgar D. Endress[a,1,2] **and Maureen de Seyssel**[b,c,1]

[a]Department of Psychology, City, 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 de Paris, CNRS, Paris, France

This manuscript was compiled on October 26, 2021

Learning statistical regularities from the environment is ubiquitous across domains and species (1–7). Such learning mechanisms are often remarkably tuned to ecological learning situations (8–13), and separable from declarative memory mechanisms (14–19). Statistical Learning mechanisms might be particularly critical for the earliest stages of language acquisition, notably to identify and memorize words from fluent speech (i.e., for word-segmentation; 20, 21). Here, we ask if the Statistical Learning mechanisms involved in word-segmentation are tuned to specific learning situations, and how they interact with the (declarative) memory mechanisms needed to remember words. We show that these mechanisms predominantly operate in continuous speech sequences similar to those used in prior word-segmentation experiments (1–3), but not in discrete chunk sequences, even though the latter are likely encountered during language acquisition (due to the prosodic organization of language; 22–24). Conversely, when exposed to continuous sequences in a memory recall experiment, participants are sensitive to probable syllable transitions, but, to the extent that they remember any items at all, they tend to initiate their productions with random syllables rather than with word onsets. In contrast, familiarization with discrete sequences produces reliable memories of actual, high-probability forms. This dissociation between Statistical Learning and (declarative) memory mechanisms suggests that Statistical Learning might have a specialized role when distributional information can be accumulated (e.g., for predictive processing), and that it is separable from the (declarative) memory mechanisms needed to acquire words.

Statistical Learning | Declarative Memory | Predictive Processing | Language Acquisition

1. Introduction

The ability to learn statistical regularities from the environment is remarkably widespread across species and domains (1–7), and might support a wide range of computations, especially during language acquisition (20, 21). However, such Statistical Learning abilities are also remarkably modular (9). Humans have independent statistical learning abilities in superficially similar domains, including associations of objects with landmarks vs. boundaries (10), associations among social vs. non-social objects (11) and associations among consonants vs. vowels (12, 13).

Such preferential associations abound, can evolve in just 40 generations in fruit flies (25), and likely reflect ecological constraints. For example, rats readily associate tastes with sickness and external stimuli with pain, but cannot associate taste with pain or external stimuli with sickness (8). However, taste-sickness associations (but not other associations) are blocked in a suckling context if rat pups had no exposure to solid food (26, 27), presumably because avoidance of the *only* food source is costly. Over evolutionary times, Statistical Learning mechanisms can thus become tuned to specific learning situations (though they still might be a "spandrel" (28) that evolved as a side effect of local neural processing and might undergo positive, negative or no selection in different brain pathways).

Here, we ask if the Statistical Learning mechanisms thought to be involved in learning words from fluent speech (i.e., in word-segmentation) show similar specializations, and how they relate to their presumed computational function, namely to store words in (declarative) memory. In brief, we suggest that these mechanisms are critical for predicting speech material and operate predominantly under conditions where prediction is possible. However, we also suggest that separate mechanisms are required to form (declarative) memories of the words learners need to acquire.

Speech is thought to be a continuous signal (and often perceived as such in unknown languages), and before learners can commit any words to memory, they need to learn where words start and where they end. They might rely on Transitional Probabilities (TPs) among syllables, that is, the conditional probability of a syllable σ_{i+1} given a preceding syllable σ_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 (29) showed that human adults are sensitive to such distributional information. Subsequent work demonstrated that infants and non-human animals share this ability (1–7), and that it might reflect simple associative mechanisms such as Hebbian learning (30).

Statistical Learning therefore supports predictive processing (31, 32), a critical ability for language (33, 34) and other cognitive processes (35–37). However, while words are clearly

Significance Statement

Many organisms learn statistical regularities from their environment. In humans, such Statistical Learning mechanisms may support many cognitive processes, especially language acquisition. However, while such mechanisms are often remarkably tuned to specific learning situations, their function during language acquisition is debated. Focusing on how words are learned from fluent speech, we find reliable Statistical Learning under conditions similar to earlier investigations of statistical word learning, presumably because these conditions are conducive for predicting upcoming elements. However, we do not observe Statistical Learning under conditions where words actually need to be acquired and show that it is dissociable from the memory mechanisms required to acquire words. Statistical Learning and (declarative) memory might thus have distinct and specialized functions during language acquisition.

Both authors performed research for Experiment 1 and wrote the paper. ADE performed research for Experiment 2.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: ansgar.endressm4x.org

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 them and thus could not have memorized them (because the items are played backwards with respective to the familiarization sequence; 6, 38). Sometimes, observers even report greater familiarity with high-TP items they have *never* encountered than with low-TP items they have heard or seen (39).

Dissociations between Statistical Learning and declarative memory have long been documented behaviorally (14), developmentally (15) and neuropsychologically (16–19), to the extent that statistical predictions can *impair* declarative memory encoding in healthy adults (31). If Statistical Learning operates similarly in a word-segmentation context as in other learning situations, one would expect it to be dissociable from declarative Long-Term Memory, a view that is reinforced by the suggestion that the format of the representations created by Statistical Learning differs from that used for linguistic stimuli (39, 40).

Here, we explore the computational function of Statistical Learning in word-segmentation, focusing on the conditions under which it operates and its relation to memory processes. To explore its operating conditions, we note that speech does not come as a continuous signal but rather as a sequence of smaller units due to its prosodic organization (22–24). This prosodic organization is perceived in unfamiliar languages (41–43) and even by newborns (44). It might affect the usefulness of Statistical Learning, because Statistical Learning primarily operates *within* rather than across major prosodic boundaries (45, 46). 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; 23).

In Experiment 1, we thus ask whether Statistical Learning operates in such smaller chunks, or only in longer stretches of continuous speech. In Experiment 2, we seek to elucidate the function of Statistical Learning, asking (adult) participants to recall what they remember after being exposed to the speech stream from Saffran et al.'s (1) classic experiment, again with a sequence of pre-segmented "words" or with a continuous speech stream.

2. Results

In Experiment 1, participants listened to a speech sequence of tri-syllabic non-sense words synthesized using mbrola (47). The words were either *pre-segmented* (i.e., with a silence after each word) or continuously concatenated. These continuous vs. pre-segmented presentation modes trigger different sets of memory processes (48, 49), but it is unknown if either of these processes involves declarative memory.

For half of the participants, both the TPs and the chunk frequency was higher between the 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 a triplet 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.

When the familiarization stream was pre-segmented, participants failed to split smaller utterances into their underlying components. As shown in Figure 1, the average performance did not differ significantly from the chance level of 50%, (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 S7, performance did not depend on the language condition. As shown in SI5, this failure was replicated using a second voice (*en1*, British English male). 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 (SI6).

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 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 Figure 1, 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 S7, performance did not depend on the language condition, and was marginally better than in the pre-segmented condition (p = .08).

We replicated the successful tracking of statistical information using a new sample of participants. As shown in Figure 1, 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 S7, performance did not depend on the language condition, and was significantly better than in the pre-segmented condition (p = .013).

(This result could not be replicated using a different voice (*en1*, male British English; see SI5); participants seemed to prefer specific items, irrespective of the language they had been familiarized with, presumably because the synthesizer produced click-like sounds for some stops and fricatives that likely affected syllable grouping.)

Taken together, these results thus suggest that Statistical Learning mechanisms predominantly operate in continuous sequences, but less so in pre-segmented sequences (see also 45, 46). 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.

In Experiment 2, we explored the computational function of Statistical Learning, and asked if participants would remember

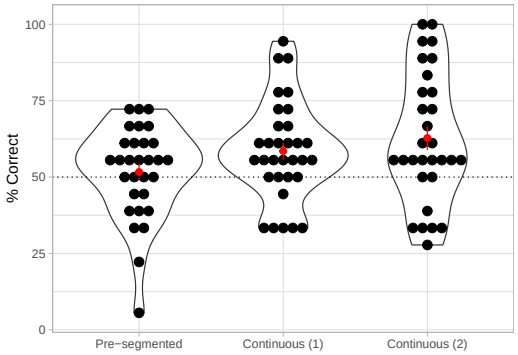

Fig. 1. Results of Experiment 1. Each dot represents a participant. The central red dot is the sample mean; error bars represent standard errors from the mean. The results show the percentage of correct choices in the recognition test after familiarization with (left) a pre-segmented familiarization stream or (middle, right) a continuous familiarization stream. The two continuous conditions are replications of one another.

the items that occurred in a speech stream. Adult participants listened to the artificial languages from Saffran et al's (1) experiments with 8-months-old infants, except that we doubled the exposure to increase the opportunity for learning the statistical structure of the speech stream. 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 both a lab-based and an online version of the experiment. Lab-based participants just listened to the speech stream, while online participants watched a video of a nebula at the same time.

Following a retention interval, participants had to repeat back the words they remembered from the speech stream. Lab-based participants responded vocally, while online participants typed their answer into a comment field. Finally, participant 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 track TPs.

In the analyses below, we removed single syllable responses (and participants who did not produce any other other items). To focus the analyses on participants who learned the statistical structure of the speech stream, we also removed participants who did not perform at least 50% during the final recognition test.

As shown in Table 1 and Figures 2a 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 1 and Figures 2c and d, forward and backward TPs in the participants' responses were significantly greater than the chance level of .083 in both segmentation conditions. 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. (TPs were somewhat higher in the pre-segmented condition. This finding does not contradict the results from the Experiment 1 above; after all, if participants faithfully recall familiarization items, the resulting TPs will be high as well.)

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 items have a 66% probability of being a high-TP chunk. As shown in Figure 3b, the proportion of high-TP among chunks high- and low-TP chunks exceeded chance in both the pre-segmented condition and the continuous condition (at least in the online version), 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).

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. 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.

As shown in Table 1 and in Figure 3a, the proportion of words among words and part-words was close to 100% in the pre-segmented condition, but did not differ from the chance level of 33% in the continuous condition. Likelihood ratio analysis suggests that, in the continuous condition, participants were 3.5 times more likely to perform at the chance level of 33% (2.6 for the lab-based experiments) than to perform at a level different from chance. These results thus suggest that participants in the continuous condition initiate their productions at random positions in the stream, and that they did not remember any word forms.

However, inspection of Figure 3a shows that the distribution in the continuous condition is bimodal, with some participants producing only words, and others producing only 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 SI4, 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 Figure S1, these results might be misleading because, in the continuous condition, many participants produce neither words *nor* part-words. In fact, on average, they produce only .4 words and part-words combined, respectively. (In the pre-segmented condition, most participants produce at least one word, with an average of 1.26.)

We thus turn to question of whether participants know where words start and end, asking if participants produce correct initial and final syllables. If participants use Statistical Learning to remember words, they should know where words start and where they end. In contrast, if they just track TPs, they should initiate the responses with random syllables. 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 1/3 should have correct final syllables. Given that participants tend to produce high-TP two-syllable chunks (i.e., *AB* and *BC* rather than *CD* chunks), the actual baseline level is somewhat higher. (For example, 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 $75\%/2 = 3/8$ rather than 1/3 items with correct initial syllables.) However, to evaluate the group performance, we keep the baseline of 1/3.

As shown in Table 1 and Figure 3c and d, participants produced items with correct initial or final syllables at greater than chance level only in the segmented condition, but not the continuous condition. In the continuous condition, the likelihood ratio in favor of the null hypothesis was 0.785 for initial syllables (3.61 for the lab-based experiment) and 4.06 for final syllables (2.14 for the lab-based experiment). While it is possible that performance in the continuous condition might exceed the chance-level of 1/3 with more than the 78 participants currently included, the actual chance-level is somewhat higher (about 38.4%). Critically, only 42% of the productions have 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 word at all, these results thus suggest that TPs do not allow learners to reliably detect onsets and offsets of words.

3. Discussion

Taken together, Experiments 1 and 2 suggest that Statistical Learning and (declarative) memory might fulfill different computational functions in the process of word-segmentation. In Experiment 1, participants tracked statistical dependencies predominantly when they were embedded in a continuous speech stream, but not across pre-segmented chunk sequences. This is consistent with earlier findings that Statistical Learning predominantly occurs within major prosodic groups, and, within these groups, predominantly at the edges of those groups (45). We show that, with shorter and better separated groups, Statistical Learning can be abolished altogether. In line with results from conditioning experiments (8, 26, 27, 50), 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

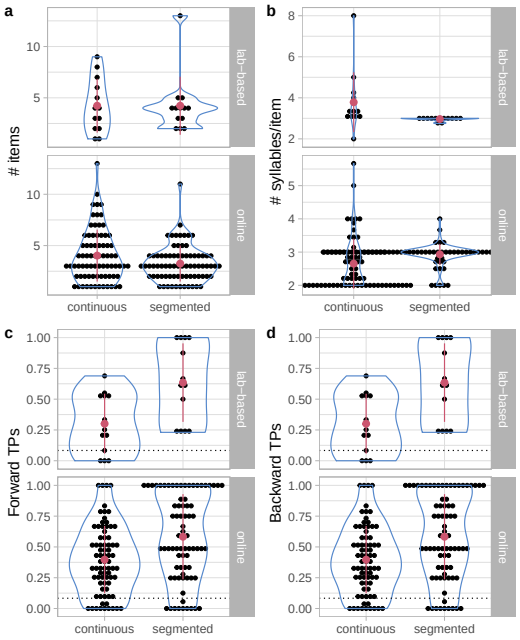

Fig. 2. Number of items produced, number of syllables per item and forward and backward TPs. The dotted line represents the chance level for a randomly ordered syllable sequence.

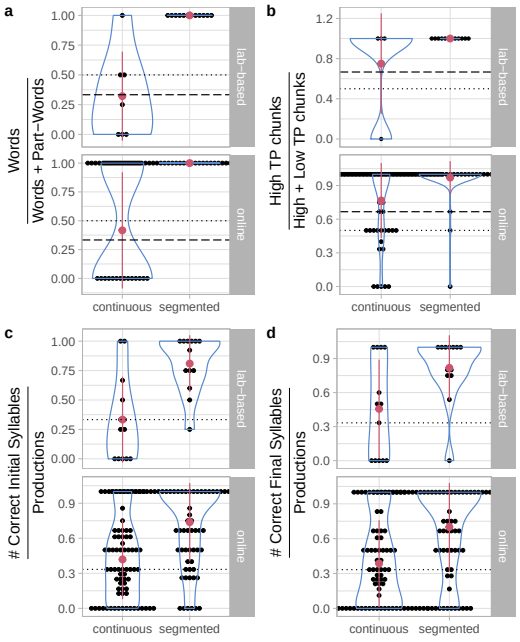

Fig. 3. Analyses of the participants' productions. (a) Proportion of words among words and part-words. The dotted line represents the chance level of 50% in a two-alternative forced-choice task, while the dashed line represents the chance level of 33% that an attested 3 syllable-chunk is a word rather than a part-word. (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. (c) proportion of productions with correct initial syllables and (d) with correct final syllables. The dotted line represents the chance level of 33%.

Table 1. Various analyses pertaining to the productions as well as test against their chances levels. The p value in the rightmost column reflects a Wilcoxon test comparing the continuous and the pre-segmented conditions.

	Continuous	Pre-segmented	*p*(Continuous vs. pre-segmented)
Recognition accuracy			
lab-based	*M*= 0.615, *SE*= 0.048, *p*= 0.048	*M*= 0.923, *SE*= 0.046, *p*= 0.0012	0.012
online	*M*= 0.628, *SE*= 0.0318, *p*= 7.84e-05	*M*= 0.911, *SE*= 0.0193, *p*= 7.08e-15	< 0.001
Number of items			
lab-based	*M*= 4.23, *SE*= 0.756, *p*= 0.0016	*M*= 4.23, *SE*= 0.818, *p*= 0.00152	0.812
online	*M*= 4.03, *SE*= 0.292, *p*= 3.17e-14	*M*= 3.25, *SE*= 0.202, *p*= 2.74e-14	0.099
Number of syllables/item			
lab-based	*M*= 3.79, *SE*= 0.421, *p*= 0.0016	*M*= 2.97, *SE*= 0.0246, *p*= 0.0007	0.026
online	*M*= 2.65, *SE*= 0.0869, *p*= 2.29e-14	*M*= 2.93, *SE*= 0.0364, *p*= 1.04e-15	< 0.001
Forward TPs			
lab-based	*M*= 0.301, *SE*= 0.0702, *p*= 0.0107	*M*= 0.634, *SE*= 0.092, *p*= 0.00159	0.006
online	*M*= 0.397, *SE*= 0.0316, *p*= 6.26e-12	*M*= 0.583, *SE*= 0.04, *p*= 3.82e-13	0.001
Backward TPs			
lab-based	*M*= 0.301, *SE*= 0.0702, *p*= 0.0107	*M*= 0.634, *SE*= 0.092, *p*= 0.00159	0.006
online	*M*= 0.397, *SE*= 0.0316, *p*= 6.26e-12	*M*= 0.583, *SE*= 0.04, *p*= 3.82e-13	0.001
Proportion of High-TP chunks among High- and Low-TP chunks			
lab-based	*M*= 0.75, *SE*= 0.289, *p*= 0.85 (vs. 2/3)	*M*= 1, *SE*= 0, *p*= 0.0006 (vs. 2/3)	1.000
online	*M*= 0.767, *SE*= 0.0459, *p*= 0.00154 (vs. 2/3)	*M*= 0.97, *SE*= 0.0187, *p*= 6.75e-13 (vs. 2/3)	< 0.001
Proportion of words among words and part-words (or concatenations thereof)			
lab-based	*M*= 0.321, *SE*= 0.153, 0.798 (vs. 1/3)	*M*= 1, *SE*= 0, *p*= 0.0006 (vs. 1/3)	0.034
online	*M*= 0.417, *SE*= 0.105, *p*= 0.189 (vs. 1/3)	*M*= 1, *SE*= 0, *p*= 2.08e-13 (vs. 1/3)	< 0.001
Proportion of items with correct initial syllables			
lab-based	*M*= 0.333, *SE*= 0.105, *p*= 0.856	*M*= 0.809, *SE*= 0.0694, *p*= 0.00186	0.016
online	*M*= 0.419, *SE*= 0.0392, *p*= 0.0864	*M*= 0.738, *SE*= 0.0387, *p*= 1.58e-11	0.000
Proportion of items with correct final syllables			
lab-based	*M*= 0.456, *SE*= 0.125, *p*= 0.5	*M*= 0.818, *SE*= 0.0829, *p*= 0.00222	0.025
online	*M*= 0.386, *SE*= 0.043, *p*= 0.456	*M*= 0.7, *SE*= 0.0437, *p*= 4.14e-10	0.000

sequences is consistent with the view that Statistical Learning is important for predictive processing (31, 32), 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 utterances, especially given that, due to its prosodic organization, speech tends to be pre-segmented into smaller groups (22–24, 41–44).

Experiment 2 provided the first direct test of the contents of the participants' (episodic or semantic) declarative memory after exposure to an 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 started with random syllables rather than actual word onsets. Given that the memory representations of linguistic items are based on their initial and final syllables (39, 40), this result thus suggests that Statistical Learning did not lead to the creation of declarative memory representations.

Contrary to this conclusion, some authors suggest that Statistical Learning might lead to declarative memories for chunks (51, 52). 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 (39). 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 combined results of Experiments 1 and 2 echo dissociations between associative learning and declarative memory (14–19), suggesting that the (cortical) declarative memory system might be independent of a (neostriatal) system for associative learning (16–18), though other authors propose that both types of memory involve the hippocampus (31, 53). In line with earlier proposals (31, 32), we thus suggest that the computational function of associative learning might be distinct from that of (declarative) memory encoding, and that associative 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.

These results also have implications for the more specific problem of word segmentation. 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 (54–56), attentional allocation to beginnings and ends of utterances (45, 57, 58), legal sound sequences (59) and universal aspects of prosody (41–44). Such cues might plausible 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, where linguistic sequences are encoded with reference to their first and their last element (39, 40).

To the extent that Statistical Learning reflects implicit memory systems (60), this suggestion mirrors earlier proposals that implicit and declarative memory systems might have different roles during language acquisition, with declarative memory systems supporting the acquisition of words and implicit memory system supporting the grammar-like regularities (61, 62). While we are agnostic about the extent to which Statistical Learning can support grammar acquisition, such results, together with the current ones, suggest that Statistical Learning and declarative memory might have separable functions, the former for predictive processing and the latter for remembering objects and episodes.

4. Methods summary

Unless otherwise stated, stimuli were synthesized using mbrola (47) and the *us3* (American English male) voice. Lab-based experiments were run using Psyscope X (http://psy.ck.sissa.it) in a quiet room. Online experiments were run on https://testable.org.

A. Participants. In Experiment 1, 30, 30 and 31 participants were retained for analysis for the pre-segmented condition, the continuous condition and its replication. In Experiment 2, 26 participants were retained for the lab-based version, and 157 for the online version. Participants reported to be native speakers of English.

B. Experiment 1. Participants were instructed to listen to a monologue in "Martian", and to remember the Martian words. Following this, they listened to a sequence of tri-syllabic words (Language 1: *w3:legu:*, *w3:levOI*, *w3:lenA:*, *faIzO:gu:*, *faIzO:vOI*, *faIzO:nA:*, *rVb{gu:*, *rVb{vOI*, *rVb{nA:*; Language 2: *w3:legu:*, *faIlegu:*, *rVlegu:*, *w3:zO:vOI*, *faIzO:vOI*, *rVzO:vOI*, *w3:b{nA:*, *faIb{nA:*, *rVb{nA:*). In Language 1 and 2, both TPs and the chunk frequency favored *AB+C* and *A+BC* patterns, respectively (TPs of 1.0 vs. 1/3; see main text). Segments lasted 60 ms and had an F_0 of 120 Hz. Sequences (45 repetitions/word) were either continuous or had 540 ms silences between words. Sequences were then played thrice (total familiarization: 7 min 17s (continuous); 18 min 14 s (pre-segmented)).

Following this familiarization, participants listened to pairs of items and had to choose the more "Martian" one. One item comprised the *first two* syllables of a word, one the *last two* syllables. The three items of each kind were combined into 9 test pairs. The test pairs were presented twice.

C. Experiment 2. Participants were instructed to listen to a monologue in "Martian", and to remember the Martian words. The languages were those from Saffran et al.'s (1) Experiment 2 (Language 1: *pAbiku*, *tibudO*, *dArOpi*, *gOLAtu*; Language 2: *bikuti*, *pigOLA*, *tudArO*, *budOpA*). Segments lasted 108 ms at an F_0 of 120 Hz. The words were combined into 20 sequences (45 repetitions/word) with different random orders, either continuously or with 222 ms silences between words. Sequences were played twice (total familiarization: 3 min 53 (continuous)

and 5 min 13 (pre-segmented)). Online participants watched a nebula during familiarization.

Following the familiarization and a 30 s filled retention interval, participants completed the recall test. Lab-based participants had 45 s to repeat back the words they remembered; their vocalizations were recorded for offline analysis. Online participants had 60 s to type their answer into a comment field. Finally, participants completed a recognition test during which we pitted words against part-words.

D. Analysis of productions. The responses were 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/). Finally, we selected the best matches to the familiarization stimuli (see SI2).

ACKNOWLEDGMENTS. We are grateful to E. Dupoux and L. Leisten for helpful discussions about earlier versions of this manuscript. ADE was supported by grant PSI2012-32533 from Spanish Ministerio de Economía y Competitividad and Marie Curie Incoming Fellowship 303163-COMINTENT.

1. JR Saffran, RN Aslin, EL Newport, Statistical learning by 8-month-old infants. *Science* **274**, 1926–8 (1996).
2. RN Aslin, JR Saffran, EL Newport, Computation of conditional probability statistics by 8-month-old infants. *Psychol Sci* **9**, 321–324 (1998).
3. MD Hauser, EL Newport, RN Aslin, Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition* **78**, B53–64 (2001).
4. NZ Kirkham, JA Slemmer, SP Johnson, Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* **83**, B35–B42 (2002).
5. JM Toro, JB Trobalon, N Sebastián-Gallés, Effects of backward speech and speaker variability in language discrimination by rats. *J Exp Psychol Anim Behav Process.* **31**, 95–100 (2005).
6. NB Turk-Browne, BJ Scholl, Flexible visual statistical learning: Transfer across space and time. *J Exp Psychol: Hum Perc Perf* **35**, 195–202 (2009).
7. J Chen, C Ten Cate, Zebra finches can use positional and transitional cues to distinguish vocal element strings. *Behav Process.* **117**, 29–34 (2015).
8. J Garcia, WG Hankins, KW Rusiniak, Behavioral regulation of the milieu interne in man and rat. *Science* **185**, 824–31 (1974).
9. AD Endress, Duplications and domain-generality. *Psychol. Bull.* **145** (2019).
10. CF Doeller, N Burgess, Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. *Proc Natl Acad Sci U S A* **105**, 5909–14 (2008).
11. SH Tompson, AE Kahn, EB Falk, JM Vettel, DS Bassett, Individual differences in learning social and nonsocial network structures. *J. experimental psychology. Learn. memory, cognition* **45**, 253–271 (2019).
12. LL Bonatti, M Peña, M Nespor, J Mehler, Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychol Sci* **16**, 451–459 (2005).
13. JM Toro, L Bonatti, M Nespor, J Mehler, Finding words and rules in a speech stream: functional differences between vowels and consonants. *Psychol Sci* **19**, 137–144 (2008).
14. P Graf, G Mandler, Activation makes words more accessible, but not necessarily more retrievable. *J. Verbal Learn. Verbal Behav.* **23**, 553–568 (1984).
15. AS Finn, et al., Developmental dissociation between the maturation of procedural memory and declarative memory. *J. Exp. Child Psychol.* **142**, 212–220 (2016).
16. BJ Knowlton, JA Mangels, LR Squire, A neostriatal habit learning system in humans. *Science* **273**, 1399–1402 (1996).
17. RA Poldrack, et al., Interactive memory systems in the human brain. *Nature* **414**, 546–550 (2001).
18. LR Squire, Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.* **99**, 195–231 (1992).
19. N Rungratsameetaweemana, LR Squire, JT Serences, Preserved capacity for learning statistical regularities and directing selective attention after hippocampal lesions. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 19705–19710 (2019).
20. RN Aslin, EL Newport, Statistical learning. *Curr. Dir. Psychol. Sci.* **21**, 170–176 (2012).
21. MS Seidenberg, MC MacDonald, JR Saffran, Does grammar start where statistics stop? *Science* **298**, 553–554 (2002).
22. A Cutler, D Oahan, W van Donselaar, Prosody in the comprehension of spoken language: A literature review. *Lang Speech* **40**, 141–201 (1997).
23. M Nespor, I Vogel, *Prosodic Phonology.* (Dordrecht, Foris), (1986).
24. S Shattuck-Hufnagel, AE Turk, A prosody tutorial for investigators of auditory sentence processing. *J Psycholinguist Res* **25**, 193–247 (1996).
25. AS Dunlap, DW Stephens, Experimental evolution of prepared learning. *Proc. Natl. Acad. Sci.* **111**, 11750–11755 (2014).
26. LT Martin, JR Alberts, Taste aversions to mother's milk: the age-related role of nursing in acquisition and expression of a learned association. *J. comparative physiological psychology* **93**, 430–445 (1979).
27. JR Alberts, DJ Gubernick, Early learning as ontogenetic adaptation for ingestion by rats. *Learn. Motiv.* **15**, 334 – 359 (1984).
28. SJ Gould, RC Lewontin, J Maynard Smith, R Holliday, The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. Royal Soc. London. Ser. B. Biol. Sci.* **205**, 581–598 (1979).
29. CE Shannon, Prediction and entropy of printed english. *Bell Syst. Tech. J.* **30**, 50–64 (1951).
30. AD Endress, SP Johnson, When forgetting fosters learning: A neural network model for statistical learning. *Cognition* **104621** (2021).
31. BE Sherman, NB Turk-Browne, Statistical prediction of the future impairs episodic encoding of the present. *Proc. Natl. Acad. Sci. United States Am.* **117**, 22760–22770 (2020).
32. NB Turk-Browne, BJ Scholl, MK Johnson, MM Chun, Implicit perceptual anticipation triggered by statistical learning. *J. neuroscience* **30**, 11177–11187 (2010).
33. R Levy, Expectation-based syntactic comprehension. *Cognition* **106**, 1126–1177 (2008).
34. JC Trueswell, I Sekerina, NM Hill, ML Logrip, The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition* **73**, 89–134 (1999).
35. A Clark, Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204 (2013).
36. K Friston, The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11**, 127–138 (2010).
37. GB Keller, TD Mrsic-Flogel, Predictive processing: A canonical cortical computation. *Neuron* **100**, 424–435 (2018).
38. J Jones, H Pashler, Is the mind inherently forward looking? comparing prediction and retrodiction. *Psychon. Bull. & Rev.* **14**, 295–300 (2007).
39. AD Endress, A Langus, Transitional probabilities count more than frequency, but might not be used for memorization. *Cogn. Psychol.* **92**, 37–64 (2017).
40. S Fischer-Baum, J Charny, M McCloskey, Both-edges representation of letter position in reading. *Psychon Bull Rev* **18**, 1083–1089 (2011).
41. D Brentari, C González, A Seidl, R Wilbur, Sensitivity to visual prosodic cues in signers and nonsigners. *Lang Speech* **54**, 49–72 (2011).
42. AD Endress, MD Hauser, Word segmentation with universal prosodic cues. *Cogn. Psychol* **61**, 177–199 (2010).
43. R Pilon, Segmentation of speech in a foreign language. *J. Psycholinguist. Res.* **10**, 113 – 122 (1981).
44. A Christophe, J Mehler, N Sebastian-Galles, Perception of prosodic boundary correlates by newborn infants. *Infancy* **2**, 385–394 (2001).
45. M Shukla, M Nespor, J Mehler, An interaction between prosody and statistics in the segmentation of fluent speech. *Cogn. Psychol* **54**, 1–32 (2007).
46. M Shukla, KS White, RN Aslin, Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proc Natl Acad Sci U S A* **108**, 6038–6043 (2011).
47. T Dutoit, V Pagel, N Pierret, F Bataille, O van der Vreken, The MBROLA 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 Processing.* (Philadelphia), Vol. 3, pp. 1393–1396 (1996).
48. M Peña, LL Bonatti, M Nespor, J Mehler, Signal-driven computations in speech processing. *Science* **298**, 604–7 (2002).
49. AD Endress, LL Bonatti, Words, rules, and mechanisms of language acquisition. *Wiley Interdiscip. Rev. Cogn. Sci.* **7**, 19–35 (2016).
50. DJ Gubernick, JR Alberts, A specialization of taste aversion learning during suckling and its weaning-associated transformation. *Dev Psychobiol* **17**, 613–628 (1984).
51. K Graf-Estes, JL Evans, MW Alibali, JR Saffran, Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol Sci* **18**, 254–60 (2007).
52. ES Isbilen, SM McCauley, E Kidd, MH Christiansen, Statistically induced chunking recall: A memory-based approach to statistical learning. *Cogn. science* **44**, e12848 (2020).
53. CT Ellis, et al., Evidence of hippocampal learning in human infants. *Curr Biol* **31**, 3358–3364.e4 (2021).
54. H Bortfeld, JL Morgan, RM Golinkoff, K Rathbun, Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychol Sci* **16**, 298–304 (2005).
55. M Brent, J Siskind, The role of exposure to isolated words in early vocabulary development. *Cognition* **81**, B33–44 (2001).
56. K Mersad, T Nazzi, When mommy comes to the rescue of statistics: Infants combine top-down and bottom-up cues to segment speech. *Lang. Learn. Dev.* **8**, 303–315 (2012).
57. P Monaghan, MH Christiansen, Words in puddles of sound: modelling psycholinguistic effects in speech segmentation. *J Child Lang* **37**, 545–564 (2010).
58. A Seidl, EK Johnson, Boundary alignment enables 11-month-olds to segment vowel initial words from speech. *J Child Lang* **35**, 1–24 (2008).
59. JM McQueen, Segmentation of continuous speech using phonotactics. *J Mem Lang* **39**, 21–46 (1998).
60. MH Christiansen, Implicit statistical learning: A tale of two literatures. *Top. Cogn. Sci.* **11**, 468–481 (2018).
61. MT Ullman, A neurocognitive perspective on language: The declarative/procedural model. *Nat Rev Neurosci* **2**, 717–26 (2001).
62. S Pinker, MT Ullman, The past and future of the past tense. *Trends Cogn Sci* **6**, 456–463 (2002).