Cognition 258 (2025) 106088

Contents lists available at ScienceDirect

Cognition

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

Implicit prediction as a consequence of statistical learning

Laura J. Batterink [*], Sarah Hsiung, Daniela Herrera-Chaves, Stefan Köhler

Department of Psychology, Western Centre for Brain and Mind, Western Institute for Neuroscience, University of Western Ontario, Canada

ARTICLE INFO

Keywords:
Statistical learning
Prediction
Implicit learning
Speech segmentation

ABSTRACT

The sensory input that we encounter while navigating through each day is highly structured, containing patterns that repeat over time. Statistical learning is the process of becoming attuned to these patterns and can facilitate online processing. These online facilitation effects are often ascribed to prediction, in which information about an upcoming event is represented before it occurs. However, previously observed facilitation effects could also be due to retrospective processing. Here, using a speech-based segmentation paradigm, we tested whether statistical learning leads to the prediction of upcoming syllables. Specifically, we probed for a behavioural hallmark of genuine prediction, in which a given prediction benefits online processing when confirmed, but incurs costs if disconfirmed. In line with the idea that prediction is a key outcome of statistical learning, we found a trade-off in which a greater benefit for processing predictable syllables was associated with a greater cost in processing syllables that occurred in a "mismatch" context, outside of their expected positions. This trade-off in making predictions was evident at both the participant and the item (i.e., individual syllable) level. Further, we found that prediction did not emerge indiscriminately to all syllables in the input stream, but was deployed selectively according to the trial-by-trial demands of the task. Explicit knowledge of a given word was not required for prediction to occur, suggesting that prediction operates largely implicitly. Overall, these results provide novel behavioural evidence that prediction arises as a natural consequence of statistical learning.

1. Introduction

Our environment is full of patterns that repeat over time, both within and across sensory modalities. For example, lightning is frequently followed by thunder, the chime of a doorbell is often followed by a dog's bark, and certain words regularly appear together in everyday speech, forming common phrases (e.g., "good" followed by "morning"). *Statistical learning* allows learners to become sensitive to these types of repeating patterns simply through exposure to structured input, a process that occurs without conscious effort or feedback (Aslin, 2017). In the first laboratory demonstrations of statistical learning (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), infants, children and adults were presented with a continuous speech stream composed of repeating trisyllabic nonsense words (e.g., *bidakupadoti*…). In a subsequent test period, learners showed sensitivity to the statistical patterns of the speech stream, revealed in infants through differential looking times to words versus recombined foil items (Saffran, Aslin, & Newport, 1996), and in older learners through recognition performance on a forced-choice task (Saffran et al., 1997; Saffran, Newport, & Aslin, 1996). These findings suggested that statistical learning may help to support speech segmentation, a crucial process in language acquisition in which words are identified from continuous speech (Bates & Elman, 1996; Saffran, Aslin, & Newport, 1996). Subsequent research showed that statistical learning extends into other domains, operating over diverse types of stimuli such as auditory tones (Creel, Newport, & Aslin, 2004; Saffran, Johnson, Aslin, & Newport, 1999), environmental sounds (Ordin, Polyanskaya, & Samuel, 2021; Siegelman, Bogaerts, Elazar, Arciuli, & Frost, 2018), abstract shapes (Fiser & Aslin, 2001, 2002; Turk-Browne, Jungé, & Scholl, 2005), cartoon aliens (Arciuli & Simpson, 2012), everyday scenes (Brady & Oliva, 2008), tactile patterns (Conway & Christiansen, 2005) and cross-modal audio-visual associations (Mitchel & Weiss, 2011).

An important theoretical idea is that a key function of statistical learning is to predict future events (e.g., De Lange, Heilbron, & Kok, 2018; Sherman et al., 2022; Sherman & Turk-Browne, 2020; Turk-Browne, 2012). Prediction in the context of statistical learning is said to occur if the acquired regularities are used *prospectively*, to anticipate upcoming events before they have occurred, which in turn facilitates the perception and/or detection of incoming input (Turk-Browne, 2012).

[*] Corresponding author at: 1151 Richmond St WIRB 6124, London, ON N6A 5B7, Canada.
E-mail address: lbatter@uwo.ca (L.J. Batterink).

https://doi.org/10.1016/j.cognition.2025.106088
Received 23 September 2024; Received in revised form 12 February 2025; Accepted 14 February 2025
Available online 21 February 2025

This idea makes sense intuitively; by extracting patterns in previously encountered input, we may apply this knowledge to novel situations and predict items or events that have not yet occurred. However, as we will review further below, there is limited direct evidence for prediction in the current literature on statistical learning; many previous findings taken to support the notion of prediction in statistical learning could in principle be explained by alternative accounts. Thus, the goal of the current study was to directly examine the assumption that statistical learning enables prediction, and in particular, to test whether statistical learning enables *genuine* prediction (Kuperberg & Jaeger, 2016), in which a commitment is made to a preactivated candidate prior to its occurrence.

1.1. Conceptualizing prediction

A common idea is that prediction entails pre-activation, in which information about an upcoming stimulus is activated or represented before it is encountered (Bubic., 2010; Kuperberg & Jaeger, 2016; Kutas, Federmeier, & Urbach, 2014). For example, when hearing the beginning of the sentence "She went to the bakery to buy a loaf of…" a listener may activate the representation of the word "bread" prior to hearing the final word. Prediction is graded, probabilistic and can occur at multiple representational levels; in language processing contexts, these levels include semantic meaning, syntactic structure and phonological representations (Kray, Sommerfeld, Borovsky, & Häuser, 2024; Kuperberg & Jaeger, 2016). However, it has been argued by some that "true" prediction goes beyond mere pre-activation and requires some level of "commitment" to preactivated candidates, prior to encountering new input (Kamide, 2008; Lau, Holcomb, & Kuperberg, 2013; for review, see Kuperberg & Jaeger, 2016). Critically, by this view, if a prediction is not met, this leads to a processing cost relative to a "neutral" condition in which no predictions are made. For example, in the earlier sentence, a commitment to the word "bread" would result in greater difficulty integrating a less-expected word, such as "cake," compared to a sentence where no commitment was made. In other words, a hallmark of true prediction is that any particular prediction produces benefits when confirmed, but incurs costs if it turns out to be incorrect (Van Petten & Luka, 2012). A classic illustration of this trade-off comes from the Posner cueing paradigm, in which a valid cue of a target's location leads to faster responses, but an invalid cue leads to slower responses relative to a neutral condition of no prior information (Posner, 1980).

Predictive processing contrasts with *retrospective* processing, which occurs only after the current stimulus is encountered (Bubic., 2010; Dale, Duran, & Morehead, 2012). In a retrospective model, the generation of a new response is carried out once information about the current stimulus is integrated with relevant information about the prior stimuli (Dale et al., 2012). If relevant recent representations are still active when encountering the current stimulus, this can facilitate responses, without requiring any forward-looking predictive mechanisms. For example, in the previous illustrative sentence, processing of the word "bread" might be facilitated due to easier integration with the prior congruent word "bakery", but not because it is actively predicted. As others have noted (Bubic., 2010; Dale et al., 2012), it is not trivial to demonstrate that a certain phenomenon is driven by true predictive processing rather than retrospective processing. For example, faster response times to predictable items in a sequence could reflect true prediction, but could also reflect activated memory traces for prior congruent information that facilitates responses only after the current item is presented (Dale et al., 2012). Nonetheless, some studies have overcome these challenges through clever experimental designs and successfully demonstrated true prediction across different domains, including prediction of upcoming words during sentence processing using visual world experiments (Altmann & Kamide, 1999; Arai & Keller, 2013) as well as prediction of upcoming locations of a visual cue in a sequence through continuous computer-mouse tracking (Dale et al., 2012).

Similar distinctions between prospective and retrospective processes have been made to account for semantic priming effects (Yap, Hutchison, & Tan, 2017). For example, the prime word "cat" may automatically preactivate related nodes (such as "dog") through automatic spreading activation (Posner & Snyder, 1975), facilitating identification of the target word once it is presented. Priming effects may also reflect retrospective processes such as semantic matching, in which finding a semantic match between the target and the prime facilitates decision-making (Neely, Keefe, & Ross, 1989). In the context of statistical learning, using as an example the word *bidaku* (as in Saffran, Aslin, & Newport, 1996), a true predictive mechanism would entail that a learner commits to the syllable "ku" as soon the syllable string "bida.." is presented. In contrast, a retrospective mechanism would integrate the syllable "ku" with the prior syllables "bida" only after the full word is encountered.

1.2. Re-examining past evidence for prediction in statistical learning

The theoretical idea that statistical learning supports prediction (e.g., De Lange et al., 2018; Sherman et al., 2022; Sherman & Turk-Browne, 2020; Turk-Browne, 2012) is consistent with a signature outcome of statistical learning studies – the finding that later-occurring items within a triplet or unit are processed more efficiently than initial items (e.g., "ku" is faster than "bi" in the word *bidaku*). This type of facilitation is typically demonstrated using a target detection task, in which participants are asked to detect pre-defined targets embedded within shortened versions of a previous familiarization stream. In the visual modality, participants have been shown to respond more quickly to images occurring within later (2nd and 3rd) positions of visual triplets compared to initial items (Barakat, Seitz, & Shams, 2013; Bays, Turk-Browne, & Seitz, 2015; Bertels, Franco, & Destrebecqz, 2012; Campbell, Zimerman, Healey, Lee, & Hasher, 2012; Kim, Seitz, Feenstra, & Shams, 2009; Musz, Weber, & Thompson-Schill, 2015; Turk-Browne et al., 2005). Notably, the second and third items of a triplet are characterized by having higher *transitional probabilities* (Saffran, Newport, & Aslin, 1996)—appearing consistently after their preceding neighbour in the sequence—and are thus more predictable than initial items within a triplet that follow a triplet boundary. Similarly, in the auditory modality, participants respond more quickly to the second and third syllables of a word compared to word-initial syllables (Batterink, 2017; Batterink, Mulgrew, & Gibbings, 2024; Batterink & Paller, 2017, 2019; Batterink, Reber, Neville, & Paller, 2015; Batterink, Reber, & Paller, 2015; Batterink & Zhang, 2022; Franco, Eberlen, Destrebecqz, Cleeremans, & Bertels, 2015; Kiai & Melloni, 2021; Liu, Forest, Duncan, & Finn, 2023; Luo, Cao, & Wang, 2024; Moreau, Joanisse, Mulgrew, & Batterink, 2022; Sweet, Van Hedger, & Batterink, 2024; Wang et al., 2023; Wang, Köhler, & Batterink, 2023). This RT advantage for later-occurring syllables appears to be dissociable from participants' explicit knowledge of the statistical regularities, possibly because both explicit and implicit representations are acquired in parallel during learning (Batterink, Reber, Neville, & Paller, 2015; Franco et al., 2015; Kim et al., 2009; Liu et al., 2023). In addition, it has been demonstrated that these RT facilitation effects emerge very rapidly, as quickly as the second presentation of a triplet (Batterink, 2017; Luo et al., 2024). Overall, these findings show that learners can rapidly use their statistical knowledge to optimize processing of predictable items.

At face value, these temporal cueing benefits can be—and often are—taken as evidence for prediction (e.g., Batterink, Reber, & Paller, 2015; Luo et al., 2024; Turk-Browne, 2012). Participants who encounter the initial part of a word (e.g. "bida…") may predict the subsequent syllable "ku" even before it occurs, enabling them to more quickly prepare their motor response for target detection. However, retrospective processes could also account for these observed RT facilitation effects, without needing to invoke forward-looking processes. As mentioned previously, a retrospective model can process a current stimulus more quickly if the activated memory traces for the prior context are congruent or otherwise support processing of the current stimulus (Dale et al., 2012). For example, during the target detection task, once a participant encounters a syllable in the third position of a word (e.g., "ku"), their processing of this syllable may be facilitated by considering the two prior syllables ("bida"). The congruency of the prior context may increase the participant's confidence that the target syllable was indeed just encountered, allowing them to more rapidly reach a decision threshold for making a response. By contrast, the initial syllable in the word ("bi") would not benefit from retrospective processing of a prior congruent context, and thus responses would be relatively slower. It has also been shown that predictable items are perceptually enhanced as a result of statistical learning, at least in the visual modality, with second items in a pair being detected more easily than first items even when presented outside of their usual pairing (Barakat et al., 2013). This type of perceptual enhancement could also account for the faster responses to later items, without the need to invoke prediction. Thus, behavioural results alone currently do not provide conclusive support for genuine prediction effects in statistical learning.

Some additional evidence for prediction to consider is offered by neuroimaging data in visual statistical learning paradigms. In a fMRI study, Turk-Browne and colleagues (Turk-Browne, Scholl, Johnson, & Chun, 2010) presented participants with a stream of faces and scenes, containing both regularly occurring pairs (a face followed by a scene, or scene followed by a face) as well as unpaired images that were not predictive. Relative to unpaired images, the initial image of a pair elicited greater activation of the right anterior hippocampus, suggestive of anticipation. In addition, faces that were paired with subsequent scenes elicited enhanced neural activation in a scene-selective ROI within the parahippocampal place area. These effects provide evidence of prediction, whereby anticipation of the next image recruits the hippocampus and prospectively modulates activity in relevant visual cortex. In a more recent fMRI study using neural decoding, participants viewed a continuous stream that contained pairs of scenes, in which the category of the first scene of the pair predicted the second (e.g. beach – mountain) (Sherman & Turk-Browne, 2020). Within the hippocampus, the category of the upcoming second image could be decoded prior to its onset, while the first image was still being shown, again providing evidence of prediction. A subsequent intracranial EEG study that used a similar scene pair paradigm (Sherman et al., 2022) found that electrode contacts within visual cortex showed category evidence for the upcoming second image of a pair before it was presented. Although shown within a different region of the brain, this result converges with the prior (2020) fMRI finding showing significant neural decoding for predictable yet-to-be-shown items.

At first glance, these results again appear to provide evidence that statistical learning can generate representations to be used predictively—at least within the visual modality. Yet, there may be alternative explanations for these results. As described by (Endress, 2024), pattern similarity analyses of units presented in a typical statistical learning stream may reflect the co-activation of the *necessarily* similar contexts for items that belong to the same unit. For example, in the intracranial EEG study by Sherman and colleagues (Sherman et al., 2022), the structured stream consisted of only three scene exemplar pairs (e.g., AB, CD, and EF) in a given block. Given that back-to-back repetitions of pairs were not allowed, a single pair (e.g., AB) would necessarily follow only two items (e.g., D or F). It is possible that the prestimulus category decoding evidence for the second item of a given pair (e.g., B) may reflect the similar context that precedes this item (i.e., items DA and FA) (Endress, 2024). Thus, given the evidence reviewed, additional approaches would be useful to more conclusively address the question of whether statistical learning leads to genuine prediction, whereby a commitment is made to an expected item or event (Kuperberg & Jaeger, 2016).

1.3. The current study

To address this question, we used a behavioural approach to test whether statistical learning involves prediction of upcoming events. Here, we capitalized on a signature of prediction – the finding that predictions produce benefits when confirmed, but costs when disconfirmed, as previously described (Van Petten & Luka, 2012). If statistical learning produces representations that can be deployed in the service of prediction, we expected that learners' processing would benefit from these predictions when correct, while being impaired when predictions are violated.

To test this hypothesis, we modified the standard target detection task such that a target syllable could occur not only within its expected position in its usual word (as in previous versions of this task, e.g., (Batterink, Reber, Neville, & Paller, 2015), but also in unexpected positions, inserted into another word where it normally did not occur. We referred to these two types of words as "standard" words and "mismatch" words. At a basic level, we predicted that (1) later-occurring syllables within standard words would show progressively faster reaction times, replicating prior work; and (2) syllables occurring within mismatch words (i.e., outside of their expected context) would be detected more slowly and less accurately than syllables within standard words. Evidence for both (1) and (2) would indicate that more predictable syllables are facilitated during online processing. Critically, as a direct test of our hypothesis, we also predicted that (3) syllables that are more strongly facilitated within expected positions in standard words will be correspondingly more strongly impaired when they occur in mismatch contexts, outside of their expected positions. That is, if active prediction is supporting processing of the syllables, there should be a trade-off whereby committing to an expectation of when a syllable will occur should have both benefits (when the syllable does occur in the expected position) and costs (when the syllable occurs outside of its expected context).

As a preview of our design, participants listened to an artificial language stream made up of repeating trisyllabic words, followed by the modified target detection task. In addition, they completed two explicit measures of statistical learning—the familiarity rating task and the two-alternative forced-choice (2AFC) task (Batterink & Paller, 2017; Batterink, Reber, Neville, & Paller, 2015)—in order to provide a thorough characterization of knowledge acquired through statistical learning. Reaction times on the target detection task provided the key test of our hypothesis that there should be at trade-off in making predictions.

2. Methods

2.1. Participants

Based on previous research showing robust effects on the target detection task with sample sizes ranging from 19 to 22 participants per group (Batterink, 2017; Batterink, Reber, Neville, & Paller, 2015; Batterink, Reber, & Paller, 2015), we considered that 30 participants would provide a well-powered sample size for the current study. A total of 33 young adults aged 18–35 ($M = 18.9$, $SD = 2.9$) completed the experiment, of which 3 were excluded from analysis due to poor performance on the target detection task (failing to detect at least 60 % of the targets and/or showing a false alarm rate in excess of 15 % relative to detected targets). Exclusion criteria were based on previous results from the target detection task, which reported mean detection accuracy in the range of 83 % - 89 % with a false alarm rate around 10 % (Batterink & Paller, 2017, 2019). After these exclusions, we obtained a final sample size of 30 participants, meeting our goal sample size. All participants self-reported normal hearing and were fluent English speakers. Participants were compensated in the form of course credits as part of their enrollment in a university psychology course.

2.2. Stimuli

The main stimuli used for the experiment were 12 auditory syllables (*ba*, *fe*, *fu*, *ge*, *ko*, *me*, *ni*, *pu*, *re*, *ru*, *su*, and *ti*, taken from previous statistical learning studies (Batterink & Paller, 2019; Wang, Rosenbaum, et al., 2023). These syllables were originally produced by a male native English speaker using neutral intonation and no co-articulation between syllables. Syllables were combined to form four trisyllabic nonsense words (*bafuko, regeme, rupuni, fetisu*), henceforth referred to in this study as "standard" words. As described further below, syllables were also presented as part of "mismatch" words, which were only introduced during the target detection task (not during the initial exposure phase). Mismatch words consisted of standard words with one syllable from a different word replacing one of the original syllables (e.g., *su-ge-me*, *ba-ti-ko*, *fe-ti-ru)*. Mismatch syllables could appear in any position within a mismatch word (1st, 2nd, or 3rd). All participants were exposed to the same words. Within the exposure period, initial syllables in standard words had transitional probabilities of 0.33, and second and third syllables in standard words had transitional probabilities of 1.0. Mismatch syllables within mismatch words had transitional probabilities of 0.

2.3. Procedure

Participants completed the experimental session in person. The experimental procedure consisted of the exposure phase, followed by the modified target detection task that included both standard and mismatch words. After the target detection task, participants completed two additional measures of statistical learning, allowing us to assess participants' explicit knowledge of the regularities (Batterink & Paller, 2017; Batterink, Reber, Neville, & Paller, 2015): the familiarity rating task, and the two-alternative forced-choice task (2AFC task). The entire experiment took approximately 50 min to complete. An overview of the entire experimental procedure is outlined in Fig. 1. Auditory stimuli were presented at a comfortable listening level from two speakers. The experiment was run from a laptop computer using the program Psychopy (Peirce et al., 2019).

2.4. Exposure phase

Participants were presented with a continuous auditory stream composed of the four standard words concatenated together in pseudorandom order, with the constraint that the same word did not appear consecutively. Each syllable was presented at a rate of 300 ms, with total of 1080 syllables (360 words) presented. Participants were instructed to listen to the stream, but were not instructed that there were underlying patterns or that the syllables were arranged into trisyllabic words. Syllable volume for the first 3 syllables and last 3 syllables of the stream gradually ramped up and down, respectively, in order to prevent the onset and offset of the stream from being used as cues for word boundaries.

2.5. Target detection task

In this task, participants made speeded responses to syllable targets embedded in continuous streams of the artificial language. The task included a total of 48 streams, with each stream consisting of 144 syllables in total (48 triplets), arranged together in a similar manner and presented at the same rate as in the exposure phase. However, unlike in the exposure phase, the streams in this task contained occasional mismatch words, in which a syllable from a standard word was replaced by a syllable from another word.

For each stream, participants were required to detect a specific target syllable, which varied from stream to stream. Both RT and accuracy were emphasized. Prior to the onset of each stream, the target syllable was presented twice auditorily, and its written form was displayed on the screen. Participants then initiated the syllable stream via button press. The written form of the target syllable remained on the screen while the stream was presented. In each stream, a total of 12 syllable targets occurred within standard words, while an additional 2–3 target syllables replaced a syllable within a word where it did not normally occur (creating a mismatch word). Therefore, in each stream, there were either 14 or 15 total targets to detect. The reason that the number of mismatch words per stream was not consistent per stream was because the total number of unique mismatch words that could be constructed (108) did not divide evenly into our 48 streams, as described further below. Other than the 2–3 mismatch words per stream containing target syllables, all other syllables were organized into standard words. As in the exposure stream, there was a ramp-up in volume over the first 3 syllables of each stream and a ramp-down for the final 3 syllables. Again, this was done so that the onset and offset of the streams could not provide a cue for word segmentation.

There were multiple constraints that guided the construction of each stream. No mismatch word could occur back-to-back with another mismatch word, and words containing a target syllable (whether a standard word or mismatch word) could also not be presented back-to-back. In addition, the target could not occur within the first or last word of the stream, to avoid any idiosyncratic effects on RTs related to stimulus onset and offset. The first word appearing in a stream could not be the same as the last word that appeared in the previous stream. In constructing the mismatch words, across the entire task, we ensured that each of the 12 syllables in the inventory replaced each of the 9 syllables from the other three words (not including its own word) a single time, yielding a total of 108 unique mismatch words. Mismatch syllables notwithstanding, each word was represented the same number of times in each stream. While following these constraints, word order was pseudorandomized within each stream.

The task was subdivided into 4 blocks of 12 streams each, with each syllable serving as the target syllables once per block. Across the entire task, there were a total of 576 target syllables embedded within standard words (192 targets in each triplet position) and 108 target syllables embedded within mismatch words (36 mismatch targets in each triplet position). Prior to the target detection task, there was a practice phase that consisted of two streams, with a similar structure to the main task, but composed of a separate inventory of syllables. For this practice phase, participants were given feedback at the end of each stream, indicating the number of targets out of 5 that they detected, as well as their average reaction time. In the main task, participants were not given feedback; however, after each stream, they were shown the number of streams completed out of 48 to give them a sense of their progress. Participants were informed that they could take breaks between the streams if needed. Individual syllables were presented at the same rate as in the Exposure phase (300 ms/syllable), and the entire task took approximately 45 min to complete.

2.6. Familiarity rating task

The familiarity rating task was designed to assess participants' explicit memory of the nonsense words in the language. On each trial, participants were presented with a trisyllabic sequence (a word, part-word, or nonword), and asked to rate how familiar each item sounded to them on a scale from 1 (not familiar) to 4 (very familiar) through button press response. Words consisted of standard words presented in the exposure stream (e.g., *re-ge-me*). Partwords consisted of a syllable pair from a word plus a syllable from a different word (e.g., *ba-fu-me*),[1] and are a small subset of the total set of "mismatch" words. Lastly, nonwords consisted of three syllables that were each from a different word (e.g., *ni-su-ba*). Individual syllables were presented within each sequence at the same rate as in the Exposure phase (300 ms/syllable)

[1] One of the part-words contained a syllable pair from an original standard word, but presented in the wrong order from what was initially intended (*rusuti* rather than *rutisu*).

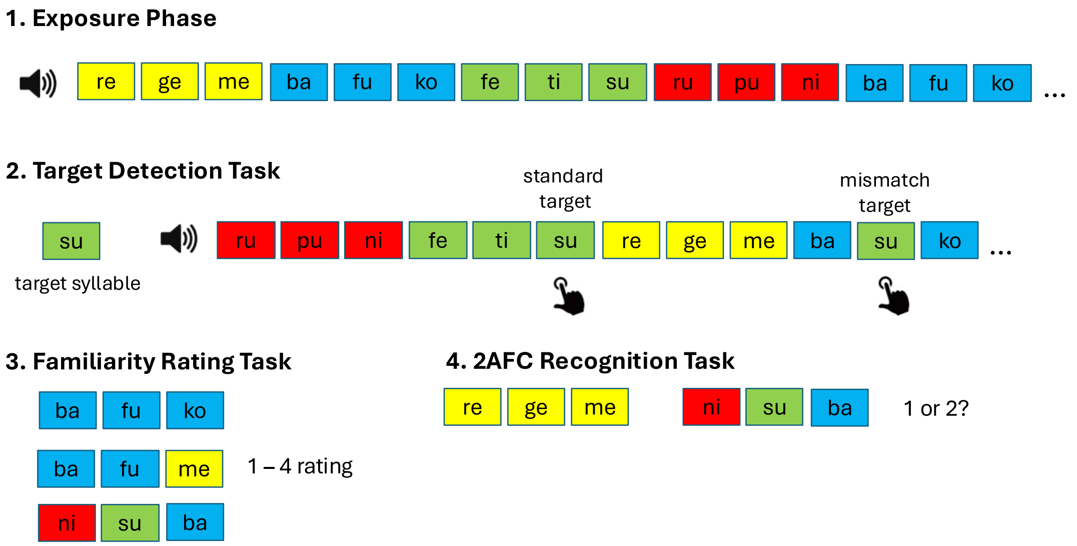

Fig. 1. Overview of experimental procedure with example trials. All syllables were presented in the auditory modality. Participants completed an initial exposure period, followed by the target detection task, and two explicit measures of statistical learning, the familiarity rating task and the 2AFC recognition task.

and responses were given without time pressure. There were 12 trials in total, with 4 of each type (word, partword, and nonword). Trial order was randomized.

2.7. *Two-Alternative Forced Choice (2AFC) recognition task*

The two-alternative forced choice (2AFC) recognition task provided an additional assessment of learner's explicit memory of the words from the stream. On each trial, participants were presented with a word and a nonword, and asked to indicate which word sounded more familiar (the 1st or 2nd item presented) by pressing 1 or 2 on a keypad. There were a total of 4 words and 4 nonwords, paired exhaustively to produce a total of 16 trials. The two trisyllabic items within each trial were separated by a 1000 ms interval. Participants provided their responses without time pressure. Trial order was randomized and whether a word or nonword appeared first on a given trial was counterbalanced across participants.

2.8. *Data analysis*

2.8.1. *Target detection task*

2.8.1.1. Reaction times. Our main dependent measure was reaction time (RT) on the target detection task. Responses that occurred within 1200 msec after a target were considered to be hits, while all other responses were considered to be false alarms; this is the same criterion used in our past studies (e.g., Batterink & Paller, 2017, 2019; Batterink, Reber, Neville, & Paller, 2015; Batterink, Reber, & Paller, 2015). We noted that there was a special scenario in which regular first position syllables also appeared in the first position in a mismatch word. These syllables cannot be considered "true" mismatch syllables. For example, when the syllable *ru* (which typically appears in the word *rupuni*) occurs in the mismatch word *ru-ge-me* (i.e., replacing the syllable *re* in the standard word *re-ge-me*), it would not be registered as occurring within a mismatch context until *after its own presentation*, only once the subsequent syllable (*ge*) is presented. For analysis purposes, we excluded these special cases from our main RT-based analyses. As such, there were fewer total trials for analysis in the initial triplet position compared to the second and third positions (24 versus 36 trials), within the mismatch condition.

In our initial analysis, we tested the prediction that RTs would be faster for later occurring syllables within a word, for standard words only and not for mismatch words. RTs were modeled using a linear mixed-effects model, with independent variables including word type (standard word versus mismatch word), triplet position (1–3), and their interaction, in addition to target position within the stream (4–141) and stream number across the task (1–48). Target position within the stream and stream number were not variables of direct interest but were included in the model to account for possible fluctuations in RTs over time, both within and across the streams. In addition, by including target position within the stream in the model, we ensured that any effect of triplet position was not due to a general speeding-up of responses over the course of the stream. We expected a linear effect of triplet position on RTs for standard words (with later occurring targets within a triplet eliciting progressively faster RTs). All independent variables were modeled as continuous predictors except for word type, which was categorical. Treatment coding was used for word type, with standard words set as the reference. Random intercepts were included for participant and for syllable token (i.e., the 12 syllables).

Next, we tested our main hypothesis that there should be a trade-off in making predictions, such that greater facilitation to expected, predictable syllables will correspond to increased costs in processing syllables that violate these expectations. We tested this hypothesis at both the participant and item (syllable) level. At the participant level, we computed the "RT prediction effect" by subtracting the average RT for the final syllable position from the average RT for the initial syllable position and dividing it by the average RT for the initial syllable position, within standard words only ($[\text{RT}_1 - \text{RT}_3]/\text{RT}_1$;(Batterink & Paller, 2019). This measure controls for participants' individual baseline RTs and provides an index of each individual's relative RT facilitation for predictable syllables. We also computed a "RT mismatch cost" by subtracting each participant's mean RTs across all standard syllables from their mean RTs to mismatch syllables, and dividing by the RT to standard syllables ($[\text{RT}_{\text{mismatch}} - \text{RT}_{\text{all-standard}}]/\text{RT}_{\text{all-standard}}$). The RT mismatch cost provides an index of an individual's relative processing impairment to unexpected, unpredictable syllables relative to those occurring in their usual positions and words, again controlling for individual baseline RTs. We pooled the mismatch syllables across all three triplet positions in this measure because our initial analyses revealed that RTs to mismatch syllables were not modulated by triplet position (see Results). We then computed the Pearson correlation between the RT prediction effect and the RT mismatch cost. We expected that greater prediction would be associated with greater mismatch processing costs (i.e. slower RTs to mismatch syllables).

Similarly, at the item level, we expected that if a given syllable is strongly expected to occur in a given context, it should be processed with greater difficulty when it occurs outside that context. To evaluate this hypothesis, across all participants and syllables, we computed the partial correlation between the mean RT for standard occurrences and mismatch occurrences at the syllable level (i.e., 360 individual standard-mismatch pairs), while controlling for each participant's overall mean RT across all syllables in standard words. We expected that a given syllable that shows relatively stronger RT facilitation within a standard word would be associated with a proportionately *slower* RT when it occurs within a mismatch word.

As an additional exploratory analysis, we tested whether facilitation effects associated with statistical learning reflect learning of specific syllable-to-syllable transitions. To the extent that a learner has successfully extracted the predictable relationship between syllables in a word, faster detection of the initial syllable (due to perceptual or other idiosyncratic factors) could serve to facilitate the detection of the subsequent syllables within the word, by alerting the learner sooner to the presence of the upcoming target syllable. Such a finding would be compatible with a prediction account. In contrast, if the learner did not learn the relationship between neighbouring syllables within a word, no relationship would be expected between response times to earlier and later syllables within the same word. We therefore examined the impact of average RTs to previous syllables on a current (predictable) syllable within the same standard word. Average RTs to the previous syllable were computed from other streams in which it served as the target. Within standard words only, we modeled RTs at the individual trial level for predictable syllables only (n = syllables occurring in position 2 and 3 in standard words) in a separate linear mixed-effects model, using the mean RT for the previous syllable (n−1) as a predictor. We also included stream position (4–141) and random intercept for syllable and participant in the model.

2.8.1.2. Accuracy. To fully characterize performance on the target detection task, we also assessed accuracy as a function of triplet position and word type. For each participant, we computed the mean accuracy within each triplet position (1,2,3) and word type (standard words, mismatch words). Accuracy was then analyzed using a repeated-measures ANOVA with the factors of word type and triplet position.

2.9. Familiarity rating task and 2AFC recognition task

Ratings on the familiarity rating task were analyzed with a repeated measures ANOVA using trial type (word, part-word, and nonword) as a within-subjects factor. As a single metric of performance for subsequent correlation analyses, we computed a familiarity rating score by subtracting the average familiarity rating for partword and nonwords from the average rating for words (Batterink & Paller, 2017). Performance on the 2AFC task was assessed with a one-sample *t*-test against chance level.

2.10. Correlations in performance across tasks

In the final set of analyses, to understand the degree to which prediction operates implicitly or explicitly, we explored correlations in performance between our three experimental tasks at the individual participant level. We computed Pearsons' correlations between the RT prediction effect, the RT mismatch cost, the familiarity rating score, and recognition accuracy.

We further explored whether participants' explicit knowledge of the specific word-level regularities was related to facilitation of predictable syllables. For each participant, we computed (1) the accuracy for each individual word on the 2AFC recognition measure (out of a maximum of 4 trials), and (2) the "baselined" or mean-centered RT for each standard syllable by subtracting the average RT across all syllables from the mean RT for a given syllable (within standard words only); this controls for individual differences in response times. Across participants, we then computed the correlation between (1) each word's recognition accuracy and (2) the "baselined" mean RT to the third (most predictable) syllable within the given word. Next, these four correlation values (one for each word) were averaged, yielding a mean observed correlation value between recognition and RT facilitation to third syllables at the word level. We used nonparametric permutation testing to assess statistical significance of the observed correlation value, while controlling for position-specific effects. Specifically, for each participant, we shuffled their baselined mean RT values within each triplet position (i.e., third syllable values *ni, me, su* and *ko* were shuffled and relabeled with a randomly assigned syllable name within this set). We then recomputed the word-level correlations between recognition accuracy and baselined RTs with these shuffled values. This procedure was repeated 1000 times, creating a surrogate distribution of shuffled (position-controlled) correlation values. The corresponding *p* value of the observed correlation was then computed with reference to this surrogate distribution of shuffled values, using a two-tailed test with significance set at $p < 0.05$. An above chance correlation indicates a relationship between performance on the 2 AFC task and RT speeding to the third syllable, over and above any effects attributable to general RT facilitation to third position syllables. The same procedure was also conducted for the first and second syllable positions of a word. In addition, using a similar procedure, we also assessed the correlation between recognition accuracy to a given word and the RT mismatch cost for the word's predictable (2nd and 3rd) syllables. In this analysis, we computed the correlation between each word's recognition accuracy and the baselined mean RT to mismatch syllables within a given word, computed by subtracting the mean RT across all syllables from a given syllable's mean RT to occurrences within mismatch words. We then examined whether this observed correlation differed significantly from a distribution of 1000 surrogate correlations, computed by shuffling syllable labels across the 12 syllables for each iteration.

Finally, to assess the related question as to whether explicit knowledge is necessary for online prediction effects on the target detection task, we conducted an item subset analysis, as follows. For each word in turn, we excluded all participants who showed above-chance performance on the 2 AFC recognition task for that given word. This resulted in a list of "unknown" words, defined as words that failed to be recognized at above-chance levels on the recognition task (Batterink, Reber, Neville, & Paller, 2015). We then extracted the RT prediction effect for each unknown word (computed in the same way as before, i.e., $[RT_1 - RT_3]/RT_1$, but considering only the 1st and 3rd syllables within the given word) and collated these word-specific prediction effects. We also compiled the RT mismatch cost for the predictable (2nd and 3rd syllables) of each unknown word, computed as before ($RT_{mismatch} - RT_{all\text{-}standard}$). If the RT prediction effect and the RT mismatch cost for 2nd and 3rd syllables of these unknown words exceeded 0, as tested using a one-sample *t*-test for each test, this would indicate that online prediction effects occur even in the absence of explicit knowledge.

3. Results

3.1. Expected RT pattern observed within standard words

Across the three triplet positions, RTs were faster to syllables that occurred within standard words (estimated marginal mean = 544 ms, SE = 12.7 ms) compared to those occurring within mismatch words (estimated marginal mean = 663 ms, SE = 13.1 ms; z ratio = −33.2, $p <$ 0.001; see Fig. 2A). In addition, as we expected, reaction times showed significant differences as a function of triplet position between standard words and mismatch words (Word Type x Triplet Position: t(1069) = 8.58, $p < 0.001$). To understand this interaction, we examined the separate linear trends of triplet position by word type. In standard words, there was a significant decrease across triplet positions, as expected based on prior results (trend estimate = −61.6 ms, SE = 4.17, p < 0.001). In contrast, within the mismatch word condition, RTs were stable across triplet positions, with no significant differences as a

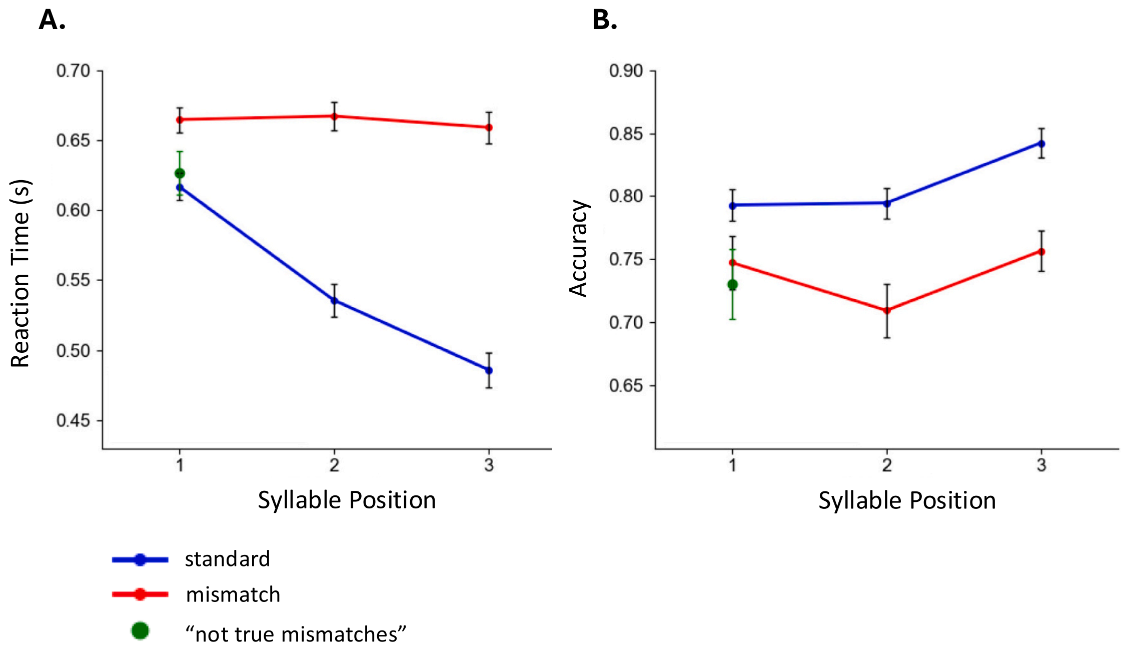

Fig. 2. Overall behavioural performance on the target detection task. A) Mean response times for syllables in standard words and mismatch words, as a function of syllable position. B) Mean accuracy as a function of condition and syllable position. The "not true mismatches" point only includes special cases in which regular first position syllables also appeared in the first position of a mismatch word and are included for visualization purposes only; these words were not included in our main analyses. The error bars represent the standard error of the mean.

function of triplet position (trend estimate = −5.4, SE = 4.26, p = 0.21). These results establish that facilitation to later-occurring positions within a triplet occurred only within standard words, with no effect of triplet position in the mismatch word condition. Although not effects of direct interest, we also found that RTs increased significantly as a function of target position within a stream (estimate for overall target position within the stream = 0.19 ms, SE = 0.029 ms; t(16060) = 6.32, p < 0.001), and also decreased as the task progressed (estimate for stream number: −0.28 ms, SE = 0.086, t(16060) = −3.28, p = 0.001). Interestingly, in a separate targeted analysis, we also found that mismatch syllables were significantly slower overall than even word-initial syllables (i.e., the least predictable syllables) within standard words (F(1,29) = 61.2, p < 0.001).

In addition, a separate model that included Block as a predictor showed that the triplet position effect for standard words significantly increased over the course of the task, suggesting that some additional learning occurred over the course of the target detection task itself (Triplet Position x Block: t(16050) = −3.43, p < 0.001; see Supplementary Materials and Supplementary Fig. 1).

3.2. Evidence for a prediction trade-off

Next, we evaluated whether participants who show proportionately greater facilitation to predictable syllables occurring in standard words also show a stronger *cost* in processing syllables that unexpectedly occur in mismatch words. We found that this was indeed the case. The RT prediction effect showed a strong correlation with the RT mismatch cost (r = 0.72, p < 0.001; Fig. 3). This finding indicates that participants who experienced the greatest facilitation for highly predictable syllables also showed the largest cost for syllables occurring outside of their usual contexts.

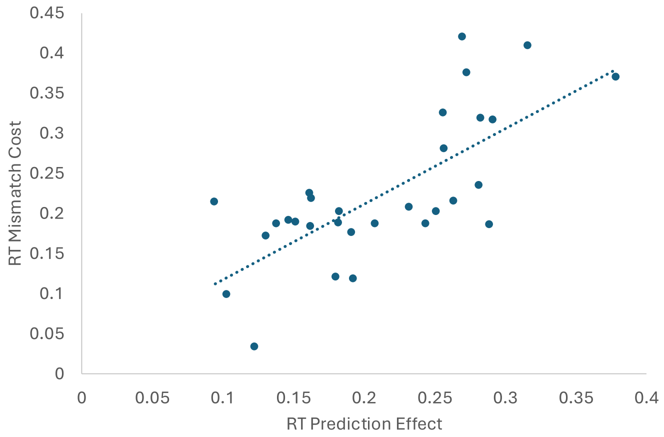

Fig. 3. Correlation at the individual participant level, showing relation between the RT Prediction Effect and the RT Mismatch Cost. Participants showing relatively stronger facilitation for standard words (greater RT Prediction effect) also showed correspondingly slower responses to syllables in the mismatch condition (as indexed by stronger RT Mismatch costs).

We further examined whether this trade-off, whereby facilitation for predictable syllables is associated with a cost in processing mismatch syllables, was present at the individual syllable level. Controlling for each participant's mean RT across standard syllables, we found that the average RTs to a given syllable occurring within a predictable position in a standard word (triplet position 2 and 3) negatively correlated with its mean RT occurring within a mismatch word (position 3: r = −0.41, p < 0.001; position 2: r = −0.23, p = 0.014). The effect was numerically larger in magnitude for position 3 syllables compared to position 2, though not significantly so (z = −0.74, p = 0.23). In contrast, syllables occurring in word-initial positions within standard words (i.e., a position that is not predictable) showed positive RT correlations with their occurrences in mismatch words (r = 0.20, p = 0.029). These results provide additional evidence of a cost of prediction; the more facilitated a given syllable is when it occurs within a predictable context (position 2 and 3), the more inhibited that syllable is when it occurs outside of its predictable context. In contrast, for initial-position syllables that cannot be specifically predicted, RTs are positively correlated between standard and mismatch occurrences. This positive association could at least partially reflect perceptual factors associated with the individual syllable (e.g., perceptual discriminability) that would be consistent across contexts. Overall, these findings provide evidence for our hypothesis that there should be a trade-off between ease of processing expected syllables and cost of processing unexpected syllables, at both the participant and individual syllable level.

3.3. Positive correlation in RTs to syllables within the same word

Next, we examined whether facilitation effects associated with statistical learning reflect learning of specific syllable-to-syllable transitions with a word. For predictable syllables in standard words (syllables occurring in positions 2 and 3), there was a significant positive relationship between the RT to a given syllable and its preceding neighbour, such that faster RTs to the preceding syllable corresponded to faster RTs to the current syllable (estimate = 0.27, SE = 0.038, $p < 0.001$). This relationship held for both position 2 (estimate = 0.50, SE = 0.069, p < 0.001) and position 3 syllables separately (estimate = 0.32, SE 0.047, p < 0.001). Furthermore, conducting the same analysis but using the mean RT to the syllable occurring in *two* previous positions as a predictor, a positive relationship was also found between position 1 and position 3 syllables within the same word (estimate = 0.19, SE = 0.06, p = 0.004). This result indicates that the faster detection of one syllable in a word predicts faster detection of subsequent syllables within the same word. In other words, taking as an example the word “bafuko”, a participant who is relatively fast in responding to “ba” and “fu” (relative to their own RT baseline) will also be fast in responding to “ko.” We confirmed this result using a separate nonparametric permutation test, in which for each of 1000 iterations we randomly shuffled the labels of each syllable with other syllables sharing the same triplet position, and then assessed the mean RT correlations for syllables occurring within the same (randomly created) word. This result showed that the true observed average within-word RT correlation was significantly greater than the correlations that emerge for syllables in randomly ordered words that still control for triplet position ($p < 0.001$). These results are compatible with a prediction account, whereby the presence of an initial syllable within a word (e.g., “ba”) may help to signal the upcoming presence of the subsequent syllables (e.g., “fu” and “ko”), such that faster detection of earlier syllables leads to corresponding faster RTs to later syllables.

3.4. Accuracy

As we would expect, targets occurring within standard words were detected significantly better than those within mismatch words (81.0 % versus 73.8 % detection rate; effect of word type: F(1,29) = 43.7, $p <$ 0.001; Fig. 2B). There was also an overall effect of position on accuracy rates, with accuracy rates increasing for later-occurring syllables within a word (effect of position: F(2,58) = 9.60, p < 0.001). Word type and word position did not significantly interact (F(2,58) = 2.03, p = 0.14). However, linear contrasts indicated that there was a linear trend by syllable position for standard words F(1,29) = 27.1, $p < 0.001$, but not for mismatch words (F(1,29) = 0.21, p = 0.65).

3.5. Familiarity rating task

As expected, words were rated as most familiar, followed by part-words, with non-words being rated as the least familiar (Effect of Word Type: F(2,58) = 31.0, $p < 0.001$, word > partword: t(58) = 5.80, p < 0.001; word > nonword: t(58) = 7.51, p < 0.001; Fig. 4A). This result indicates that participants acquired explicit knowledge of the words.

3.6. Recognition task

Participants performed significantly above-chance on this task, with a mean accuracy of 75.4 % (SD = 13.8 %; (t(29) = 10.1, $p < 0.001$; Fig. 4B). Again, this finding provides evidence that participants acquired explicit knowledge of the words.

3.7. Correlations in performance across tasks

The RT prediction effect did not correlate with our performance measures on either of the two explicit tasks (recognition accuracy: r = 0.23, p = 0.23; familiarity rating score: r = 0.13, p = 0.49), suggesting that RT facilitation is not strongly associated with explicit knowledge. Likewise, the RT mismatch cost also did not correlate with performance on the explicit tasks (recognition accuracy: r = 0.21, p = 0.27; familiarity rating score: r = 0.30, p = 0.095). By contrast, a positive correlation was found between the familiarity rating score and recognition accuracy on 2AFC recognition task (r = 0.41, p = 0.026), which is in line with our expectations, given that both tasks are considered explicit measures of statistical learning.

At the item level, a somewhat different picture emerged. On the recognition task, higher accuracy for a given word was associated with faster RTs to the word’s third syllable (occurring within standard words; r = −0.33, $p < 0.001$). By comparison, no significant relationship occurred between recognition accuracy at the item level and RTs for second or first syllables within a given word (2nd syllables: r = −0.058, p = 0.64; 1st syllables: r = −0.087, p = 0.07). There was also no relationship between recognition accuracy to a given word and the RT

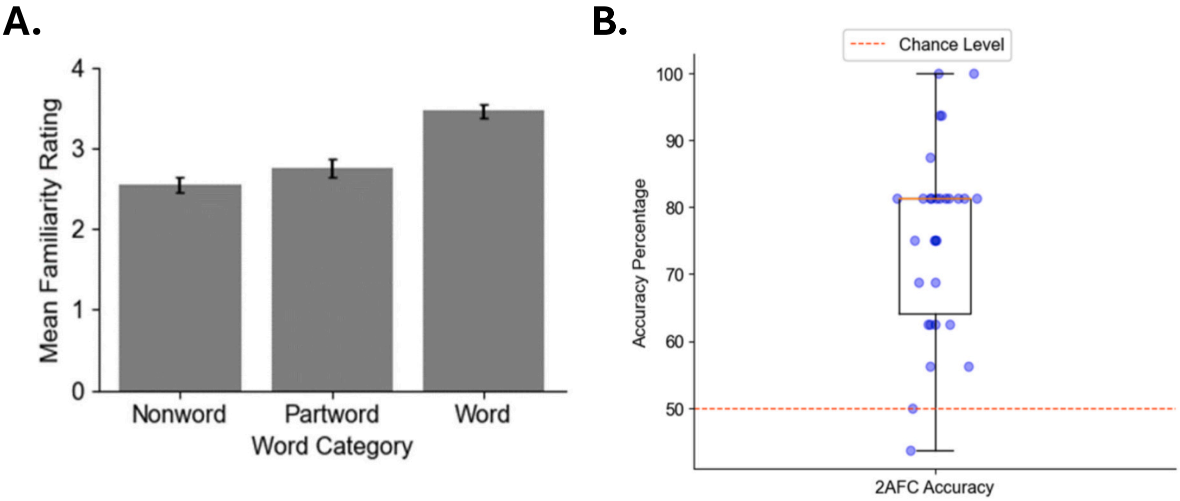

Fig. 4. Group-level performance on the explicit measures of statistical learning. A) Performance on the familiarity task, with rating scores range from 1 (not familiar at all) to 4 (very familiar). B) Individual performance on the 2AFC recognition task. Error bars represent the standard error.

mismatch cost for the word's predictable (2nd and 3rd) syllables (r = 0.094, p = 0.75). This suggests that explicit knowledge, as assessed through recognition performance, is uniquely linked to faster RTs to third syllables.

Nonetheless, when considering only "unknown words" (i.e., the subset of words for which recognition was at chance level or below), we still observed a highly reliable facilitation effect to predictable syllables (RT prediction effect: M = 0.15, SD = 0.11, t(32) = 8.01, p < 0.001), as well as a highly significant RT cost in processing 2nd and 3rd syllables of these unknown words when they occurred within mismatch contexts (RT mismatch cost: 2nd syllables: M = 76 ms; SD = 74 ms; t(32) = 5.90, p < 0.001; 3rd syllables: M = 75 ms, SD = 76 ms; t(32) = 6.82, p < 0.001). Overall, the findings show that, although explicit knowledge is associated with stronger priming effects to third syllables at the item level, robust prediction effects occur even in the absence of explicit knowledge.

4. Discussion

Our findings provide novel behavioural evidence for the idea that genuine prediction is an important consequence of statistical learning. By probing for a trade-off in RTs between expected and unexpected syllables during a target detection task after an initial statistical learning period, we examined whether statistical learning produces representations of regularities that can be used prospectively. In line with the idea that prediction is supported by statistical learning, we found that participants who show the strongest prediction effects – as assessed by relatively faster RTs to third syllable targets compared to first syllable targets – show the greatest cost when syllables occur outside of their usual context, in mismatch words. Similarly, at the individual syllable level, syllables that were more strongly facilitated when they occur within their typical position in standard words were more difficult to detect when they occur in a mismatch context. These findings do not follow from a retrospective processing model, but are directly supported by a prediction account: the more weight a given prediction is given, the greater the enhancement in processing a confirmed prediction, but the greater the cost in processing a violation of the prediction (Van Petten & Luka, 2012). Also in line with a prediction account is the finding of positive correlations in RTs between syllables within the same word, suggesting that earlier-occurring syllable in a word may provide a cue for later-occurring syllables. Finally, we found evidence that explicit knowledge of a given word is positively correlated with response times to third-position syllables, but not to other aspects of prediction, and is not necessary for prediction to occur.

The trade-off in processing expected versus unexpected syllables was evident at both the participant and the item level. At the participant level, a highly robust correlation was observed between the RT prediction effect, indexing relative facilitation to predictable syllables within standard words, and the RT mismatch cost, indexing the relative cost associated with processing unexpected syllable. In other words, individuals who showed the strongest facilitation effects within standard words were the most impaired at processing syllables that occurred in unexpected contexts. This result suggests that participants vary systematically in the degree to which they engage predictive mechanisms during online processing of the speech stream, which in turn confers both benefits and costs for online processing of individual syllables, depending on whether the prediction is correct. Why some participants engage more strongly in prediction than others is not yet clear, and remains a direction for future research. A processing trade-off between expected and unexpected syllables was also observed at the specific syllable level. After accounting for participants' baseline RTs (i.e., the fact that some participants have generally faster response times than others), we showed that the faster the response to an individual syllable occurring within its expected position, the slower the response to the syllable within a mismatch word. It appears that each individual participant anticipates some idiosyncratic subset of syllables more strongly than others, with more strongly predicted syllables being more difficult to process in mismatch contexts.

Interestingly, better explicit memory for a given word, as assessed through recognition performance, was associated with faster RTs to the word's final syllable. This result suggests that explicit knowledge of a given regularity may boost prediction of the final syllable over and above "baseline" priming effects, and echoes our previous finding showing that participants who were explicitly trained on the regularities responded significantly more quickly to final-position targets compared to participants who were learned the regularities through a typical statistical learning paradigm (Batterink, Reber, & Paller, 2015). In that study, we also found that explicitly trained participants made a significant number of "anticipatory" or early responses to targets occurring in the third position. In this context, it is worth noting that in the current study, participants generally achieved better recognition performance (~75 %) than in many previous studies of statistical learning, which has been previously estimated to average around 60 % (Isbilen, McCauley, Kidd, & Christiansen, 2020;)see(Batterink & Paller, 2017; Batterink, Reber, Neville, & Paller, 2015; Batterink, Reber, & Paller, 2015; Franco, Cleeremans, & Destrebecqz, 2011; Siegelman & Frost, 2015; Smalle, Daikoku, Szmalec, Duyck, & Möttönen, 2022)for examples of studies with 2 AFC performance <70 %). It may be that for words that are most strongly learned, participants are sometimes able to explicitly make predictions about upcoming syllables. However, and in contrast to the final position effect, explicit knowledge did not predict facilitation of second syllable targets, which are also predictable, nor the mismatch cost for a given word's predictable syllables (occurring in either the second or final position). In addition, the syllables composing unknown words (words for which performance on the recognition task was at or below chance) still elicited robust RT prediction effects and mismatch costs. These findings are consistent with other studies that have reported dissociations between implicit and explicit measures of auditory statistical learning (Batterink, Reber, Neville, & Paller, 2015; Franco et al., 2015; Kim et al., 2009; Krishnan, Carey, Dick, & Pearce, 2022; Liu et al., 2023). In particular, this result aligns with our previous finding that both known and unknown words, as measured by recognition performance, elicit similar levels of RT priming on the target detection task (Batterink et al., 2015). Overall, these findings indicate that the contribution of explicit knowledge uniquely facilitates processing of the third-position targets, rather than other aspects of prediction, and suggests that explicit knowledge per se is not the main driver of prediction effects.

4.1. Attentional mechanisms underlying prediction

Other aspects of our results suggest that prediction—at least at the level observed in our RT measure within the target detection task—is not elicited automatically but may be modulated by selective attention to individual targets at the trial level. As shown in Fig. 2A, RTs to syllables in mismatch words were stable across the three triplet positions. In other words, once a syllable occurred outside of its expected context, it was processed with similar difficulty regardless of whether it replaced the first, second or third syllable of another word. Given that participants had to detect just one specific assigned target for each stream, this result suggests that participants did not generate ongoing predictions for task-irrelevant syllables that were not generally part of the relevant constituent word that would typically precede the assigned target syllable. If participants had formed real-time predictions for all syllables contained in the speech stream, we would have expected to see that targets occurring within the 2nd and 3rd positions of a mismatch word were progressively more delayed relative to the initial position, given that participants would have to overcome their expectations of the usual syllable in these positions to successfully detect the swapped mismatch syllable. The fact that this result was not observed suggests that participants selectively "upregulated" or attended only to the particular syllables that they expected to immediately precede the relevant target.

For example, if participants were asked to detect the syllable "su" (which occurs within the word "fetisu"), they may selectively attend to the syllables "fe" and "ti" as they occur within the stream, and may ignore or downregulate the other syllables that are not part of this "syllable family." Overall, these findings suggest that prediction is "target-centric," rather than occurring indiscriminately to all syllables, at least in the context of the target detection task used here. Given the finding that our prediction effects – both the RT prediction effect and the RT mismatch cost – occur robustly to unknown words, it appears that the engagement of prediction does not require conscious, explicit knowledge of the regularities, and thus is unlikely to represent an intentional or top-down strategy employed by participants. Rather, this type of mechanism may reflect an implicit "priority map" operating over words in the temporal domain, perhaps analogous to spatial attentional priority maps that emerge following the implicit learning of spatial regularities (Duncan, Van Moorselaar, & Theeuwes, 2023).

If attention is automatically deployed to earlier occurring syllables occurring within the same constituent word as a target syllable, this could also explain our observed positive RT relation between syllables in the same word. By this account, initial syllables that are more readily perceptible within the stream would themselves be easier to detect, and would also speed responses to subsequent syllables by providing a more rapidly available cue to the upcoming target syllable. A similar finding was reported in a recent statistical learning study, which used a representational similarity analysis of reaction-time data and reported greater representational similarity for syllables within the same word grouping compared to syllables from different words (Kiai & Melloni, 2021). Positive correlations for syllables within the same word would be expected only if participants had extracted something about the relationship between the three syllables of a word, and thus can serve as a marker of word-specific learning. Alongside the overall progressive reduction in RTs to later syllables within a triplet (Fig. 2A), this finding also provides additional evidence of statistical learning. Incidentally, the finding that RTs differ between syllables of the same triplet position cannot be explained by a pure statistical learning account that considers only transitional probabilities, but could be driven by language-related factors (such as resemblance between the word and participants' native language; e.g., Elazar et al., 2022; Siegelman et al., 2018) or perceptual factors such as syllable discriminability.

To further understand how prediction at the item level may contribute to facilitation effects in statistical learning, future studies could leverage event-related potentials or other neuroimaging measures to probe for neural evidence of neural pre-activation in the window prior to target onset. For example, in the current task, participants are asked to detect a predefined target syllable that differs across streams. This variable target assignment on each stream would allow for a direct comparison of neural responses occurring prior to targets versus non-targets, while keeping the syllables themselves identical between conditions. Based on the current behavioural results and interpretation, we would expect to see an enhanced neural response to earlier syllables preceding a target syllable, relative to the exact same syllables occurring in other streams where a different syllable is serving as a target. Such an effect would provide evidence that prediction is deployed specifically for syllables that are relevant to an upcoming target syllable, rather than engaged throughout processing. In contrast, if we saw no differences in the prestimulus window between targets and nontargets, this would call our current interpretation into question, suggesting that perhaps prediction is not specific to target identity. Additional approaches, such as representational similarity analysis, would allow for assessing whether prediction that results from speech-based statistical learning serves to broadly enhance attention to upcoming input or whether it may specifically preactivate representations of yet-to-be-presented syllables, as suggested by studies of visual statistical learning (Sherman et al., 2022; Sherman & Turk-Browne, 2020).

Our conclusions are necessarily limited by our measures. While the current results provide evidence for a predictive mechanism that is shaped by the trial-by-trial demands of the target detection task, prediction operates at multiple levels (Bar, 2009; Bubic., 2010; De Lange et al., 2018) and it is possible that our RT-based measure is insensitive to some aspects of prediction that may occur indiscriminately to all syllables and/or across different task contexts. To further understand how prediction may result from statistical learning, another direction for future research would be to compare neural markers of prediction during passive listening versus an active detection task. Sensitive brain-based measures may reveal more subtle prediction effects to upcoming syllables that are not captured by our behavioural RT measure. In other words, the current results leave open the possibility that there are aspects of prediction supported by statistical learning that may operate more automatically, independently of task demands.

Finally, it is interesting to note that reaction times to syllables in mismatch words were significantly slower even compared to syllables occurring in the initial position of standard words, generally considered to be unpredictable (Fig. 2A). This finding highlights that in the context of artificial language studies with highly limited number of words, even the initial syllables of a word are still weakly predictive (Endress, 2024). In the current study, a given word-initial syllable could only follow one of the three word-final syllables from the other words, given that there were only four words in the stream and that the same word could not repeat consecutively. One implication of these findings is that RT-based measures of statistical learning, which typically quantify learning effects as the difference in RTs between initial and final items of the learned units, may actually be underestimating statistical learning effects. Although RT effects are generally robust and more consistent at the individual level compared to explicit measures of learning (Batterink, Reber, Neville, & Paller, 2015; Kiai & Melloni, 2021), the inclusion of a 'mismatch' condition in RT-based tasks could be worth pursuing as a future methodological improvement, which may produce an even more robust measure of statistical learning.

5. Conclusions

To summarize, by relying on a behavioural hallmark of prediction in which any prediction leads to either benefits or costs depending on whether it is accurate (Van Petten & Luka, 2012), we show that genuine prediction emerges as a consequence of statistical learning. Our results further suggest that the engagement of prediction is constrained by task demands, rather than occurring indiscriminately to all stimuli in input. We show that learners make predictions selectively, when it is useful to do so, suggesting that prediction is selectively and dynamically employed after regularities are acquired through statistical learning. Overall, our findings highlight the functional utility of statistical learning and the flexible and adaptive nature of representations acquired through this form of learning.

CRediT authorship contribution statement

Laura J. Batterink: Writing – original draft, Visualization, Supervision, Funding acquisition, Formal analysis, Conceptualization. **Sarah Hsiung:** Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis. **Daniela Herrera-Chaves:** Writing – review & editing, Methodology, Investigation, Formal analysis, Conceptualization. **Stefan Köhler:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgements

This research was supported by a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant (2019-05132) to Laura

Batterink.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2025.106088.

Data availability

All data and experimental task files associated with this manuscript are available on Open Science Framework: https://osf.io/a4xue/?view_only=0c68ead882e444229cff853f27fbbccc).

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language & Cognitive Processes, 28*(4), 525–560. https://doi.org/10.1080/01690965.2012.658072

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to Reading ability in children and adults. *Cognitive Science, 36*(2), 286–304. https://doi.org/10.1111/j.1551-6709.2011.01200.x

Aslin, R. N. (2017). Statistical learning: A powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science, 8*(1–2). https://doi.org/10.1002/wcs.1373

Barakat, B. K., Seitz, A. R., & Shams, L. (2013). The effect of statistical learning on internal stimulus representations: Predictable items are enhanced even when not predicted. *Cognition, 129*(2), 205–211. https://doi.org/10.1016/j.cognition.2013.07.003

Bar, M. (2009). Predictions: a universal principle in the operation of the human brain. Introduction. *Philos Trans R Soc Lond B Biol Sci., 364*(1521), 1181–1182. https://doi.org/10.1098/rstb.2008.0321. PMID: 19527998; PMCID: PMC2666718.

Bates, E., & Elman, J. (1996). Learning rediscovered. *Science, 274*, 1849–1850.

Batterink, L. J. (2017). Rapid statistical learning supporting word extraction from continuous speech. *Psychological Science, 28*(7), 921–928. https://doi.org/10.1177/0956797617698226

Batterink, L. J., Mulgrew, J., & Gibbings, A. (2024). Rhythmically modulating neural entrainment during exposure to regularities influences statistical learning. *Journal of Cognitive Neuroscience, 36*(1), 107–127. https://doi.org/10.1162/jocn_a_02079

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

Batterink, L. J., & Paller, K. A. (2019). Statistical learning of speech regularities can occur outside the focus of attention. *Cortex, 115*, 56–71. https://doi.org/10.1016/j.cortex.2019.01.013

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

Batterink, L. J., Reber, P. J., & Paller, K. A. (2015). Functional differences between statistical learning with and without explicit training. *Learning & Memory, 22*(11), 544–556. https://doi.org/10.1101/lm.037986.114

Batterink, L. J., & Zhang, S. (2022). Simple statistical regularities presented during sleep are detected but not retained. *Neuropsychologia, 164*, Article 108106. https://doi.org/10.1016/j.neuropsychologia.2021.108106

Bays, B. C., Turk-Browne, N. B., & Seitz, A. R. (2015). Dissociable behavioural outcomes of visual statistical learning. *Visual Cognition, 23*(9–10), 1072–1097. https://doi.org/10.1080/13506285.2016.1139647

Bertels, J., Franco, A., & Destrebecqz, A. (2012). How implicit is visual statistical learning? *Journal of Experimental Psychology. Learning, Memory, and Cognition, 38*(5), 1425–1431. https://doi.org/10.1037/a0027210

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science, 19*(7), 678–685. https://doi.org/10.1111/j.1467-9280.2008.02142.x

Bubic.. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2010.00025

Campbell, K. L., Zimerman, S., Healey, M. K., Lee, M. M. S., & Hasher, L. (2012). Age differences in visual statistical learning. *Psychology and Aging, 27*(3), 650–656. https://doi.org/10.1037/a0026780

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

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

Dale, R., Duran, N., & Morehead, R. (2012). Prediction during statistical learning, and implications for the implicit/explicit divide. *Advances in Cognitive Psychology, 8*(2), 196–209. https://doi.org/10.5709/acp-0115-z

De Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences, 22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Duncan, D. H., Van Moorselaar, D., & Theeuwes, J. (2023). Pinging the brain to reveal the hidden attentional priority map using encephalography. *Nature Communications, 14*(1), 4749. https://doi.org/10.1038/s41467-023-40405-8

Elazar, A., Alhama, R. G., Bogaerts, L., Siegelman, N., Baus, C., & Frost, R. (2022). When the "tabula" is anything but "rasa:" what determines performance in the auditory statistical learning task? *Cognitive Science, 46*(2), Article e13102. https://doi.org/10.1111/cogs.13102

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

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 28*(3), 458–467. https://doi.org/10.1037//0278-7393.28.3.458

Franco, A., Cleeremans, A., & Destrebecqz, A. (2011). Statistical learning of two artificial languages presented successively: How conscious? *Frontiers in Psychology, 2*. https://doi.org/10.3389/fpsyg.2011.00229

Franco, A., Eberlen, J., Destrebecqz, A., Cleeremans, A., & Bertels, J. (2015). Rapid serial auditory presentation: A new measure of statistical learning in speech segmentation. *Experimental Psychology, 62*(5), 346–351. https://doi.org/10.1027/1618-3169/a000295

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*(7), Article e12848. https://doi.org/10.1111/cogs.12848

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Lang & Ling Compass, 2* (4), 647–670. https://doi.org/10.1111/j.1749-818X.2008.00072.x

Kiai, A., & Melloni, L. (2021). *What canonical online and offline measures of statistical learning can and cannot tell us* (p. 2021.04.19.440449). *bioRxiv*. https://doi.org/10.1101/2021.04.19.440449

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters, 461*(2), 145–149. https://doi.org/10.1016/j.neulet.2009.06.030

Kray, J., Sommerfeld, L., Borovsky, A., & Häuser, K. (2024). The role of prediction error in the development of language learning and memory. *Child Development Perspectives, 18*(4), 190–203. https://doi.org/10.1111/cdep.12515

Krishnan, S., Carey, D., Dick, F., & Pearce, M. T. (2022). Effects of statistical learning in passive and active contexts on reproduction and recognition of auditory sequences. *Journal of Experimental Psychology: General, 151*(3), 555–577. https://doi.org/10.1037/xge0001091

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M., Federmeier, K. D., & Urbach, T. P. (2014). The "negatives" and "positives" of prediction in language. In M. S. Gazzaniga, & G. R. Mangun (Eds.), *The cognitive neurosciences* (5th ed.). The MIT Press. https://doi.org/10.7551/mitpress/9504.003.0071.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience, 25*(3), 484–502. https://doi.org/10.1162/jocn_a_00328

Liu, H., Forest, T. A., Duncan, K., & Finn, A. S. (2023). What sticks after statistical learning: The persistence of implicit versus explicit memory traces. *Cognition, 236*, Article 105439. https://doi.org/10.1016/j.cognition.2023.105439

Luo, M., Cao, R., & Wang, F. H. (2024). The speed of detection vs. segmentation from continuous sequences: Evidence for an anticipation mechanism for detection through a computational model. *eLife*, 13. https://doi.org/10.7554/eLife.95761.1

Mitchel, A. D., & Weiss, D. J. (2011). Learning across senses: Cross-modal effects in multisensory statistical learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 37*(5), 1081–1091. https://doi.org/10.1037/a0023700

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

Musz, E., Weber, M. J., & Thompson-Schill, S. L. (2015). Visual statistical learning is not reliably modulated by selective attention to isolated events. *Attention, Perception, & Psychophysics, 77*(1), 78–96. https://doi.org/10.3758/s13414-014-0757-5

Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 15*(6), 1003–1019. https://doi.org/10.1037//0278-7393.15.6.1003

Ordin, M., Polyanskaya, L., & Samuel, A. G. (2021). An evolutionary account of intermodality differences in statistical learning. *Annals of the New York Academy of Sciences, 1486*(1), 76–89. https://doi.org/10.1111/nyas.14502

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behav Res Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y. PMID: 30734206; PMCID: PMC6420413.

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology, 32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Posner, & Snyder, C. R. R. (1975). Attention and cognitive control. In *Information processing and cognition* (pp. 55–85). Erlbaum.

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

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*(4), 606–621. https://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 corner of your ear. *Psychological Science, 8*(2), 101–105. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x

Sherman, B. E., Graves, K. N., Huberdeau, D. M., Quraishi, I. H., Damisah, E. C., & Turk-Browne, N. B. (2022). Temporal dynamics of competition between statistical learning and episodic memory in intracranial recordings of human visual cortex. *Journal of Neuroscience, 42*(48), 9053–9068. https://doi.org/10.1523/JNEUROSCI.0708-22.2022

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*(37), 22760–22770. https://doi.org/10.1073/pnas.2013291117

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition, 177*, 198–213. https://doi.org/10.1016/j.cognition.2018.04.011

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language, 81*, 105–120. https://doi.org/10.1016/j.jml.2015.02.001

Smalle, E. H. M., Daikoku, T., Szmalec, A., Duyck, W., & Möttönen, R. (2022). Unlocking adults' implicit statistical learning by cognitive depletion. *Proceedings of the National Academy of Sciences, 119*(2), Article e2026011119. https://doi.org/10.1073/pnas.2026011119

Sweet, S. J., Van Hedger, S. C., & Batterink, L. J. (2024). Of words and whistles: Statistical learning operates similarly for identical sounds perceived as speech and non-speech. *Cognition, 242*, Article 105649. https://doi.org/10.1016/j.cognition.2023.105649

Turk-Browne, N. B. (2012). Statistical learning and its consequences. In *, 59. Nebraska symposium on motivation* (pp. 117–146). https://doi.org/10.1007/978-1-4614-4794-8_6

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

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience, 30*(33), 11177–11187. https://doi.org/10.1523/JNEUROSCI.0858-10.2010

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Wang, H. S., Köhler, S., & Batterink, L. J. (2023). Separate but not independent: Behavioral pattern separation and statistical learning are differentially affected by aging. *Cognition, 239*, Article 105564. https://doi.org/10.1016/j.cognition.2023.105564

Wang, H. S., Rosenbaum, R. S., Baker, S., Lauzon, C., Batterink, L. J., & Köhler, S. (2023). Dentate gyrus integrity is necessary for behavioral pattern separation but not statistical learning. *Journal of Cognitive Neuroscience, 35*(5), 900–917. https://doi.org/10.1162/jocn_a_01981

Yap, M. J., Hutchison, K. A., & Tan, L. C. (2017). Individual differences in semantic priming performance: Insights from the semantic priming project. In *Big data in cognitive science* (pp. 203–226). Routledge/Taylor & Francis Group.