

Beta-Band Activity Is a Signature of Statistical Learning

2035

 Louisa Bogaerts,¹  Craig G. Richter,²  Ayelet N. Landau,¹ and  Ram Frost^{1,2,3}

¹Department of Psychology, The Hebrew University, Jerusalem, 91905, Israel, ²Basque Center on Cognition, Brain and Language, Donostia-San Sebastian, 2009, Spain, and ³Haskins Laboratories, New Haven, Connecticut 06511

Through statistical learning (SL), cognitive systems may discover the underlying regularities in the environment. Testing human adults ($n = 35$, 21 females), we document, in the context of a classical visual SL task, divergent rhythmic EEG activity in the interstimulus delay periods within patterns versus between patterns (i.e., pattern transitions). Our findings reveal increased oscillatory activity in the beta band (~ 20 Hz) at triplet transitions that indexes learning: it emerges with increased pattern repetitions; and importantly, it is highly correlated with behavioral learning outcomes. These findings hold the promise of converging on an online measure of learning regularities and provide important theoretical insights regarding the mechanisms of SL and prediction.

Key words: electroencephalography; neurobiological signature; prediction; statistical learning

Significance Statement

Statistical learning has become a major theoretical construct in cognitive science, providing the primary means by which organisms learn about regularities in the environment. As such, it is a critical building block for basic and higher-order cognitive functions. Here we identify, for the first time, a spectral neural index in the time window before stimulus presentation, which evolves with increased pattern exposure, and is predictive of learning performance. The manifestation of learning that is revealed, not in stimulus processing but in the blank interval between stimuli, makes a direct link between the fields of statistical learning on the one hand and either prediction or consolidation on the other hand, suggesting a possible mechanistic account of visual statistical learning.

Introduction

Sensory information is often structured, both in time and in space. Statistical learning (SL) is the ability to incidentally discover regularities in the environment. Seminal work by Saffran et al. (1996) demonstrated that infants can extract syllable patterns presented in a continuous speech stream based solely on the extent of the syllables' co-occurrences. Ever since, SL has been extensively investigated and was documented in individuals of all ages, over different types of stimuli and sensory modalities (for a recent review, see Frost et al., 2019). Similar to linguistic information, our visual environment contains

extensive statistical regularities (e.g., probabilistic relations between objects, prevalent sequences of letters, etc.). Indeed, recent studies of visual SL demonstrated robust learning of temporal relationships among sequentially presented ordered stimuli (Fiser and Aslin, 2002; Kirkham et al., 2002; Bogaerts et al., 2016). SL is unsupervised and may occur without intent or awareness of the structured nature of the input (Aslin et al., 1998; Fiser and Aslin, 2001; Turk-Browne et al., 2005).

While SL has become an important construct in cognitive science, little is known about the neural mechanisms underlying it. Most neuroimaging studies that identified brain regions sensitive to the statistical regularities in sensory input have summarized brain activity during structured versus unstructured blocks, lasting several seconds or minutes. These studies associated SL effects with domain-general regions involved in binding temporal and spatial contingencies, such as the hippocampus, medial temporal lobe, and inferior frontal gyrus (e.g., Turk-Browne et al., 2009; Karuza et al., 2013; Shohamy and Turk-Browne, 2013). Investigating item-specific hemodynamic responses, Turk-Browne et al. (2010) also reported increased hippocampal activity in response to visual stimuli which predict subsequent visual stimuli. At the same time, imaging work identified regions in the early visual and auditory cortices that are sensitive to regularities in vision and audition (e.g., McNealy et al., 2006; Turk-Browne et al., 2009; for discussion, see Frost et al., 2015). A handful of fMRI studies did go beyond the neurobiological "where in the

Received Apr. 1, 2020; revised July 26, 2020; accepted Aug. 4, 2020.

Author contributions: L.B., A.N.L., and R.F. designed research; L.B. and A.N.L. performed research; L.B., C.G.R., and A.N.L. analyzed data; L.B. wrote the first draft of the paper; L.B., C.G.R., A.N.L., and R.F. edited the paper; L.B. and R.F. wrote the paper.

The authors declare no competing financial interests.

This work was supported by ERC Advanced Grant Project 692502-L2STAT to R.F., Marie Skłodowska-Curie Grant 743528 (IF-EF, European Union's Horizon 2020 Research and Innovation Program) to L.B., the Basque Government BERG 2018-2021 program, and the Spanish State Research Agency BCBL Severo Ochoa excellence accreditation SEV-2015-0490. We thank Limor Shedlesky, Cassi Gewer, and Michelle Schechter for help with subject recruitment and data collection.

L. Bogaerts' present address: Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, 1081BT Amsterdam, the Netherlands.

Correspondence should be addressed to Louisa Bogaerts at bog.louisa@gmail.com.

<https://doi.org/10.1523/JNEUROSCI.0771-20.2020>

Copyright © 2020 the authors

brain” of visual SL. For example, Schapiro et al. (2012) demonstrated that representations of visual elements in the human medial temporal lobe converge when they frequently co-occur. In the same vein, Turk-Browne et al. (2010) found greater hippocampal hemodynamic responses to predictive stimuli in a patterned stream, thus providing evidence for implicit anticipation.

The current study aimed to deepen our understanding of the underlying neurobiological mechanisms of visual SL, investigating whether brain rhythms play a role in the implicit learning of statistical regularities. So far, attempts to identify neural indices of SL with electrophysiology centered on the consequences that learning has on stimulus processing. Consequences of learning were typically assessed by contrasting evoked responses to stimuli that are predictable to those of stimuli that cannot be predicted (*N400* in the visual domain: Abia and Okanoya, 2009; *N400*, *N100*, *P300*, and *P200* in the auditory domain: e.g., Sanders et al., 2002; Cunillera et al., 2006; De Diego Balaguer et al., 2007; Abia et al., 2008; Batterink et al., 2015; for a recent review, see Daikoku, 2018). Abia and Okanoya (2009), for example, demonstrated that participants who show behavioral evidence for good learning have an increased *N400* in response to unpredictable shapes in a sequence, compared with the middle and final shapes within a sequence, which are fully predictable. ERPs were also shown to differentiate between visual stimuli that carry high versus low predictability for a subsequent target. Highly reliable predictors were shown to elicit either a larger centro-parietal late positivity (Daltrozzo et al., 2017) or a larger *P300* component (Jost et al., 2015). More recently, the predictability of visual stimuli was shown to modulate the low-frequency activity associated with stimulus presentation: compared with predictable second items of a learned pair, unexpected shapes elicited stronger activity in the alpha range (7–14 Hz) (Zhou et al., 2019). Auditory SL has also been indexed considering longer segments of structured versus unstructured input, by quantifying the neural entrainment to the temporal structure created by repeating regular sequences (Batterink and Paller, 2017; Farthouat et al., 2017). This entrainment is postulated to reflect the perceptual binding of stimuli into familiar composites (Batterink and Paller, 2017).

In the present study, by contrast, our main interest was in electroencephalography (EEG) activity in the time window leading up to stimulus presentation, targeting the mechanisms of prediction. Underlying our approach are two notions on the nature of SL. First, the learning process itself is likely not a passive process; therefore, the online manifestations of learning need not be limited to stimulus processing but could impact anticipatory moments in the learning episode. Second, the outcome of learning, once again, might entail an active state of anticipation whereby learned regularities lead to active predictions (e.g., Engel et al., 2001; Turk-Browne et al., 2010; see also Tollman, 1932). In previous studies on SL, the process of prediction has often been implied but not operationally measured (e.g., Sanders et al., 2002; Abia and Okanoya, 2009). Our present investigation uses EEG to focus on the prestimulus epoch in the absence of stimulus processing, aiming to measure anticipation (and/or any other processes associated with pattern learning) directly. In doing so, we quantify spectral signatures of brain activity (i.e., brain rhythms). Indeed, experimental work with animals as well as humans has demonstrated that rhythmic brain activity from the delta to gamma range (1–100 Hz) has functional relevance for several sensory and cognitive processes (Buzsáki, 2016). For example, activity and oscillatory synchrony in the beta-frequency

(13–30 Hz) has been associated with sensory prediction (Arnal and Giraud, 2012) and top-down modulation (e.g., Hipp et al., 2011; for review, see Bressler and Richter, 2015). In addition, theta-band activity (4–7 Hz) has been associated with the categorical prediction of upcoming images (e.g., Cashdollar et al., 2017) and locations in visual-search displays (e.g., Spaak and de Lange, manuscript under review).

Here, we document, for the first time, divergent rhythmic prestimulus EEG activity within patterns versus between patterns (i.e., pattern transitions). We demonstrate increased power in the beta-band at pattern transitions, in the context of a typical visual SL task. We further find that this differential prestimulus beta-band activity is a signature of learning: we show that it emerges with increased pattern repetitions; and importantly, we show that it is highly correlated with behavioral learning outcomes.

Materials and Methods

Participants

Thirty-five healthy individuals (21 females) participated in the study for payment or for course credit. Participants had a mean age of 24.85 years (range 18–33 years) and reported normal or corrected-to-normal vision and no history of neurologic or psychiatric disease. Written informed consent was obtained from all participants in line with the institutional review board approval from the Hebrew University of Jerusalem.

Experimental design

The experiment consisted of a structured familiarization stream with embedded patterns, directly followed by a test, and of a random stream. The random stream either followed the test or preceded the structured stream (counterbalanced across participants).

The task included 24 abstract shapes (Fig. 1) in equiluminant dark orange ($R=205$, $G=85$, $B=50$), displayed on a gray background ($R=110$, $G=110$, $B=110$). The latent structure of the structured visual input stream was similar to that of multiple previously used SL tasks (e.g., Turk-Browne et al., 2005; Frost et al., 2013; Glicksohn and Cohen, 2013). Shapes were randomly organized for each participant to create 8 triplets, with a transitional probability of 1 between shapes within each triplet. The structured stream consisted of 54 repetition blocks, with all 8 triplets appearing once (in a random order) in each repetition block, with the constraint that a same triplet could not appear twice in a row. A self-paced break was included after every 6 repetition blocks, dividing the structured familiarization stream into 9 equal periods. Given our interest in anticipatory brain activation, we presented stimuli for a short duration (0.2 s) with a fixed interstimulus interval of 1.1 s. Before exposure to the structured stream, participants received the instruction “In this part there are shapes that follow each other, pay attention to the sequence. Following this part you will be asked questions about what you have seen” (see, e.g., Siegelman et al., 2018).

Following exposure to the structured stream, participants completed a test consisting of 32 two-alternative forced-choice questions. For each question, participants were instructed to select which of two possible triplet sequences they believed had occurred during the structured familiarization stream. For each question, participants were sequentially presented with the following: (1) a target: three shapes that formed a triplet in the structured stream; and (2) a foil: three shapes that never appeared in sequence in the structured stream. Foils were constructed without violating the position of the shapes within the triplets (e.g., given the triplets ABC, DEF, and GHI, a possible foil could be, e.g., AEI or GBF but not BID). Shapes making up a target or foil appeared sequentially with a fixed presentation rate and interstimulus interval as during exposure. A blank screen of 1.5 s separated the two three-item sequences. The offline test score, defined as the number of correct identifications of targets, ranged from 0 to 32. Chance performance corresponds to a score of 16 of 32.

The random stream consisted of 18 repetition blocks, with the same 24 shapes appearing once in each repetition block, in a random order (with timing parameters identical to those in the structured stream).

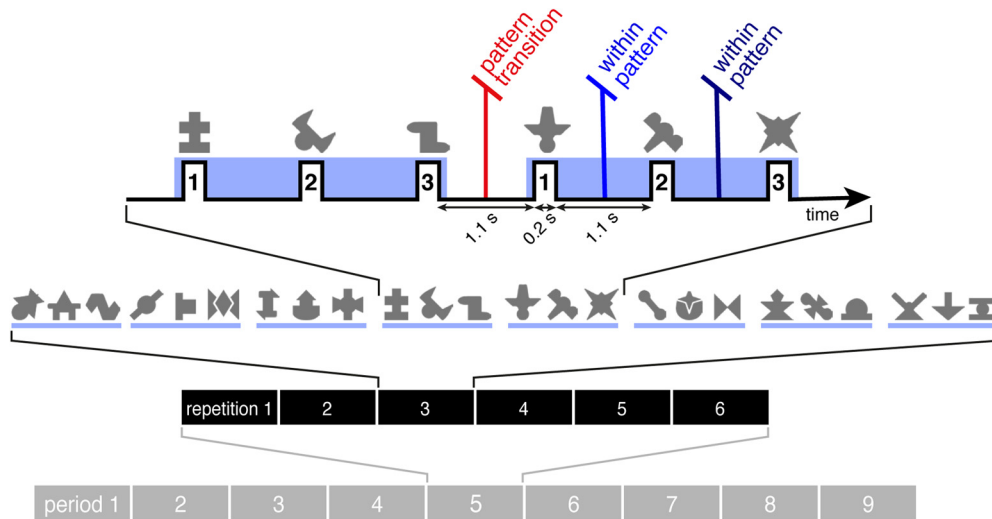


Figure 1. Schematic depiction of the structured familiarization stream, containing 9 exposure periods each consisting of 6 repetitions of 8 embedded triplets.

Before their exposure to the random stream, participants received the instructions: “In this part we want to test whether your brain recognizes individual shapes. We will therefore present the same shapes again and again. Your task is simply to watch the shapes attentively.” We opted for nonidentical instructions before the random versus structured stream given the evidence that exposure to a random stream can impair the subsequent learning of structure (e.g., Jungé et al., 2007; see also Turk-Browne et al., 2009). In order to be able to demonstrate predictive validity, we prioritized finding robust behavioral evidence for behavioral learning above absolute equality.

EEG recording

EEG was recorded from 64 Ag/AgCl active electrodes (g.tec medical engineering, 62 scalp electrodes and 2 electrode earclips). Four EOG electrodes were placed at the outer canthi of both eyes (horizontal EOG), and above and below the left eye (vertical EOG). Eye position and pupil area were monitored (binocularly) with a video-based eye tracker (Eyelink 1000, SR Research). A chinrest was used to reduce head movements. We used a Simulink model in MATLAB (2015a, The MathWorks) as the data acquisition software. EEG was recorded using a sampling frequency of 512 Hz, and eye data were coregistered by the model at that same sampling frequency.

EEG data preprocessing

All EEG analyses were conducted using the Fieldtrip toolbox (Oostenveld et al., 2011) for MATLAB (2016b, The MathWorks). EEG preprocessing was conducted using the following analysis pipeline: (1) Scalp signals were referenced to the average of the left and right earlobes. (2) Signals were bandpass filtered between 0.1 and 140 Hz with 50 and 100 Hz line noise removal. (3) The data were segmented into 1.3 s epochs (0.8 s before 0.5 s after shape onset). (4) We identified bad channels (mean number per participant = 1.86, SD = 2.21) and eliminated epochs containing large artifacts (including eye movement artifacts), based on an absolute amplitude threshold and a variance threshold using Fieldtrip’s artifact detection routines (mean threshold absolute amplitude = $237.06 \mu\text{V}$, SD = $103.30 \mu\text{V}$; mean threshold variance = $3381.243 \mu\text{V}$, SD = $3231.28 \mu\text{V}$). On average, 4.30% of epochs were removed for each participant (range 0.87%–12.15%). (5) Large muscle artifacts were identified using automatic artifact detection, and were subsequently visually inspected. Epochs containing confirmed muscle artifacts were rejected (average of 1.78% of epochs rejected, range 0.23%–5.79%). (6) Blink artifacts were corrected using independent component analysis (see Jung et al., 2000) (mean number of components = 1). (7) The data without large artifacts were cleaned a last time using the threshold approach of Step 4 (mean threshold absolute amplitude = $136.61 \mu\text{V}$, SD = 45.34; mean threshold range $212.35 \mu\text{V}$, SD = 75.72;

mean threshold variance = $2012.12 \mu\text{V}$, SD = 2567.40), rejecting on average 3.91% of epochs (range 0.64%–10.07%). This step also took care of smaller eye movement artifacts. (8) Finally, bad channels were interpolated using the average of all neighbors (with regularization parameter $\lambda = 1e-5$).

The total number of epochs rejected was small (average of 10.00% of epochs rejected, range 3.99%–25.75%) and the proportion of remaining trials did not differ significantly between conditions (for all comparisons: $\chi^2_{(1)} < 2.29$, $p < 0.78$, Bonferroni-corrected for multiple comparisons). The average remaining number of trials was 355.14 for the first triplet position, 355.40 for the second, 357.14 for the third, and 339.57 for random.

Statistical analyses

Classification of learners. Participants were divided into two groups, “learners” and “other,” using a conservative criterion. Learners were defined as individuals who scored on the behavioral offline test 22 or more of 32. According to the binomial distribution, this is the minimal score needed to present significantly above-chance learning at the individual level (with $\alpha = 0.05$) (see, e.g., Bogaerts et al., 2016; Siegelman et al., 2017b). All participants that did not meet this criterion are considered “other.”

Spectral decomposition. All spectral estimates were based on time-frequency representations of power derived via a sliding-window FFT. Spectral estimates were computed for frequencies between 1 and 30 Hz on 0.5 s. Hann tapered windows covering the interval between -0.8 and 0.5 s (17 windows with a step size of 0.05 s). Each window was zero-padded to 1 s, providing a spectrally interpolated frequency resolution of 1 Hz. For our analysis of the prestimulus window, we computed a single prestimulus power spectral density estimate for each participant, channel, frequency, and epoch by averaging the estimates for the first 7 windows, covering the -0.8 to 0 s interval. This prestimulus time window is largely uncontaminated by the visual evoked response for the preceding shape, which onsets 0.6 s before this window of interest. Averaging in this way over each FFT window results in a more stable spectral estimate via Welch’s method with a consecutive window overlap of 90% (Welch, 1967). The focus on oscillatory activity in the window leading up to stimuli (i.e., no stimulation) was also our motivation to analyze the frequency range up to 30 Hz. Indeed, higher frequency signatures have typically been associated with the response to stimulus presentation (e.g., Muthukumaraswamy and Singh, 2013; Landau et al., 2015).

Grand-average power spectrum. To calculate the grand-average power spectrum (shown in Fig. 2), we averaged the spectral estimates for all trials, including trials from Positions 1–3 as well as random trials. We then averaged over subjects and electrodes. Spectral peaks were identified as maxima in the 1–30 Hz grand-averaged power spectrum. Band-

limited estimates were derived based on these peaks as the average of the maximum ± 1 Hz, which we refer to as frequencies of interest (FOI). A simple normalization by frequency was applied dividing each power bin by $1/f$. The identified spectral peak frequencies for each of the trial types were virtually identical to those of the reported grand-average power spectrum. In addition, they fell within a significant frequency cluster, identified by cluster-based permutation, based on an F -statistic computed from 1 to 30 Hz across the three positions.

Spectral content of different prestimulus intervals. To investigate differences in spectral power of the prestimulus windows before the 3 triplet position stimuli, we computed average power per stimulus position for each channel and FOI for each of the 9 structured exposure periods. As a control, we also calculated average power for yoked 1, 2, and 3 positions in the random stream (e.g., each third stimulus is treated as a Position 3), for which no differences were predicted. To mitigate individual differences in overall raw EEG power, a normalization was performed separately for each stream type (random vs structured). The power for each position (triplet positions for the structured stream; yoked positions for the random stream) was expressed as the modulation of average power (i.e., percentage) over all three positions and channels.

To contrast the spectral power of the prestimulus windows before the 3 triplet position stimuli, taking into account the expected progression of learning over the course of the experiment, we used a cluster-based approach based on Monte-Carlo estimates (Maris and Oostenveld, 2007). The 9 exposure periods of the structured stream were treated as the time dimension in this analysis. For each FOI, a nonparametric permutation test clustered data samples of adjacent electrodes and time points (i.e., periods within the structured familiarization) simultaneously and compared the sum of the descriptive statistic used, across each cluster. This approach is similar to the standard cluster-based analysis over electrode-time pairs within a trial, except here time is considered not at the trial level but at the larger scale of the 9 exposure periods of the structured familiarization stream (Maris and Oostenveld, 2007). Here and in subsequent analyses, cluster-based tests were dependent samples and parameterized with a minimum number of neighboring channels of three and the cluster threshold F -value (or t -value) corresponding to a p -value of 0.25. This cluster threshold parameter does not affect the false alarm rate of the cluster test (for which we maintained the standard value of $\alpha = 0.05$, with p -values Bonferroni-corrected for the number of FOIs tested). It merely sets the threshold for considering a sample as a candidate member of a cluster: smaller values favor highly localized clusters, and larger cluster threshold p -values favor clusters with large spatiotemporal extent (Maris and Oostenveld, 2007). We do, however, demonstrate robustness of our main cluster results across a range of values for this parameter (Steegen et al., 2016). All permutation tests were based on distributions formed from 100,000 permutations. We hypothesized spectral power differences as a result of assimilating the patterned structure of the stream; hence, for our initial analysis, we only included the subset of individuals who demonstrated learning in the test phase ($n = 25$). We further show that, for these same individuals, no position differences are observed in the random stream.

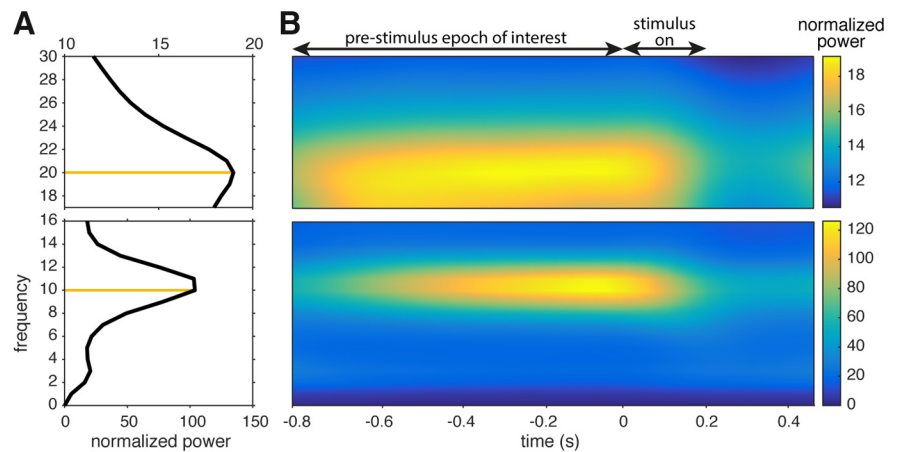


Figure 2. *A*, Peaks in the grand-average raw power spectrum of all prestimulus epochs within the structured and random stream. Yellow lines indicate the peak frequency. *B*, Time-frequency plots for the grand-average for the entire trial epoch. *A*, *B*, Bottom, Lower frequencies; Top, Higher frequencies (with separate axes).

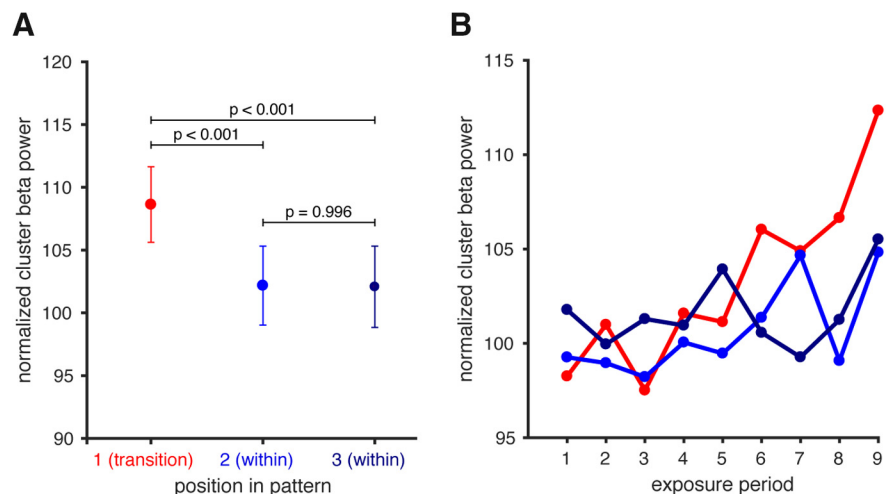


Figure 3. *A*, Average normalized beta power for the significant cluster (based on learners). Error bars indicate the between-subject SE. p -values are corrected for multiple comparisons. *B*, Temporal evolution of the cluster beta power for each of the prestimulus intervals, across the structured familiarization stream. Exposure Period 1 indicates the start of the structured familiarization stream. Period 9 indicates the end of the structured familiarization stream.

Following the cluster-based F -tests, *post hoc* permutation-based paired t -tests were performed between each stimulus position, for each FOI that showed a significant difference for the cluster-based test. For these tests, the average power over subjects was computed for the set of all electrodes that were members of the significant cluster over the temporal extent of the cluster (in practice, this meant averaging across 35 electrodes and across the last 5 exposure periods). No more than one significant cluster was identified for any of the cluster-based tests. Each of the three t -tests was controlled for multiple comparisons using a maximum-based approach, where an omnibus null distribution was constructed for the three t -tests based on the maximum value over t -tests for each permutation (Nichols and Holmes, 2002; Maris and Oostenveld, 2007). For the beta band, we determined via the *post hoc* t -tests that there was no difference between the two predictable Positions 2 and 3, whereas both of these predictable positions significantly differed from unpredictable Position 1 (Fig. 3). This motivated us to pool Positions 2 and 3, and to perform a cluster-based t -test between this average and Position 1. Desiring to generalize results to all subjects with their varying levels of learning, and investigate the brain-behavior correlation, this test was performed on the

“learner” group but also on all subjects. The cluster obtained for all subjects served our neural index for subsequent analyses that included the full sample of participants.

Brain-behavior correlation. The neural index of learning in the frequency domain (Position 1 minus average of Positions 2 and 3) was calculated per individual. This index was based on the average value over each electrode that was present in the significant cluster, across the temporal extent of the cluster (in practice, this meant averaging across 46 electrodes and across the last 6 exposure periods). We tested whether this index is higher for the “learners” compared with the “other” group using an independent-samples *t*-test. Given that the distribution of the post-test data deviates significantly from normality (Shapiro-Wilk statistic = 0.793, $p < 0.001$), we assessed the relation between our neural index of learning and offline test performance using a Spearman rank correlation coefficient. Finally, we tested whether neural indices quantified separately for each of the 9 exposure periods (averaging across the same 46 cluster electrodes) show a monotonic increase in their correlation with test performance, which would be expected if the neural index reflects a learning trajectory.

Stimulus-evoked activity. To examine poststimulus ERPs, the pre-processed data were bandpass filtered from 0.15 to 30 Hz, segmented into epochs ranging from stimulus onset to 0.5 s, and baseline corrected to a 200 ms prestimulus interval. Mean ERPs in the 3 triplet positions were calculated for each participant. To contrast the ERPs for the 3 triplet positions, we took a cluster-based approach, clustering over channels and time points within the epoch using a nonparametric dependent-samples *t* statistic, with a minimum number of neighboring channels of 3. This approach does not restrict the analysis to a particular time window within the epoch or a particular electrode location. A similar approach was taken to investigate spectral differences in the poststimulus intervals of predictable Positions 2 and 3 versus Position 1: we performed a standard cluster-based analysis on the FFT time-frequency results, limited to the epoch from stimulus onset to 0.5 s. For both the ERP and the time-frequency analyses, the reported results are for learners, but qualitatively identical results were obtained looking at the full sample.

Data sharing and code accessibility

The behavioral data reported in this paper, experimental materials, and analysis code will be made available via the Open Science Framework (<https://osf.io/>). EEG data will be provided to any scientist on request. This study has not been preregistered.

Results

Behavioral results

Average group performance on the offline test was 26.46 of 32 (SD = 6.41), which is significantly above chance (one-sample *t*-test comparing mean performance to a score of 16, corresponding to 50% chance level, $t_{(34)} = 9.65$, $p < 0.0001$). Learning scores did not differ significantly between participants who were first exposed to the structured stream ($M = 27.21$, $SD = 5.64$) versus first exposed to the random stream ($M = 25.56$, $SD = 7.31$) ($t_{(33)} = 0.752$, $p = 0.457$). Based on a conservative individual criterion (see Classification of learners), 25 of 35 participants were classified as learners.

Spectral results

Two analyses support our choice of FOIs. A cluster-based *F*-test, with frequency and time as clustering dimensions (averaging over electrodes), was used to identify the frequency range of differences in the spectral content of the prestimulus interval preceding the first, second, and third stimuli of a triplet. We observed a single significant cluster ($p = 0.01$) spanning an 8–25 Hz frequency range (over exposure Periods 4–9). Within this range, we identified two maxima in the average power spectrum across all participants and trials: at 10 and 20 Hz (Fig. 2).

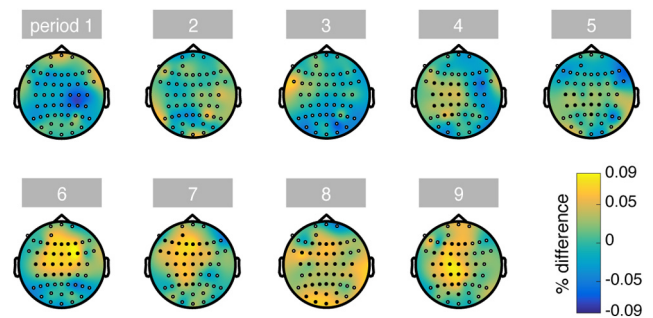


Figure 4. Temporal evolution of the topography of the difference between prestimulus beta-band power for within versus between triplet transitions (based on the full sample). Period 1 indicates the start of the structured familiarization stream. Period 9 indicates the end of the structured familiarization stream. Electrodes that are part of the significant cluster are filled black.

Subsequent analyses focused on these peak frequencies (FOIs) ± 1 Hz. We refer to the 9–11 Hz range as the alpha range and the 19–21 Hz band as the beta range.

Prestimulus beta power is dependent on stimulus position

Focusing on the data of learners, we performed cluster-based *F*-tests for alpha and beta FOIs, comparing spectral content averaged over the prestimulus interval of the different triplet positions, and the FOI ± 1 Hz range (electrodes and exposure period as clustering dimensions). We detected one significant positive cluster with a broad central scalp distribution across exposure Periods 5–9 for the beta range indicating a difference between the three positions (Bonferroni-corrected $p = 0.01$), but none in the alpha range (Bonferroni-corrected $p = 0.11$). A multiverse analysis (Steenen et al., 2016) evaluating the robustness of this beta result across cluster threshold parameter values revealed a significant cluster with all threshold values between 0.05 and 0.46.

A control analysis for the beta finding contrasted the yoked positions in the random stream and revealed no significant clusters ($p = 0.385$).

Post hoc permutation-based paired *t*-tests for learner's beta power in the significant beta-range cluster revealed significantly higher power for the interval preceding a triplet transition (i.e., the interval before shapes in Position 1), which by design was unpredictable (Fig. 3). No difference was found between the spectral power of the prestimulus interval preceding the two predictable Positions 2 and 3, which were virtually identical. Figure 3 also shows the evolution of average cluster beta power over the 9 exposure periods. What we observe is a gradual divergence of the beta power at pattern transitions relative to the beta power within patterns. An ANOVA with position and exposure period as repeated-measures factors indicates a significant effect of position ($F_{(2,48)} = 4.20$, $p = 0.021$) as well as a significant interaction between position and exposure period ($F_{(16,384)} = 2.10$, $p = 0.008$).

Beta power within versus between pattern transitions

Given the difference between the prestimulus beta-frequency power of the unpredictable first shape and the predictable shapes of Positions 2 and 3, we subsequently focused on the contrast of Position 1 versus the average of Positions 2 and 3. For the “learner” group, this contrast revealed, in line with the results reported above, a significant cluster ($p = 0.0038$; significant with all cluster threshold values ≥ 0.02). We performed the cluster-based *t*-test also including all participants (see Spectral content

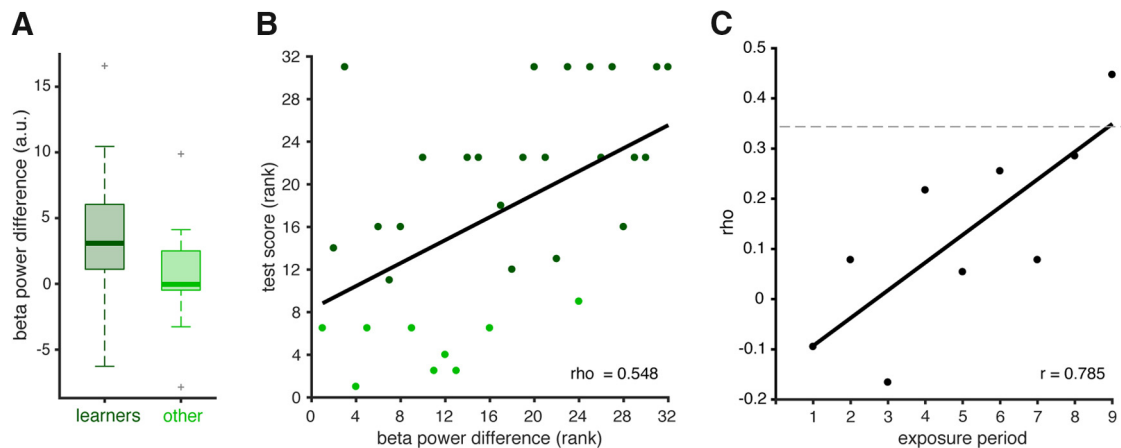


Figure 5. *A*, Boxplot summarizing the beta power modulation (average difference across exposure Periods 4–9) for participants classified as learners versus other. On each box, the central line indicates the median, and bottom and top edges of the box indicate the 25th and 75th percentiles. Dashed whiskers extend to the most extreme data points not considering outliers. +, Outliers. *B*, Relation between the size of the beta power difference and behavioral test scores. *C*, The relation between the beta power difference (calculated per exposure period) and test scores increases over time. Dashed line indicates the critical rho value.

of different prestimulus intervals). Figure 4 illustrates the significant cluster ($p = 0.0032$; significant with all cluster threshold values ≥ 0.01) we observed for the full sample, revealing a broad central scalp distribution across exposure Periods 4–9. This full sample cluster is subsequently used for quantifying the neural beta index of learning for all individual participants.

Relation between neural beta indices of learning and offline test scores

In order to validate our beta finding as a neural signature of learning, it is important to establish a link to behavior. We thus investigated whether the difference in beta power for within- versus between-pattern transitions (averaged across exposure Periods 4–9 and all cluster channels; see Brain-behavior correlation) covaried with the offline measure of learning across individuals (Fig. 5). An independent-samples *t*-test corroborated the prediction of higher beta indices for “learners” ($n = 25$) compared with the “other” group ($n = 10$), although the effect was marginal ($M_{\text{learners}} = 3.50$ $SD_{\text{learners}} = 4.88$; $M_{\text{others}} = 0.59$; $SD_{\text{others}} = 4.62$; $t_{(33)} = 1.616$; $p = 0.058$), likely because of the uneven group sizes and the smaller “other” sample. Indeed, when grouping the participants into high ($n = 17$) and low ($n = 18$) scorers based on a median-split of the behavioral learning measure, that same test showed a highly significant difference ($M_{\text{high}} = 4.86$ $SD_{\text{high}} = 4.92$; $M_{\text{low}} = 0.60$ $SD_{\text{low}} = 4.04$; $t_{(33)} = 2.81$, $p = 0.004$). It is worth noting that the beta effect was not significantly modulated by the order of presentation of the two streams ($t_{(33)} = 0.99$, $p = 0.328$). Considering learning score as a continuous variable, we observe a strong positive relation between these two indices of learning (Spearman’s $\rho = 0.55$, $p = 0.0007$). Finally, when calculating separately for each of the 9 exposure periods, the beta difference (within vs between triplets), and a Spearman’s rho correlation coefficient quantifying the strength of the relation between this difference and the offline test performance, we observe a strong linear increase of ρ values over time ($r = 0.785$, $p = 0.012$).

Stimulus-evoked results

ERP results

No significant clusters were observed when comparing the grand average ERPs for either Position 1 versus the average of Positions 2 and 3, Position 1 versus Position 2, or Position 1 versus Position 3. When calculating the ERP based on exposure Periods 4–9 of the experiment only, we again find no significant clusters for either contrast (smallest corrected p -value = 0.19).

Time-frequency analysis

A cluster-based time-frequency analysis comparing the poststimulus window for Position 1 versus the average of 2 and 3, 1 versus 2, and 1 versus 3 did not reveal any significant clusters. The same result was obtained for data from exposure Periods 4–9 only (smallest corrected p -value = 0.78).

Overall, this suggests that our frequency finding in the prestimulus interval occurs in the absence of any significant modulation of EEG activity in the poststimulus window because of the learning of statistical structure.

Discussion

In the current study, we identified a neural signature of visual SL, operationalized as the extraction of triplet patterns embedded in a continuous sequence of abstract shapes given differences in transitional probabilities. This signature comprised increased beta-band activity in the interval leading up to unpredictable shapes in the stream; thus, this effect was concurrent with triplet boundaries. Importantly, our results indicate that the differential beta-band power before unpredictable shapes increased steadily over the course of repeated exposure to the patterns. Moreover, looking at individual differences in the magnitude of this effect, we show that the differential prestimulus beta power is highly predictive of performance in an offline behavioral test of pattern recognition, which is the learning measure used by the vast majority of SL studies (see Siegelman et al., 2017a). This leads us to conclude that our proposed spectral signature is tracking the learning process of segmenting the continuous stream, and thus provides a valid online assessment of SL performance.

The identification of a neural signature of visual SL offers novel insights regarding the mechanisms underlying SL. The increased beta power at triplet transitions point to two theoretical possibilities. A first possibility is that such beta-band activity reflects anticipation of uncertainty (i.e., higher entropy) once the statistical structure of the recurrent patterns in the stream is assimilated. Moreover, in addition to uncertainty, the upcoming item, consisting of the first position of a new triplet, is highly informative of the identity of this triplet. Thus, the maximal preparation of processing resources may be expected before the first item of a triplet. This interpretation concurs with the proposed functional role of beta oscillations for attentional top-down

regulation in the literature (for review, see Engel and Fries, 2010; Bastos et al., 2012). Whereas this would be, to our knowledge, the first report on increased beta power at high entropy moments in a sequence, it has previously been shown that hippocampal BOLD activity is sensitive to the entropy of a visual stimulus stream (Strange et al., 2005; Harrison et al., 2006). In line with our first suggested interpretation of the beta finding, this hippocampal activity represents “the expected information or novelty of an event before it occurs” (Strange et al., 2005).

On a speculative note, given the direction of our effect (i.e., more beta power before pattern transitions relative to transitions within a visual pattern) and its wide scalp distribution, it might reflect a distributed, global beta network state in anticipation of high uncertainty and a focused, local state in anticipation of predictable events. Consistent with the latter state, locally increased beta activity before expected sensory events has been suggested to reflect “the mobilization of neuronal populations under predictive signals” (Arnal and Giraud, 2012; see also Bernasconi et al., 2011). Linking back to neurophysiological findings in the primate literature (Bressler and Richter, 2015), the assumption is that intracranial recordings, which are typically limited in spatial coverage, pick up on such local beta network states, whereas EEG, with its wide scalp coverage, might be more sensitive to beta synchronization within distributed states.

A second possibility is that the increased beta oscillations reflect postprocessing of the now completed triplet. If one regards a learned triplet as a cognitive set that is the target of learning, it is possible that the interval between triplet transitions is used to actively maintain this set in memory. Indeed, such functional role has been hypothesized for beta-band activity (Engel and Fries, 2010). This is further evident in findings reporting elevated beta-band activity in the delay phase of working memory tasks (Deiber et al., 2007; Siegel et al., 2009; Salazar et al., 2012). Similarly, if learning results in the grouping of stimuli into chunks and these chunks are what is stored in memory (Orbán et al., 2008; Perruchet, 2019), the oscillatory beta signal could reflect the process of memory encoding or strengthening (e.g., Berke et al., 2008). However, whereas participants in our SL task were cued about the presence of structure, they were not informed about the embedded triplet patterns or instructed to memorize them (different from typical working memory tasks). An interpretation in terms of postprocessing is also in line with studies in the event boundary literature showing that offline encoding processes occur immediately following an event offset and are predictive of subsequent memory performance for the event (e.g., Ben-Yakov and Dudai, 2011; Sols et al., 2017).

Our finding that we can track learning during passive exposure to regularities without monitoring overt responses has important implications for assessing learning. The caveats involved in assessing learning through postfamiliarization test have been discussed in length (see, e.g., Siegelman et al., 2017a). Our findings offer then an online neurobiological signature of learning that potentially reveals a learning trajectory over time. The emergence of the beta effect with increased pattern repetitions and the strong linear increase of the relation between the beta-power difference (between vs within triplets) and final test performance suggests it might indeed reflect the learning trajectory. However, given that learning was measured behaviorally only at a single time point, this interpretation needs to be further established using continuous behavioral testing during pattern exposure.

How generalizable the oscillatory beta signature of SL is to different learning situations is an important question for future research. Whereas previous findings linked theta-band activity to

the prediction of visual categories and locations during search (e.g., Cashdollar et al., 2017; Spaak and de Lange, unpublished observations), we did not observe such an effect for the predictable shape identities. This suggests that oscillatory signatures of regularity learning might be contingent on the object of learning and/or the nature of experimental tasks.

The manifestation of learning in interstimulus beta-power differences was observed in the absence of any difference in stimulus processing between positions as assessed by ERPs and time-frequency analysis. Previous literature suggests that the learning of structure modulates stimulus processing as captured by ERPs, a finding not reproduced in our study despite the relatively large sample size. One possibility is that our use of equiluminant shapes obscured differences in stimulus processing. However, Abia and Okanoya (2009), using a visual SL task most similar to ours yet with maximally contrasting shapes, reported significant *N*400 differences only for learners (*n* = 9) and only for the first of three sessions of triplet exposure. Jointly with the absence of ERP differences in our dataset, this raises the question of whether such ERP measures in tasks with passive visual exposure are reliable and, if they are, what aspect of the learning process they tap given that the effect seems not sustained across later periods of exposure. Based on the recent finding that perceptual expectations modulate activity in the alpha range (Zhou et al., 2019), one could make the prediction of more alpha power for first position shapes, which are unpredictable. We found no evidence for this prediction, suggesting that the alpha modulation observed by Zhou et al. (2019) might be specific to the contrast between expected and unexpected items, with the latter items being violations of a predictable transition given a learned pattern.

In conclusion, our findings reveal a neural signature of visual regularity learning: elevated beta-band activity at pattern transitions. This signature tracks the segmentation process during pattern exposure and is highly predictive for the behavioral learning outcome. Whether the heightened beta-band activity reflects the anticipation of a novel upcoming pattern or rather postprocessing of the completed pattern requires additional investigation, aiming to unravel the possible functional role(s) of beta-band oscillations in regularity learning.

References

- Abia D, Katahira K, Okanoya K (2008) On-line assessment of statistical learning by event-related potentials. *J Cogn Neurosci* 20:952–964.
- Abia D, Okanoya K (2009) Visual statistical learning of shape sequences: an ERP study. *Neurosci Res* 64:185–190.
- Arnal LH, Giraud AL (2012) Cortical oscillations and sensory predictions. *Trends Cogn Sci* 16:390–398.
- Aslin RN, Saffran JR, Newport EL (1998) Computation of conditional probability statistics by 8-month-old infants. *Psychol Sci* 9:321–324.
- Bastos AM, Usrey MW, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical Microcircuits for Predictive Coding. *Neuron* 76:695–711.
- Batterink LJ, Paller KA (2017) Online neural monitoring of statistical learning. *Cortex* 90:31–45.
- Batterink LJ, Reber PJ, Neville HJ, Paller KA (2015) Implicit and explicit contributions to statistical learning. *J Mem Lang* 83:62–78.
- Ben-Yakov A, Dudai Y (2011) Constructing realistic engrams: poststimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. *J Neurosci* 31:9032–9042.
- Berke JD, Hetrick V, Breck J, Greene RW (2008) Transient 23–30 Hz oscillations in mouse hippocampus during exploration of novel environments. *Hippocampus* 18:519–529.
- Bernasconi F, Manuel AL, Murray MM, Spierer L (2011) Prestimulus beta oscillations within left posterior sylvian regions impact auditory temporal order judgment accuracy. *Int J Psychophysiol* 79:244–248.

- Bogaerts L, Siegelman N, Frost R (2016) Splitting the variance of statistical learning performance: a parametric investigation of exposure duration and transitional probabilities. *Psychon Bull Rev* 23:1250–1256.
- Bressler SL, Richter CG (2015) Interareal oscillatory synchronization in top-down neocortical processing. *Curr Opin Neurobiol* 31:62–66.
- Buzsáki G (2016) *Rhythms of the Brain*. Oxford: Oxford UP.
- Cashdollar N, Ruhnau P, Weisz N, Hasson U (2017) The role of working memory in the probabilistic inference of future sensory events. *Cereb Cortex* 27:2955–2969.
- Cunillera T, Toro JM, Sebastián-Gallés N, Rodríguez-Fornells A (2006) The effects of stress and statistical cues on continuous speech segmentation: an event-related brain potential study. *Brain Res* 1123:168–178.
- Daikoku T (2018) Neurophysiological markers of statistical learning in music and language: hierarchy, entropy, and uncertainty. *Brain Sci* 8:114.
- Daltrozzo J, Emerson SN, Deocampo J, Singh S, Freggens M, Branum-Martin L, Conway CM (2017) Visual statistical learning is related to natural language ability in adults: an ERP study. *Brain Lang* 166:40–51.
- De Diego Balaguer R, Toro JM, Rodríguez-Fornells A, Bachoud-Lévi AC (2007) Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS One* 2:e1175.
- Deiber MP, Missonnier P, Bertrand O, Gold G, Fazio-Costa L, Ibañez V, Giannakopoulos P (2007) Distinction between perceptual and attentional processing in working memory tasks: a study of phase-locked and induced oscillatory brain dynamics. *J Cogn Neurosci* 19:158–172.
- Engel AK, Fries P (2010) Beta-band oscillations: signalling the status quo? *Curr Opin Neurobiol* 20:156–165.
- Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci* 2:704–716.
- Farthouat J, Franco A, Mary A, Delpouve J, Wens V, Op de Beeck M, De Tiège X, Peigneux P (2017) Auditory magnetoencephalographic frequency-tagged responses mirror the ongoing segmentation processes underlying statistical learning. *Brain Topogr* 30:220–232.
- Fiser J, Aslin RN (2001) Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol Sci* 12:499–504.
- Fiser J, Aslin RN (2002) Statistical learning of new visual feature combinations by infants. *Proc Natl Acad Sci USA* 99:15822–15826.
- Frost R, Siegelman N, Narkiss A, Afek L (2013) What predicts successful literacy acquisition in a second language? *Psychol Sci* 24:1243–1252.
- Frost R, Armstrong BC, Siegelman N, Christiansen MH (2015) Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn Sci* 19:117–125.
- Frost R, Armstrong BC, Christiansen MH (2019) Statistical learning research: a critical review and possible new directions. *Psychol Bull* 145:1128–1153.
- Glicksohn A, Cohen A (2013) The role of cross-modal associations in statistical learning. *Psychon Bull Rev* 20:1161–1169.
- Harrison LM, Duggins A, Friston KJ (2006) Encoding uncertainty in the hippocampus. *Neural Netw* 19:535–546.
- Hipp JF, Engel AK, Siegel M (2011) Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron* 69:387–396.
- Jost E, Conway CM, Purdy JD, Walk AM, Hendricks MA (2015) Exploring the neurodevelopment of visual statistical learning using event-related brain potentials. *Brain Res* 1597:95–107.
- Jung TP, Makeig S, Humphries C, Lee TW, Mckeown MJ, Iragui V, Sejnowski TJ (2000) Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37:163–178.
- Jungé JA, Scholl BJ, Chun MM (2007) How is spatial context learning integrated over signal versus noise? A primacy effect in contextual cueing. *Vis cogn* 15:1–11.
- Karuza EA, Newport EL, Aslin RN, Starling SJ, Tivarus ME, Bavelier D (2013) The neural correlates of statistical learning in a word segmentation task: an fMRI study. *Brain Lang* 127:46–54.
- Kirkham NZ, Slemmer JA, Johnson SP (2002) Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* 83:B35–B42.
- Landau AN, Schreyer HM, Van Pelt S, Fries P (2015) Distributed attention is implemented through theta-rhythmic gamma modulation. *Curr Biol* 25:2332–2337.
- Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164:177–190.
- McNealy K, Mazziotta JC, Dapretto M (2006) Cracking the language code: neural mechanisms underlying speech parsing. *J Neurosci* 26:7629–7639.
- Muthukumaraswamy SD, Singh KD (2013) Visual gamma oscillations: the effects of stimulus type, visual field coverage and stimulus motion on MEG and EEG recordings. *Neuroimage* 69:223–230.
- Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp* 15:1–25.
- Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* 2011:156869.
- Orbán G, Fiser J, Aslin RN, Lengyel M (2008) Bayesian learning of visual chunks by human observers. *Proc Natl Acad Sci USA* 105:2745–2750.
- Perruchet P (2019) What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. *Top Cogn Sci* 11:520–535.
- Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. *Science* 274:1926–1928.
- Salazar RF, Dotson NM, Bressler SL, Gray CM (2012) Content-specific fronto-parietal synchronization during visual working memory. *Science* 338:1097–1100.
- Sanders LD, Newport EL, Neville HJ (2002) Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nat Neurosci* 5:700–703.
- Schapiro AC, Kustner LV, Turk-Browne NB (2012) Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr Biol* 22:1622–1627.
- Shohamy D, Turk-Browne NB (2013) Mechanisms for widespread hippocampal involvement in cognition. *J Exp Psychol Gen* 142:1159–1170.
- Siegel M, Warden MR, Miller EK (2009) Phase-dependent neuronal coding of objects in short-term memory. *Proc Natl Acad Sci USA* 106:21341–21346.
- Siegelman N, Bogaerts L, Christiansen MH, Frost R (2017a) Towards a theory of individual differences in statistical learning. *Philos Trans R Soc Lond B Biol Sci* 372:20160059.
- Siegelman N, Bogaerts L, Frost R (2017b) Measuring individual differences in statistical learning: current pitfalls and possible solutions. *Behav Res Methods* 49:418–415.
- Siegelman N, Bogaerts L, Kronenfeld O, Frost R (2018) Redefining “learning” in statistical learning: what does an online measure reveal about the assimilation of visual regularities? *Cogn Sci* 42:692–727.
- Sols I, DuBrow S, Davachi L, Fuentesmilla L (2017) Event boundaries trigger rapid memory reinstatement of the prior events to promote their representation in long-term memory. *Current Biology* 27:3499–3504.
- Steege S, Tuerlinckx F, Gelman A, Vanpaemel W (2016) Increasing transparency through a multiverse analysis. *Perspect Psychol Sci* 11:702–712.
- Strange BA, Duggins A, Penny W, Dolan RJ, Friston KJ (2005) Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Netw* 18:225–230.
- Tollman EC (1932) *Purposive behavior in animals and men*. New York: Appleton.
- Turk-Browne NB, Jungé J, Scholl BJ (2005) The automaticity of visual statistical learning. *J Exp Psychol Gen* 134:552–564.
- Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK (2009) Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J Cogn Neurosci* 21:1934–1945.
- Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM (2010) Implicit perceptual anticipation triggered by statistical learning. *J Neurosci* 30:11177–11187.
- Welch PD (1967) The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans Audio Electroacoust* 15:70–73.
- Zhou YJ, Pérez-Bellido A, Haegens S, de Lange FP (2019) Perceptual expectations modulate low-frequency activity: a statistical learning magnetoencephalography study. *J Cogn Neurosci* 32:691–702.