

Effect of pattern awareness on the behavioral and neurophysiological correlates of visual statistical learning

Sonia Singh^{1,*}, Jerome Daltrozzo¹ and Christopher M. Conway^{1,2}

¹Department of Psychology, Georgia State University, Urban Life Building 11th Floor, 140 Decatur Street, Atlanta, GA 30303-3083, USA; ²Neuroscience Institute, Georgia State University, 880 Petit Science Center, 100 Piedmont Ave. SE, Atlanta, GA 30303, USA

*Correspondence address. Department of Psychology, Georgia State University, PO Box 5010, Atlanta, GA 30302, USA. Tel: 404-413-6217; Fax: 404-413-6218; E-mail: ssingh11@student.gsu.edu

Abstract

Statistical learning is the ability to extract predictive patterns from structured input. A common assumption is that statistical learning is a type of implicit learning that does not result in explicit awareness of learned patterns. However, there is also some evidence that statistical learning may involve explicit processing to some extent. The purpose of this study was to examine the effect of pattern awareness on behavioral and neurophysiological correlates of visual statistical learning. Participants completed a visual learning task while behavioral responses and event-related potentials were recorded. Following the completion of the task, awareness of statistical patterns was assessed through a questionnaire scored by three independent raters. Behavioral findings indicated learning only for participants exhibiting high pattern awareness levels. Neurophysiological data indicated that only the high-pattern awareness group showed expected P300 event-related potential learning effects, although there was also some indication that the low awareness groups showed a sustained mid-to late-latency negativity. Linear mixed-model analyses confirmed that only the high awareness group showed neurophysiological indications of learning. Finally, source estimation results revealed left hemispheric activation was associated with statistical learning extending from frontal to occipital and parietal regions. Further analyses suggested that left insula, left parahippocampal, and right precentral regions showed different levels of activation based on pattern awareness. To conclude, we found that pattern awareness, a dimension associated with explicit processing, strongly influences the behavioral and neurophysiological correlates of visual statistical learning.

Key words: statistical learning; implicit learning; explicit processing; awareness; P300; event-related potential

Introduction

Statistical learning is the ability to extract statistical associations or predictive patterns from structured input. Statistical learning can be used to infer sequential probabilities among ordered elements in the environment (Saffran *et al.* 1996). For example, in natural language, linguistic units (e.g. phonemes,

syllables, and words) are arranged in a non-random sequence according to the specific language's phonology, phonotactics, semantics, and syntax. Statistical learning can occur even after relatively brief exposure times (Saffran *et al.* 1996; Aslin *et al.* 1998; Fiser and Aslin 2002; Kirkham *et al.* 2002), allowing the extraction of statistical structures to anticipate and predict future events (Conway *et al.* 2010).

Received: 29 December 2016; **Revised:** 11 August 2017. **Accepted:** 14 November 2017

© The Author 2017. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Article Highlights

- The effect of pattern awareness on the behavioral and neural correlates of visual statistical learning was explored.
- Analyses of the behavioral and neurophysiological data showed that the level of awareness of the individuals for the underlying statistical patterns—assessed through questionnaire following the completion of the task—was closely associated with behavioral and neurophysiological indications of learning.
- Specifically, the findings demonstrated that only the participants with a high level of pattern awareness showed clear evidence of statistical learning as measured by response times and event-related potentials.
- Source estimation results further indicated that statistical learning was associated with left hemispheric activation in a network spanning occipital, parietal, and frontal regions, with increased activation observed for participants demonstrating high pattern awareness relative to low pattern awareness in a subset of brain regions including left parahippocampal cortex.
- Overall, these findings show that pattern awareness and learning ability are closely linked, suggesting that statistical learning in this task is mediated largely by explicit processes.

Traditionally, statistical learning is believed to be a type of implicit learning, occurring in the absence of explicit pattern awareness of the underlying structure that needs to be learned (Reber 1989; Perruchet and Pacton 2006; Reber 2013). Implicit learning can be defined as learning without the intention to learn or without conscious awareness of the knowledge that has been acquired (Cohen and Squire 1980; Reber and Squire 1994; Travers et al. 2010; Jeste et al. 2015). However, there is also evidence that statistical learning may involve explicit processing to some extent (Turk-Browne et al. 2005; Wessel et al. 2012; Daltrozzo and Conway 2014). In particular, a number of studies are in line with the assumption that statistical learning involves both implicit and explicit mechanisms (Batterink et al. 2015; for a recent review, see Daltrozzo and Conway 2014). Although some findings imply that attention to stimuli is not required for statistical learning (Saffran et al. 1997), others indicate that attention improves both visual (Turk-Browne et al. 2005) and auditory learning (Toro et al. 2005; Emberson et al. 2013). More specifically, Turk-Browne et al. (2005) concluded that visual statistical learning is both automatic and intentional, meaning, although attention is a prerequisite for relevant stimulus selection, subsequent learning occurs without intent or awareness. In addition, Hendricks et al. (2013) utilized a dual-task paradigm involving a working memory task in conjunction with an artificial grammar learning task to dissociate (visual) automatic and intentional learning. They found that although some aspects of visual statistical learning are relatively automatic, making direct grammaticality judgments at test as well as transferring knowledge to perceptually dissimilar stimulus sets both appeared to depend on explicit processing resources.

The extent to which statistical learning results in awareness of the learned information is also controversial. On the one hand, several researchers have suggested that the relationship between statistical learning performance and awareness may not be so clear, because implicit learning can occur independently of explicit awareness (Curran and Keele 1993; Goschke, 1998; Song et al. 2007). Others, such as Cleeremans (2006), propose a more nuanced view, where statistical learning affects awareness but only under specific circumstances. Specifically, mental representations obtained from exposure to a sequence might only result in awareness when the strength of activation of these representations reaches some critical level. Consequently, statistical learning without awareness may ensue whenever these representations are poorly activated. Findings from another recent study, suggested that measures of learning that primarily target the explicit knowledge of

sequences (e.g. recognition judgment and familiarity ratings) were not as sensitive as other indirect indices such as response times that do not rely solely on awareness (Batterink et al. 2015). According to another view, the relationship between statistical learning and awareness may be bidirectional. In particular, for participants who become aware of the existence of structured sequences, their level of intention to learn the sequence structures might modulate learning (Rüsseler et al. 2003).

There are other reasons to believe that awareness of the to-be-learned patterns can affect performance. For example, Cleeremans (1993) proposed an information processing model of statistical learning by building on the simple recurrent network (Cleeremans and McClelland 1991; Cleeremans 1993). The guiding hypothesis behind the model is that awareness of sequence structure alters the nature of the task in that instead of anticipating subsequent events in temporal context there is a switch to upcoming event retrieval from short-term memory. Here, performance is contingent on attentional resources, and such dependence could result in degradation of output, especially when memory representations are less reliable (during dual task performance). Their main findings were that explicit knowledge may enhance implicit learning and also that participants will attempt to utilize explicit knowledge whenever accessible. Findings from a different study by McIntosh et al. (1999) involved participants who were either aware or unaware of a tone that predicted a visual event or not. Participants who were aware of this versus those who were unaware showed different responses both behaviorally and neuroanatomically (as measured by regional cerebral blood flow). Several of the interacting brain areas (left prefrontal cortex, contralateral prefrontal cortex, sensory cortex, and cerebellum) showed changes in functional connectivity that also correlated with the awareness of participants.

The present study

The purpose of the present study was to explore the behavioral and neurophysiological effects of pattern awareness on statistical learning. To achieve this aim, we measured event-related potentials (ERPs) in response to a visual statistical learning task following a paradigm similar to that of Jost et al. (2015). The task involved the presentation of a series of visual stimuli wherein “target” stimuli could be predicted with varying levels by the preceding “predictor” stimulus. ERPs to two different types of predictor—cueing the target with either high predictability (HP) or low predictability (LP)—were recorded. Jost et al. (2015) found

that a greater P300 component was observed for the HP relative to the LP stimuli following learning. In the present study, this ERP effect was explored as a dependent variable against the effect of the independent variable pattern awareness, as assessed through a questionnaire after completion of the statistical learning task. We hypothesized that participants who show more awareness of the underlying statistical patterns would also show the largest learning effects in terms of behavioral response times and ERPs. We also incorporated source estimation analyses to explore the activation of brain areas during the statistical learning task and to determine how activation was affected by the level of pattern awareness of participants.

Methods

Participants

A total of 34 participants (27 females, aged 18–49 years; $M = 22.4$ years, $SD = 6.3$) without any language, neurological, or psychological deficits from Georgia State University participated in the study for class credits. All participants were right handed according to the Edinburgh Handedness Inventory (Oldfield 1971), except seven (3 left handed and 4 ambidextrous). All participants were native English speakers. None of them spoke, wrote, read, or understood Chinese (some of the stimuli were

Chinese characters, see Visual Statistical Learning Task section below). Participants were recruited from the local University online recruiting system and provided written informed consent to participate. The study was approved by the local ethics committee (The Institutional Review Board of Georgia State University).

Visual statistical learning task

Participants were administered a visual statistical learning task similar to that used in Jost et al. (2015) and Daltrozzo et al. (2017) while ERPs were recorded (see the Electroencephalography Acquisition section below). To discourage verbal rehearsal or naming of the stimuli (and to increase the difficulty of the learning task), the task used non-verbal and unfamiliar characters from standard Chinese script (common to Mandarin and Cantonese), and participants were only allowed to participate if they were unfamiliar with this script. As a follow-up, after the task, participants were asked whether they used verbal labels during the task.

In the statistical learning task, a set of traditional Chinese characters were presented to participants interspersed with target face stimuli (Fig. 1). The target faces could be either “happy” or “unhappy.” Target assignment was at the beginning of the

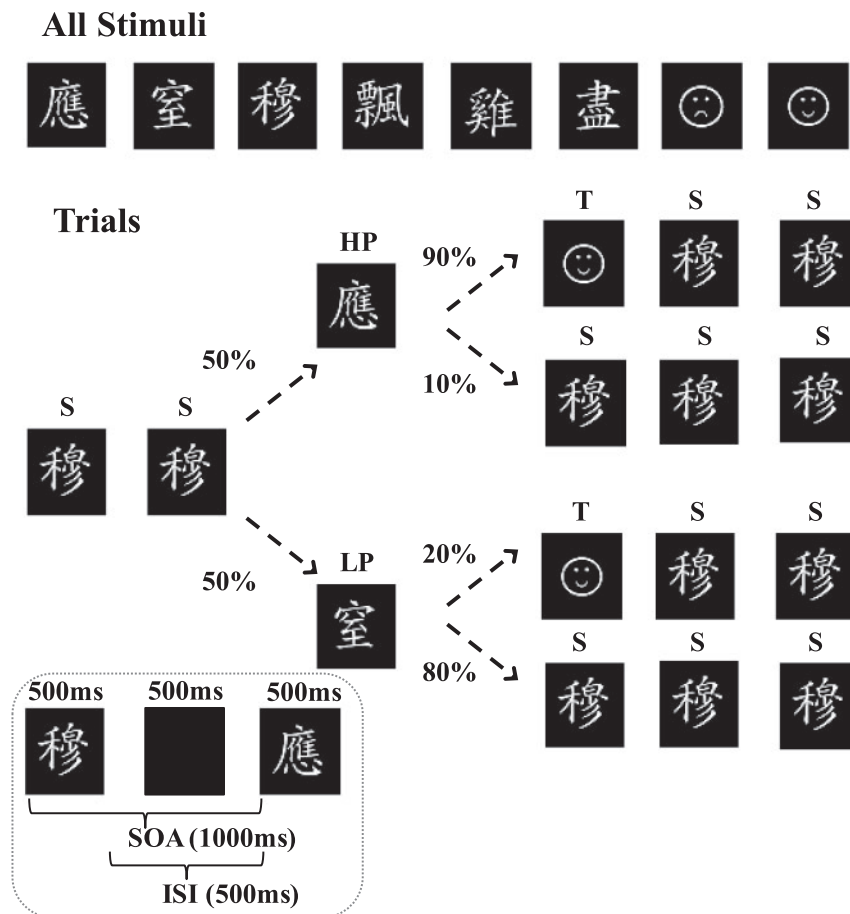


Figure 1. Statistical rules of the sequential learning task. The target (T) followed either a high predictability predictor (HP) on 90% of trials or a low predictability predictor (LP) on 20% of trials. These two items were presented within a stream of standard (S) items. For each participant, HP, LP, and S were randomly assigned to one of the six Chinese characters displayed on the top panel of the figure. For each participant, T was pseudo-randomly assigned to one of the two smileys, see top panel. Bottom panel shows stimulus onset asynchrony (SOA) and inter-stimulus intervals (ISI).

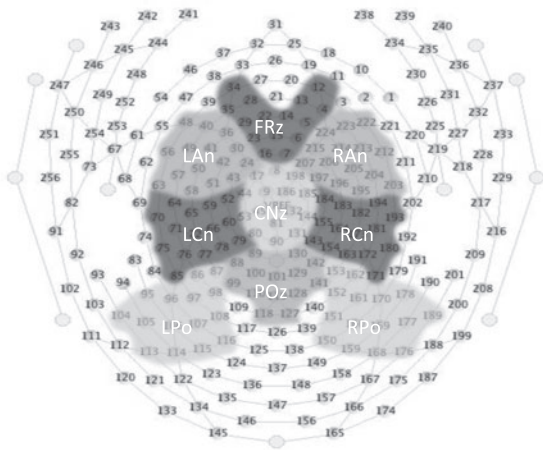


Figure 2. Electrical Geodesic Inc. sensor net with the nine ROIs highlighted [left (LAN), middle (FRz), and right anterior (RAN); left (LCn), middle (CNz), and right central (RCn); and left (LPo), middle (POz), and right posterior (RPo) regions used for recording cortical activity].

experiment and once chosen was applied across the whole experiment for that participant. Fifty percent of participants saw a happy smiley face as the target, and the remaining saw the unhappy smiley as the target (i.e. none of the participants ever saw both happy and unhappy targets during the experiment). Both happy and unhappy faces were used to balance any spontaneous emotion elicited by the stimuli (Halberstadt et al. 2009) that might affect emotion perception and in turn affect learning across participants. The task of the participants was to press a button as fast as possible when they saw the target. Unbeknownst to the participants, the target followed either a high-probability predictor (HP) on 90% of trials or a low-probability predictor (LP) on 20% of trials (Fig. 1). The predictors and the target were presented within a stream of standard (S) items. For each participant, HP, LP, and S were pseudo-randomly assigned to 1 of the 6 Chinese characters displayed on the top panel of Fig. 1. As in Jost et al. (2015), the participant was expected to learn the statistical relationship between the predictors and the target. Participants were not given information about the existence of underlying probabilities that define the co-occurrence of the predictors and the target to encourage incidental learning.

Each predictability condition (HP and LP) was presented 50 times. All trials were continuous and pseudo-randomly ordered across the two predictability conditions, so that participants encountered a seamless presentation between trials. Each participant was presented with a total of 100 trials, divided into 5 blocks of 20 trials each. A break of 30 s was given between each block. Stimuli were presented electronically using E-Prime 2.0.8.90 software (Psychology Software Tools, Pittsburgh, PA, USA) on a Dell Optiplex 755 computer. All visual stimuli were presented in white in the center of the computer screen on a dark background. Stimuli were displayed for 500 ms, followed by a dark screen, which was displayed for an additional 500 ms (inter-stimulus interval was 500 ms). Thus, the visual stimuli were presented with a 1000-ms stimulus onset asynchrony.

Electroencephalography acquisition

During the statistical learning task, electroencephalography (EEG) data were taken from 256 scalp sites using an Electrical Geodesic Inc. (EGI) sensor net (Fig. 2) and was preprocessed

using Net Station Version 4.3.1 with subsequent processing using custom scripts written in Matlab (version R2012b 8.0.0783, The MathWorks) using the EEGLAB toolbox (version 10.2.2.2.4a; Delorme and Makeig, 2004). Electrode impedances were kept below 50 k Ω . The EEG was acquired with a 0.1- to 100-Hz band-pass filter at 250 Hz with vertex reference and then rereferenced to the average reference of all sensors and low-pass filtered at 30 Hz. Signals containing non-stereotypical artifacts, including high-amplitude, high-frequency muscle noise and electrode cable movements, were rejected (~25% of trials). Prior to segmentation, stereotypical artifacts, such as vertical eye blinks and horizontal eye movements, were corrected with an extended Infomax independent component analysis using EEGLAB (Lee et al. 1999). The continuous EEG was then segmented into epochs -200 ms to $+1000$ ms with respect to the predictor onset. ERPs were baseline corrected with the 200 ms prestimulus data. Individual ERPs were computed for each participant, predictor type, and electrode. All experimental sessions were conducted in a 132-square feet double-walled, soundproof acoustic chamber.

Pattern awareness questionnaire

After the statistical learning task, the EEG electrode net was removed, and the participants completed a questionnaire to assess their level of pattern awareness (see Table 1). Pattern awareness levels were obtained from an inter-rater agreement among three independent scorers of the participants' responses. Each rater was requested to provide a score of 1 ("low awareness" to the statistical rules/patterns embedded in stimuli sequences of the statistical learning task) or 2 ("high awareness" of the pattern). The inter-rater reliability was 96.5% (Cronbach's alpha). For each participant, the final pattern awareness score was the mean of the scores of the three raters. The participants were then separated into two groups with $n = 17$ in each group based on a median split of these mean scores of pattern awareness: the group of high pattern awareness and the group of low pattern awareness.

The two groups did not differ significantly in terms of age (high-pattern awareness: $M = 22.47$, $SD = 7.15$; low-pattern awareness: $M = 22.29$, $SD = 5.86$; $U = 154.5$, $P = .74$, Mann-Whitney, two tailed]. In addition, before the statistical learning task, the executive control capacity of the participants was assessed with the Flanker task (Eriksen and Eriksen 1974), which is commonly used to test the executive control (Fan et al. 2005). There was no significant difference between Flanker performance in the high pattern ($n = 16$; $M = 2.187$, $SD = 87.817$) and the low pattern awareness groups [$n = 16$; $M = -58.000$, $SD = 89.026$; after removing 2 outliers and using the centered Flanker difference scores; $incongruent$ minus $congruent$ trials; $F(30, 1) = 0.500$, $P = 0.832$; $\eta^2 = 0.938$].

Analyses of the ERPs

We applied a linear mixed model (LMM) (West et al. 2014) to our pattern awareness and statistical learning data at the single-trial level. (The LMM was performed with R (version 3.1.2) using the lmer() function of the lme4 library, Bates et al. 2009.) The LMM is increasingly used to analyze EEG data comprising of such large data sets (Bagiella et al. 2000; Davidson and Indefrey 2007; Moratti et al. 2007; Pritchett et al. 2010; Wierda et al. 2010; Newman et al. 2012). The LMM offers advantages over traditional repeated-measures analyses of variance (ANOVAs), including richer modeling of random effects with, for instance,

Table 1. Pattern awareness questionnaire

Pattern Awareness Questionnaire	
1.	Think about the task with Chinese characters you did. Did you notice anything about the Chinese characters? Tell me about your perception of the task. [Verbatim record]
2.	Do you think the Chinese characters were occurring randomly? [If the participant says no, ask to explain how the characters were non-randomly displayed.]
3.	Was there a pattern or anything regular in the order that the Chinese characters were presented?
4.	Was there a Chinese character that usually came before the target (the smiley face you were looking for)?
5.	If you noticed a pattern, at what point did you notice it? Before the 1st break, after the 1st break, after the 2nd break, after the 3rd break, or after the 4th break?

Table 2. Table of LMM specifications (df, AIC, BIC LL, deviance) for the neurophysiological (ERP) model (within 200–700 ms) and behavioral (RT) model across all trials

Model Specifications	df	AIC	BIC	LL	Deviance	chisq	Df	P
Behavioral RT								
RT ~ Predictability + PA + Predictability:PA + (1 participant)	6	7548.4	7581.5	-3768.2	7536.4	0	0	1
Neurophysiological data (200–700 ms)								
EEG	82	-1972.002	-1971.122	986.083	-1972.166	14.257	43	<0.001***
Mean ~ Predictability + PA + Predictability:PA + Predictability:ROI + Predictability:PA:ROI + (1 + ROI participant)								

multiple, crossed, and/or nested random effects (Newman et al. 2012). As a result, this model allows increased accuracy and external validity of the parameter estimate. Another important advantage of using the LMM instead of an ANOVA approach is the ability of the LMM to handle missing data and non-sphericity issues, both of which the LMM can adequately address. Thus, unlike the ANOVA model, there is no need for subsequent correction for non-sphericity (e.g. Greenhouse–Geisser or Huynh–Feldt; Bagciella et al. 2000; Baayen et al. 2008).

To analyze the effect of the cortical topography, nine regions of interests (ROIs; see Fig. 2) were defined: left (LAN), middle (FRz), and right anterior (RAN); left (LCn), middle (CNz), and right central (RCn); and left (LPo), middle (POz), and right posterior (RPO) regions. The applied LMM was similar to the model used by Newman et al. (2012), with fixed effects defined below from predictability (HP or LP), pattern awareness (PA, i.e. high or low awareness of the statistical patterns), and ROI as well as with intercept by pattern awareness and intercept and ROI by participant random factors. According to the R syntax, the LMM was:

$$\begin{aligned} \text{EEG Mean} &\sim \text{Predictability} + \text{PA} + \text{Predictability} \\ &: \text{PA} + \text{Predictability} : \text{ROI} + \text{Predictability} : \text{PA} \\ &: \text{ROI} + (1 + \text{ROI} | \text{participant}) \end{aligned}$$

Similar to Newman et al. (2012), the LMM was applied on a single-trial data. Single-trial data were the mean values of EEG over 200–700 ms time window based on the topography of the statistical learning ERP effect in Jost et al. (2015). [To correct for the incompatibility between the additive nature of the LMM (and ANOVA models) with the multiplicative nature of interactions that could yield incorrect significant (i.e. Type I error) interactions involving ROI, McCarthy and Wood (1985) developed a correction by EEG mean scaling, see also Dien and Santuzzi 2005. In every condition, for each participant, mean EEG amplitudes are scaled by the square root of the sum of the squared mean EEG amplitudes, i.e. $X_{ij}/\sqrt{\Sigma(X_{ij}^2)}$, where X_{ij} is the EEG mean amplitude for participant i in condition j . If the scalp ROI by

condition interaction remains significant after rescaling, this allows for more confidence in the authenticity of the interaction, under certain conditions, Urbach and Kutas 2002.]

Similar to the LMM mentioned earlier, response times (RTs) were also analyzed with the LMM procedure. The model was based on the model used to analyze ERPs. Because it pertains to behavioral RT data only, the model neither contains the EEG mean variable nor the ROI factor and is adjusted appropriately. [Bonferroni-corrected pairwise comparisons were applied with the `mcposthoc.fnc()` of the `LMErConvenienceFunctions` (version 2.10) library (Tremblay and Ransijn 2015). RT data were normalized using a square root transformation.]

According to the R syntax, the LMM was:

$$\text{RT} \sim \text{Predictability} + \text{PA} + \text{Predictability} : \text{PA} + (1 | \text{participant})$$

Table 2 shows the model specifications for both RT and ERP models.

EEG source estimation

Source estimates were performed to further investigate the underlying neural mechanisms of statistical learning between the levels of pattern awareness. [For source estimation, all procedures were processed with BrainStorm software package (Tadel et al. 2011). Cortical generators of cue-locked ERP activity were reconstructed by modeling conductive head volume according to OpenMEEG BEM (Kybic et al. 2005; Gramfort et al. 2010) that is executed in the Brainstorm software package (Tadel et al. 2011). The solution space was constrained to the cerebral cortex, and cortical current source density mapping was obtained using a distributed model consisting of 15 000 fixed dipoles normally oriented to the cortical surface. Additionally, the inverse transformation was applied to Brainstorm’s default template Montreal Neurological Institute (MNI) brain (colin27 atlas) (Collins et al. 1998; Tzourio-Mazoyer et al. 2002) i.e. a canonical mesh of the cortex to approximate real anatomy (see Tadel et al.

2011 for a review). This head model was then fit to the standard geometry of the current 256 sensor net. All subsequent source analysis, and statistical estimation of the Z-scores relative to the baseline (before cue onset) was then processed. Cortical current maps were computed from the EEG time series using a linear inverse estimator called weighted minimum-norm current estimate (WMNE). WMNEs are a measure of the current density flowing at the surface of the cortex.] Source estimation was first applied at a preliminary level on the entire data set (i.e. all 34 participants of the study) to visually and statistically infer the difference between HP and LP conditions. This difference was computed as the difference between HP_i and LP_i .

To further obtain separate source estimate maps for the between-group variable, pattern awareness (high and low), each having two within conditions (HP and LP), we first computed the difference (HP – LP) in each group of high- and low pattern awareness. Thus, for each participant i ($i = 1, \dots, n = 17$), we computed: (i) A single average HP_i ; (ii) a single average for LP_i ; and (iii) the difference $D_i = HP_i - LP_i$. Then, at the group level, we computed the following: (i) $m_1 = |\text{mean}(D_i)|_{\text{high pattern awareness}}$: the absolute value of the mean of D_i (with $i = 1, \dots, n = 17$, i.e. over all participants of the high awareness group); (ii) $m_2 = |\text{mean}(D_j)|_{\text{low-pattern awareness}}$: the absolute value of the mean of D_j (with $j = 1, \dots, N = 17$, i.e. over all participants of the low awareness group); (iii) $D = m_1 - m_2$; the difference between these means. As a final result, we obtained a signed (\pm) source estimation difference D , indicating for which group the difference was more important. A non-parametric test termed permutation t-test (Maris and Oostenveld 2007) as implemented by Brainstorm software was also used to compare high and low pattern awareness to obtain a statistically significant difference

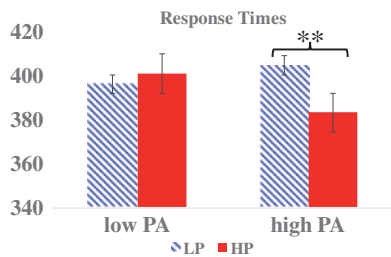


Figure 3. Response times in the high (HP, solid bar) and low predictability (LP, textured bar) conditions for the high and low pattern awareness (PA) groups. The vertical axis is mean response time in milliseconds; Horizontal axis depicts the *post hoc* tests to the Predictability \times PA interaction; ** $P < 0.001$.

using scouts (see Results section) per the Mindboggle brain atlas via Brainstorm (Klein et al. 2005).

Results

Behavioral results

Analysis of the RT data showed that predictability (i.e. HP vs. LP) interacted with pattern awareness [$F(1, 1805.03) = 9.195$, $P < 0.01$]. Figure 3 shows the average RTs to targets following the HP and LP stimuli for the high- and low pattern awareness groups. *Post hoc* tests indicated a significant effect of predictability (i.e. learning had occurred) in the high pattern awareness group [$P = 0.001$] but a non-significant effect of predictability in the low-pattern awareness group [$P = 0.627$]. That is, for the high pattern awareness group, participants responded quicker to the targets when they were preceded by the HP stimulus compared with the LP stimulus, suggesting that they had learned the predictor–target probabilities. No such behavioral facilitation was observed in the low pattern awareness group. Table 3 shows the fixed and random effects from the RT analyses across groups.

ERP results

The grand averaged ERPs across participants with high pattern awareness ($n = 17$) for the HP and LP conditions are shown in Fig. 4. Overall, the right ROI appears to show evidence of a P300-like response, a component that was also observed by Jost et al. (2015) using a similar paradigm. Specifically, there is an increased positivity for the HP stimulus relative to LP roughly

Table 3. Table of fixed and random effects from the RT analyses in the high PA and low PA groups, [$*P < 0.01$, $**P < 0.001$, $***P < 0.0001$; pattern awareness (PA); high predictability (HP); low predictability (LP)]

Model RT			
Fixed effects	Estimate	SE	P
Intercept	19.824	0.352	<0.001***
High predictability	0.075	0.155	0.627
High PA	0.184	0.498	0.713
High predictability:high PA	-0.667	0.220	0.002**
Random effects			
	SD		
Participant	1.33		
Residual	1.818		

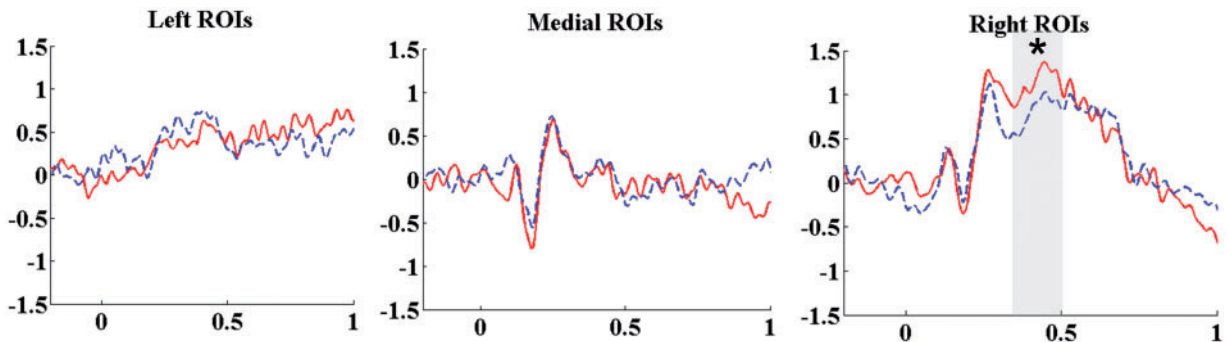


Figure 4. Grand averaged ERPs for high pattern awareness participants within left, medial, and right ROIs ($n = 17$). ERPs to the predictor in the HP (solid line) and LP (dashed line) conditions (vertical axis: electric potential in microvolts, positivity upward; horizontal axis: time in seconds). Shaded gray bar shows time window corresponding to ANOVA; * $P < 0.01$.

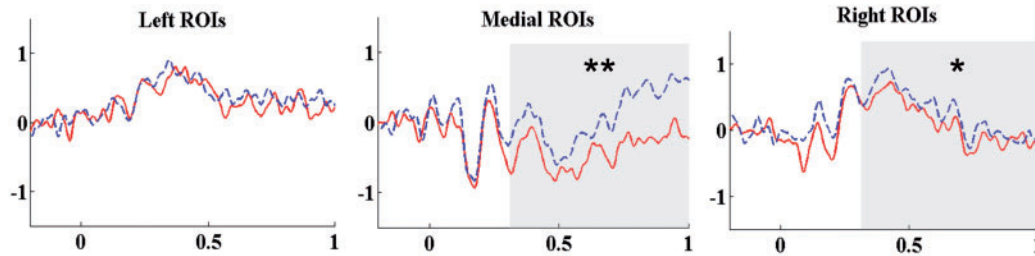


Figure 5. Grand averaged ERPs for low pattern awareness participants within left, medial, and right ROIs ($n = 17$). ERPs to the predictor in the HP (solid line) and LP (dashed line) conditions (vertical axis: electric potential in microvolts, positivity upward; horizontal axis: time in seconds). Shaded gray bar shows time window corresponding to ANOVA; * $P < 0.01$, ** $P < 0.001$.

between 250 ms and 500 ms. There also appears to be an N200 effect in the medial ROI, with a more negative peak for HP compared with LP. Based on visual inspection, one-way ANOVAs were conducted on the EEG means averaged across trials for two time windows: (i) 150–250 ms for the N200 and (ii) 300–500 ms for the P300. Results showed that there was no significant difference between high and low predictability for the N200 time window for any of the ROIs: left [$F(1, 1528) = 0.006$, $P = 0.936$], medial [$F(1, 1528) = 2.304$, $P = 0.129$], and right [$F(1, 1528) = 0.745$, $P = 0.388$]. On the other hand, the results confirmed the existence of the P300, with a significant difference between high and low-predictability conditions in the right ROI [$F(1, 1528) = 4.351$, $P = 0.037$] but not the left [$F(1, 1528) = 1.062$, $P = 0.303$] or medial [$F(1, 1528) = 0.105$, $P = 0.746$] ROI.

The grand averaged ERPs across participants with low pattern awareness ($n = 17$) for the HP and LP conditions are shown in Fig. 5. Overall, the right and, particularly, the medial ROIs appear to show differences in the waveforms between predictor stimuli in mid and late latencies, but unlike the high-pattern awareness participants, the pattern is reversed, i.e. there is higher amplitude for the LP than the HP condition. Akin to the high pattern awareness group, we performed ANOVAs for each ROI using the EEG averages. For the low awareness group, we focused on a larger 300–1000 ms time window due to the observed sustained negativity. The results showed no significant difference between high and low-predictability condition for the left ROI [$F(1, 1528) = 1.745$; $P = 0.187$]. In contrast, there was a significant difference between predictor conditions for medial [$F(1, 1528) = 40.539$; $P < 0.001$] and right [$F(1, 1528) = 4.502$; $P = 0.034$] ROIs, with the high predictability stimuli showing more negative potentials than the low-predictability stimuli (see Fig. 5).

To further explore the possible effects of pattern awareness on the ERP correlates of statistical learning, we applied an LMM on the single-trial ERP data [using the McCarthy and Wood (1985) correction] within the 200–700 ms window. This window includes most of the primary ERP effects observable by visual inspection. The results indicated a two-way interaction: Predictability \times ROI [$F(16, 62) = 8.39$, $P < 0.001$]; a second two-way interaction: Predictability \times Pattern Awareness [$F(1, 340203) = 65.763$, $P < 0.001$]; and a three-way interaction: Predictability \times Pattern Awareness \times ROI [$F(16, 62) = 2.74$, $P < 0.001$]. Because all factors of the two-way interactions (i.e. predictability, pattern awareness, and ROI) belong to the three-way interaction, we focus on the three-way interaction. Table 4 displays the fixed and random effects from the LMM analyses.

A graphical depiction of the model effects from the three-way interaction between the ERP means (McCarthy corrected) for each level of predictability and ROI is provided, for high pattern awareness (Fig. 6, right panel) and low pattern awareness (Fig. 6, left panel) groups. Figure 6 was obtained with the Effects

Table 4. Table of fixed and random effects from the ERP analyses in the high pattern awareness (HPA) and low pattern awareness (LPA) groups, [* $P < .01$, ** $P < .001$, *** $P < .0001$; pattern awareness (PA)]

Full model			
Fixed effects	Estimate	SE	P
(Intercept)	-0.002	0.001	0.003**
HP	0.000	0.000	0.948
HPA	-0.001	0.001	0.556
HP:HPA	0.000	0.000	0.180
LP:CNz	0.001	0.001	0.064
HP:CNz	0.001	0.001	0.334
LP:FRz	-0.002	0.001	0.006**
HP:FRz	-0.002	0.001	0.015*
LP:LCn	0.004	0.001	0.000***
HP:LCn	0.003	0.001	0.000***
LP:LPO	0.007	0.001	0.000***
HP:LPO	0.007	0.001	0.000***
LP:POz	0.007	0.001	0.000***
HP:POz	0.005	0.001	0.001***
LP:RAn	0.000	0.001	0.580
HP:RAn	0.000	0.001	0.915
LP:RCn	0.004	0.001	0.003**
HP:RCn	0.003	0.001	0.009**
LP:RPO	0.007	0.002	0.000***
HP:RPO	0.007	0.002	0.000***
LP:HPA:CNz	0.001	0.001	0.201
HP:HPA:CNz	0.002	0.001	0.039*
LP:HPA:FRz	0.000	0.001	0.596
HP:HPA:FRz	-0.001	0.001	0.456
LP:HPA:LCn	0.000	0.001	0.747
HP:HPA:LCn	0.001	0.001	0.384
LP:HPA:LPO	0.001	0.002	0.666
HP:HPA:LPO	0.001	0.002	0.717
LP:HPA:POz	0.002	0.002	0.266
HP:HPA:POz	0.003	0.002	0.102
LP:HPA:RAn	0.000	0.001	0.612
HP:HPA:RAn	0.000	0.001	0.633
LP:HPA:RCn	0.001	0.002	0.472
HP:HPA:RCn	0.002	0.002	0.274
LP:HPA:RPO	0.002	0.003	0.401
HP:HPA:RPO	0.002	0.003	0.471
Random effects	SD		
Participant	0.003		
CNz	0.003		
FRz	0.002		
LCn	0.003		
LPO	0.005		
POz	0.006		
RAn	0.003		
RCn	0.004		
RPO	0.007		
Residual	0.013		

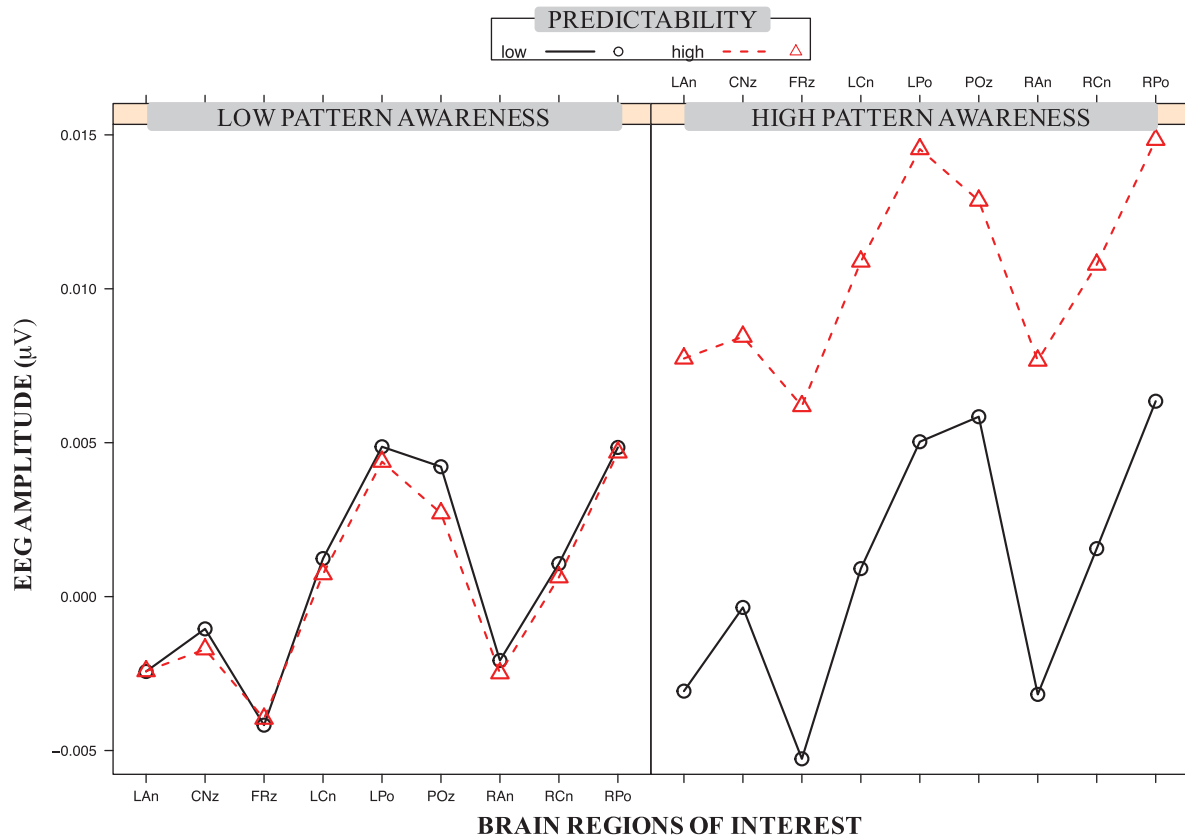


Figure 6. Predictability \times pattern awareness \times ROI interaction for all participants ($n = 34$). Right panel depicts high pattern awareness and the left panel depicts low pattern awareness. Horizontal axis are the 9 ROIs [left (LAn), middle (FRz), and right anterior (RAn); left (LCn), middle (CNz), and right central (RCn); and left (LPo), middle (POz), and right posterior (RPo)]. The vertical axis is the LMM estimated ERP mean amplitude between 200 ms and 700 ms post-predictor onset (positivity upward in microvolts; LP: solid line; HP: dashed line).

displays package in R (Fox, 2003).] Participants with high pattern awareness show an expected ERP effect across all ROIs, i.e. enhanced positivity for HP compared with LP stimuli (similar to the positivity observed in Jost et al. 2015). The low pattern awareness group shows little differentiation of the predictor conditions in any ROIs, and when it does occur, the HP predictor is more negative than LP.

Source estimation results

First, we used source estimation to examine for any differences in brain activation related to learning across all participants (see Methods section). This was performed using scouts according to the Mindboggle brain atlas (Klein et al. 2005) incorporated in Brainstorm and not the full cortical maps. (By restricting activity to the scouts, we discard any spatial resolution, and thus, the statistical results *per se* cannot be represented on a cortical map. However, we can explore the following issues: (i) the brain region where a statistically significant difference of source activity was found (upon correcting for multiple comparisons) and (ii) the direction of the difference, i.e. which condition was associated with a higher or a lower source activity.) A scout represents a region or a subset of dipoles demarcated on the cortical surface or head volume (Tadel et al. 2016). Scout selection was performed within the same time window as the ERP LMM analyses, which was 200–700 ms post-stimulus onset and the ROIs were chosen *a priori*. We checked for differences between HP and LP source activation across all 34 participants over the

entire time window and then over 50 ms bin increments. Second, we examined brain activation for high and low pattern awareness using scouts across the full 200–700 ms time window as well as at 50 ms time window increments. For time periods across and within these time windows, which are not depicted graphically, *t*-values were non-significant.

Figure 7 depicts the results of the source estimation for the difference between HP and LP (i.e. areas of the brain indicative of a learning effect) for all participants. Each depicted subfigure is showing the demarcated ROI generically on the brain template corresponding to the view (left, L and right, R), and below each template is the associated graph. All *t*-values are representative of results from a non-parametric permutation test (Maris and Oostenveld 2007) in ROIs (scouts), where significance at the 0.05 level remained even after correcting for multiple comparisons (for signal, frequency, and time) with the false discovery rate procedure (Benjamini and Hochberg 1995). (The *t*-values are only reported as statistically significant when significance remains after the false discovery rate correction; $\alpha = 0.05$; the positive or negative direction depicted in the *t*-value graph denotes which condition was associated with higher or lower source activity. For predictability, and pattern awareness, the high condition was denoted by positive values and the low by negative values.) In Fig. 7A, negative *t*-values represent the statistically significant difference for HP – LP in superior parietal regions and that activation was greater for the LP condition between 200 ms and 300 ms post-stimulus. Between 450 ms and 500 ms, a significant HP – LP difference (greater for HP) in the left lateral

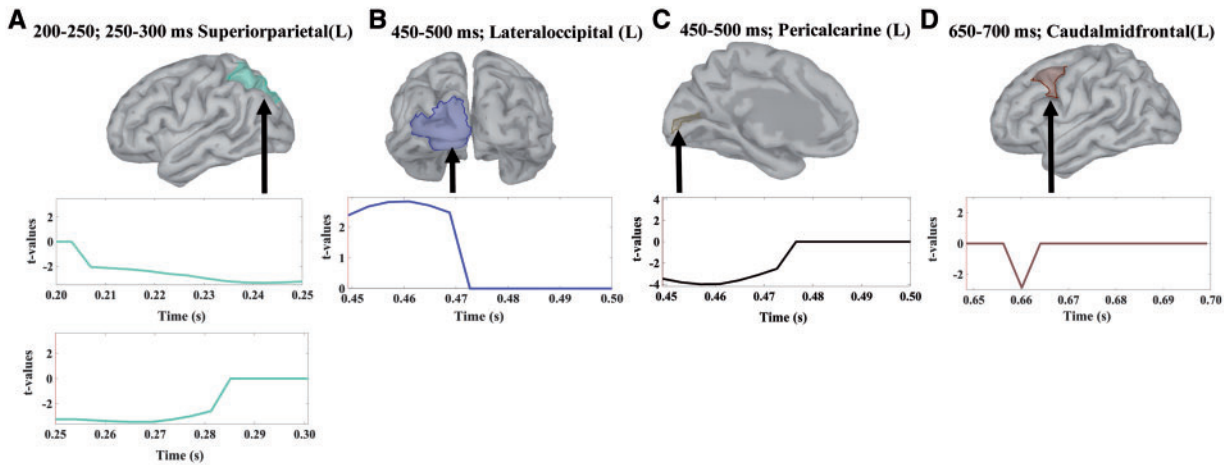


Figure 7. Cortical source estimation maps with t-values for scouts for HP minus LP ($n = 34$). (A) Highlighted scouts show significant difference for HP – LP in superior parietal region (left) at 200–250 ms and 250–300 ms. (B) Highlighted scouts show significant HP – LP difference during a 450–500 ms time window in the lateral occipital (left) region. (C) Highlighted scouts show significant HP – LP difference in a 450–500 ms time window depicted in the pericalcarine (left) area. (D) Highlighted scouts show significant HP – LP difference in a 650–700 ms in the caudal mid-frontal (left) region. (A–D) Accompanied by graph showing averaged t-values (vertical axis) and time in seconds (horizontal axis).

occipital region (Fig. 7B) as well as a significant HP – LP difference (greater for HP) in the left pericalcarine region (Fig. 7C) was observed. Between 650 ms and 700 ms, activation was found in the left caudal mid-frontal region (Fig. 7D). Thus, we see a left lateralized posterior–anterior shift over time associated with the HP–LP learning effect, with superior parietal, pericalcarine, and lateral occipital regions showing activation in the first 200–500 ms and caudal mid-frontal regions showing activation after 650 ms post-stimulus.

We then tested for statistically significant effects of pattern awareness at 50 ms increments within the 200–700 ms time windows. These are brain areas showing significant activation corresponding to learning the predictor stimuli (HP – LP) and that differed between the high and low awareness groups. As shown in (Fig. 8, the insula (L), parahippocampal (L) and precentral (R) regions were found to show significant activation differences between the high and low awareness groups.

Discussion

The aim of this study was to examine the extent to which pattern awareness influences the learning of visual statistical regularities. Our main findings were the following. First, only participants showing high levels of pattern awareness demonstrated robust behavioral learning effects as measured by RTs. Second, only participants with high pattern awareness showed the expected P300 ERP effects as well as clear indications of learning as assessed with the LMM analyses. Finally, source estimation results showed left lateralization and a caudal–rostral gradient accompanying learning across all participants. Differences in brain activation were also observed for the high and low pattern awareness groups in specific brain regions. We discuss each of these primary findings in turn.

Behavioral findings

Behavioral data revealed that pattern awareness appeared to have a strong effect on learning. For those participants with low pattern awareness, there was no difference in the response times between predictor conditions and thus no evidence of learning. When awareness was high, the response times were

much lower for HP compared with LP conditions, indicating learning. Thus, in contrast to earlier conceptualizations of statistical learning being an implicit process (Reber 1989), these findings reveal that only participants demonstrating high awareness showed behavioral indications of learning.

ERP findings

In a previous study using a similar learning paradigm, Jost et al. (2015) observed a P300-like positivity elicited by the high predictability condition. The P300 is regarded as an index of target detection and evaluation (van Zuijlen et al. 2006) and has also been observed in other learning tasks (Baldwin and Kutas 1997; Rüsseler et al. 2003; Carrión and Bly 2007). In the present study, we also obtained a P300 but only for participants with high pattern awareness (Fig. 4, right ROI). For the low pattern awareness group (Fig. 5), instead of the expected P300 effect, there was an extended negativity of the HP relative to LP conditions in the medial ROI. At least one other study from the literature has reported obtaining a similar result with unconsciously processed stimuli eliciting a reversal (i.e. negative) P300 effect (van Gaal et al. 2008), albeit with a different task (the go/no-go task). The results of their study suggested that inhibitory control functions might be influenced by unconscious events. Applied to the current findings, the extended negativity would appear to suggest that learning occurred. On the other hand, the results of the LMM analyses are much more clear in showing ERP effects only in the high pattern awareness group (Fig. 6), suggesting that only the high pattern awareness participants demonstrated learning as revealed by the ERPs.

With regard to the interpretation of the P300 in the high pattern awareness group, Jost et al. (2015) suggested that the occurrence of the P300, which is normally elicited by targets in a standard oddball paradigm, shifted from the target to the stimulus that predicted the target with a high level of probability. Ample exposure to the sequential statistics of the input array may have enabled participants to view the frequent HP stimulus as if it were the target itself, displaying the prototypical P300 response. That the P300 effect was observed only in the high pattern awareness group suggests qualitatively different neural processes occurring during the task for the two groups of

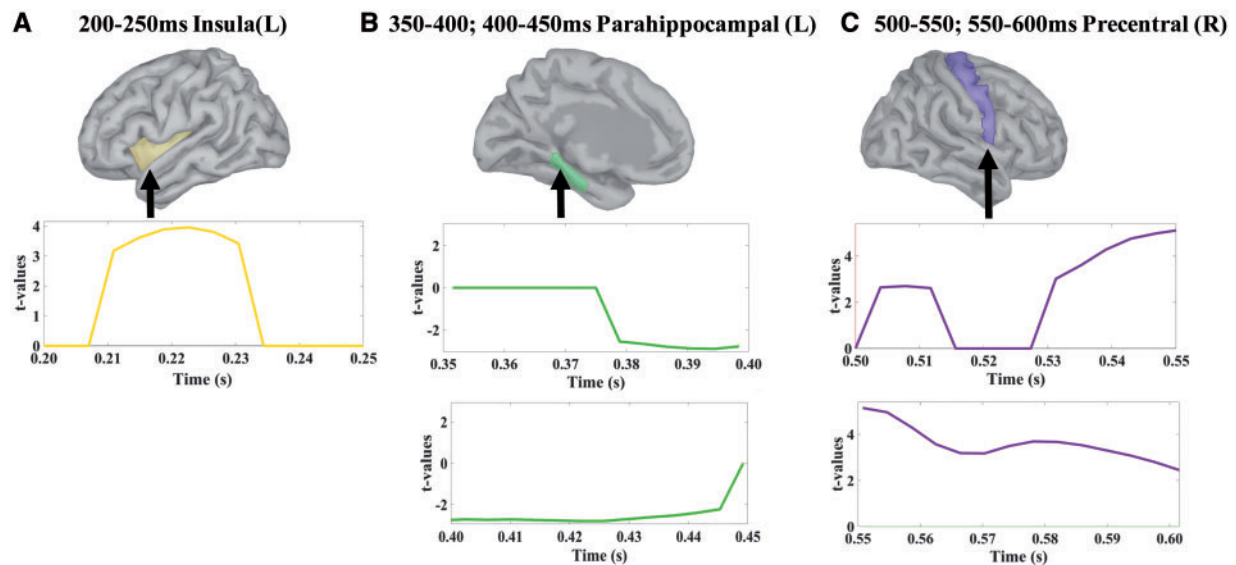


Figure 8. Cortical source estimation maps with t-values for scouts for differences in pattern awareness ($n = 17$). (A) Highlighted scouts show significant differences between high and low pattern awareness between 200 ms and 250 ms for the insula (left) region. (B) Highlighted scouts show significant differences between high and low pattern awareness between 350 ms and 400 ms and between 400 ms and 450 ms for the parahippocampal (left) region. (C) Highlighted scouts show significant differences between high and low pattern awareness between 500 ms and 550 ms and between 550 ms and 600 ms for the precentral (right) region. (A–C) Accompanied by graph showing averaged t-values (vertical axis) and time in seconds (horizontal axis).

participants. In sum, the ERP results show clear indications of learning for the high awareness participants but much less clear evidence for the low awareness participants. Taken together, with the behavioral evidence suggesting learning only for the high awareness participants, these results add to the pre-existing literature underscoring awareness as a prerequisite for or at the very least, influential, to statistical learning ability (Cleeremans 1993; McIntosh et al. 1999; McIntosh et al. 2003).

Source estimation findings

Source analysis for all 34 participants of the study indicated left hemispheric activation across the 200–700 ms time window, specifically over the superior parietal, occipital, and mid-frontal ROIs (Fig. 7). The left superior parietal region is involved in spatial orientation (Corbetta et al. 1995) and receives visual and sensory input and is closely tied to self-awareness (Goldberg et al. 2006). Other visual processing areas were also observed, specifically, left lateral occipital and left pericalcarine cortex. Lateral occipital regions are known to be involved in object processing (Grill-Spector et al. 2001) and possibly visual awareness (Ro et al. 2003). More specifically, perception at early stages of visual encoding can result from exposure to a prior visual stimulus via feedback projections in the visual cortex. Pericalcarine cortex is a part of the occipital lobe and is closely tied to the central visual field. Research shows that this area is implicated in early visual processing, which in turn is also associated with pre-attentive and attentive vision (see Lamme and Roelfsema 2000). In general, the involvement of visual processing areas during the visual statistical learning task is consistent with previous empirical findings and theory suggesting an important role of modality-specific perceptual processing during statistical learning (Conway and Christiansen 2005; Turk-Browne et al. 2009; Frost et al. 2015).

The caudal mid-frontal region also showed activation and corresponds roughly to Brodmann area 46, a part of the frontal

cortex associated with sustained attention and working memory (Curtis and D'Esposito 2003; Rypma et al. 1999). Frontal activations are consistently observed across different kinds of statistical learning and sequential learning tasks (e.g., Fletcher et al. 1999; Skosnik et al. 2002). While investigating the asymmetry, connectivity, and segmentation of the arcuate fascicle Fernández-Miranda et al. (2015) concluded that the caudal middle frontal gyrus along with other cortical areas (pars opercularis, pars triangularis, and ventral precentral gyrus) are part of a frontal trajectory that is integral to language processing. A possible overlap in neural areas supporting statistical learning and the “language network” is consistent with previous research suggesting strong links between statistical learning and language processing (Conway et al. 2010; Misyak et al. 2010; Arciuli and Simpson 2012; Christiansen et al. 2012; Tabullo et al. 2013; Daltrozzo et al. 2017). Overall, these source estimation findings point to a general left hemispheric activation pattern associated with statistical learning, across a network of areas involved in perceptual processing, working memory, sustained attention, and language, along a caudal to rostral temporal gradient (i.e. earlier activation in posterior brain regions and later activation in frontal regions). In addition, the left hemispheric pattern of activation is reminiscent of the left laterality observed in language acquisition and other aspects of learning (Friederici and Alter 2004).

Further source estimation analyses comparing high and low pattern awareness groups revealed different levels of activation for the left insula, right precentral cortex, and left parahippocampal regions. The anterior insula has been shown to be activated during performance monitoring and is also modulated by error awareness. Such activity is thought to be associated with automatic consciously perceived errors, and the encountered errors elicit responses akin to an orienting response (Ullsperger et al. 2010). Additionally, early activation of the insula (which is what we report) has also been associated with risk prediction

error and that its time course is consistent with a role in rapid updating (Preuschoff *et al.* 2008).

The precentral region is a part of the motor cortex and has been associated with sequence learning. Specifically, learning-related changes have been observed in the right precentral region, in addition to other areas (Bischoff-Grethe *et al.* 2004). The left precentral gyrus is involved in speech articulation (Itabashi *et al.* 2016) and language processing (Fernández-Miranda *et al.* 2015). In addition, some research shows that learning new words in a language is associated with increased functional connectivity of regions for learners (compared with non-learners) between the left supplementary motor area and the left precentral gyrus among other regions implicated in phonological rehearsal (Veroude *et al.* 2010).

The left parahippocampal brain region also showed differences between the low compared with high pattern awareness groups. There is evidence to show that parahippocampal activation is associated with item-based processing (Davachi and Wagner 2002) in humans. Electrophysiological findings (in rats and monkeys) have indicated that the neuronal responses in parahippocampal regions represent information about previously occurring items (Brown *et al.* 1987; Li *et al.* 1993). Preston and Gabrieli (2008) found that activations in hippocampus and parahippocampal cortex were associated with explicit memory, dissociating between subsequently remembered and forgotten repeated contexts but were unrelated to context-dependent learning. Importantly, the parahippocampal cortex has recently been implicated as playing an important role during statistical learning (Schapiro *et al.* 2012).

Limitations and future directions

One limitation of this study is that our measure of pattern awareness was taken at the end of the statistical learning task, rather than during, and as such cannot provide fine-grained information about awareness as it might have unfolded in time. Also, regarding the pattern awareness questionnaire, although we tried to quantify subjective participant self-report, there are other ways to measure awareness that are likely less subjective and reliant on participants' own reports. One such method is the process dissociation procedure (Jacoby 1991), which uses a combination of direct and indirect assessments to tease apart the contribution of explicit and implicit memory (i.e. conscious from unconscious learning). Future research could usefully use such a procedure during statistical learning and investigate how differences in conscious awareness contribute to neural patterns of activation. Another potential limitation is that in our statistical analyses we ignore the fact that the statistical significance obtained with ROIs as a predictor variable is also influenced by the spatial proximity between ROIs and as such could be modeled in terms of spatial distance, when entered as an interaction term. From a technical point of view, adding such complex interaction terms in the model involves additional parameter estimation. Provided that such technical issues can be adequately addressed, future statistical analyses using mixed models with ERP data could benefit from a more detailed analysis of each ROIs independent as well as interdependent effects on the other variables.

Another limitation to consider is to what extent the current findings will generalize to other statistical learning and implicit learning tasks. The task we used here differs from other learning tasks in important ways. First, even though we used Chinese character stimuli in non-Chinese speakers to increase the difficulty of the learning task (see Methods section), the

statistical contingencies are relatively simple and easy to learn, which might make this task easier to become aware of the patterns compared to more complex learning tasks such as artificial grammar learning tasks that require not only learning of the statistical regularities but also generalization to novel patterns. Another aspect of the current task that makes it somewhat unique (and we believe is a strength) is that the primary behavioral and neurophysiological indicators of learning are online and indirect measures. That is, at no point (except after the task is over when pattern awareness is assessed) are participants explicitly queried as to their knowledge of the patterns. This is not the case in artificial grammar learning tasks (Reber 1989) or in word segmentation/triplet tasks (Saffran *et al.* 1996), where learning is typically assessed through direct explicit measures. This task characteristic could actually make it less likely that the measures of learning themselves are contaminating the learning process. Thus, it is currently an open question to what extent the findings obtained in the current study will generalize to other measures of statistical learning commonly used in the literature.

In addition, future research might usefully attempt to disentangle the influence of attention and working memory in relation to awareness during statistical learning. In the current study, executive control was measured with the Flanker task at the start of the experimental session. A comparison of the Flanker data across the two pattern awareness groups showed that they did not differ, which suggests that executive control may not affect the relationship between awareness and statistical learning. However, one limitation with the use of the Flanker task in the current study is that it was measured before the statistical learning task and therefore does not provide an online measure of executive control during learning. It is likely that executive control and awareness interact in a complex way to affect learning processes. In their review on the neural mechanisms of attention and awareness, Tallon-Baudry (2012) discuss different ways that attention and awareness could be related: (i) the gateway hypothesis, (ii) the reverse dependence hypothesis, and (iii) the cumulative influence hypothesis. The gateway hypothesis is Dehaene *et al.*'s (2006) classical view, where attention facilitates awareness and might even be a prerequisite for awareness to emerge. According to the reverse dependence hypothesis, attentional mechanisms are only activated if a stimulus is detected at the neural level, implying awareness. In the cumulative influence hypothesis, attention and awareness are each implemented by separate mechanisms, but both independently influence the participant's report of the existence of the stimulus. In contrast, Lamme (2003) argues that, although visual attention and awareness are intimately related, the overlap between mechanisms of attention and memory are more likely than that of attention and awareness. According to Lamme (2003), the current state of the neural network characterizes attentional selection, whereas phenomenal experience ensues from the recurrent interaction between groups of neurons. Future research examining the constructs of attention, memory, and awareness in relation to statistical learning is needed.

Finally, due to the nature of this study, it is not possible to determine the nature of cause and effect between awareness and learning. One possibility is that as participants become increasingly aware of the patterns during the course of the task (possibly due to the use of explicit strategies or the deployment of attention or cognitive effort), then learning improves and the P300 effect results. This possibility would be more consistent with the gateway hypothesis discussed above. On the other

hand, it is also possible that differences in learning ability directly affect awareness of the patterns, with better learning resulting in heightened pattern awareness. This perspective seems more consistent with the reverse dependence hypothesis. Additional research is needed to better understand how these variables causally affect one another.

Conclusion

In conclusion, we have provided evidence for the influence of pattern awareness on statistical learning. Both behaviorally and neurophysiologically, our findings suggest that pattern awareness is closely associated with statistical learning ability. Neurophysiologically, we observed more distinct ERP learning effects in participants who demonstrated high pattern awareness. Across all participants, source estimation results revealed left lateral regions (superior parietal, lateral occipital, pericalcarine, and caudal mid-frontal) that were activated temporally in a caudal-to-rostral manner. Furthermore, differences in pattern awareness were associated with greater levels of activation in the left (insula and parahippocampal regions) as well as right (precentral) hemispheric regions. These findings suggest that pattern awareness influences visual statistical learning and points to an increased need to manipulate and/or measure how this construct affects learning across a variety of individuals and tasks.

Acknowledgements

We thank Georgia State University's Language and Literacy Initiative and the NIH (R01DC012037) for their financial support and J. Deocampo, S. Sims, J. Trapani, S. Emerson, and M. Freggens for their help with data acquisition and analysis; L. Branum Martin for advice on statistical analyses; and J. Deocampo for her help in designing the study.

Data Availability Statement: Data are available on request.

References

- Arciuli J, Simpson IC. Statistical learning is related to reading ability in children and adults. *Cogn Sci* 2012;**36**:286–304.
- Aslin RN, Saffran JR, Newport EL. Computation of conditional probability statistics by 8-month-old infants. *Psychol Sci* 1998;**9**:321–4.
- Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang* 2008;**59**:390–412.
- Bagiella E, Sloan RP, Heitjan DF. Mixed-effects models in psychophysiology. *Psychophysiology* 2000;**37**:13–20.
- Baldwin KB, Kutas M. An ERP analysis of implicit structured sequence learning. *Psychophysiology* 1997;**34**:74–86.
- Bates DM, Maechler M, Bolker B. lme4: Linear mixed-effects models using Eigen and Eigen. R package version 0.999375-39. The Comprehensive R Archive Network (CRAN): The Institute of Statistics and Mathematics of the Wirtshausuniversität Wien (WU); 2009.
- Batterink LJ, Reber PJ, Neville HJ et al. Implicit and explicit contributions to statistical learning. *J Mem Lang* 2015;**83**:62–78.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B* 1995;**57**:289–300.
- Bischoff-Grethe A, Goedert KM, Willingham DT et al. Neural substrates of response-based sequence learning using fMRI. *J Cogn Neurosci* 2004;**16**:127–38.
- Brown MW, Wilson FAW, Riches IP. Neuronal evidence that inferomedial temporal cortex is more important than hippocampus in certain processes underlying recognition memory. *Brain Res* 1987;**409**:158–62.
- Carrión RE, Bly BM. Event-related potential markers of expectation violation in an artificial grammar learning task. *Neuroreport* 2007;**18**:191–5.
- Christiansen MH, Conway CM, Onnis L. Similar neural correlates for language and sequential learning: evidence from event-related brain potentials. *Lang Cogn Process* 2012;**27**:231–56.
- Cleeremans A. Attention and awareness in sequence learning. In: *Proceedings of the 15th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum, p.330–335. 1993.
- Cleeremans A. *Conscious and Unconscious Cognition: A Graded, Dynamic Perspective*. In: Jing Q, Rosenzweig M, d'Ydewalle G, Zhang H, Chen HC, and Zhang K (eds.), *Progress in psychological science around the world: Vol. 1. Neural, cognitive, and developmental issues*. Hove, England: Psychology Press, 2006, 401–18.
- Cleeremans A, McClelland JL. Learning the structure of event sequences. *J Exp Psychol Gen* 1991;**120**:235.
- Cohen NJ, Squire LR. Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. *Science* 1980;**210**:207–10.
- Collins DL, Zijdenbos AP, Kollokian V et al. Design and construction of a realistic digital brain phantom. *IEEE Trans Med Imaging* 1998;**17**:463–8.
- Conway CM, Bauernschmidt A, Huang SS et al. Implicit statistical learning in language processing: word predictability is the key. *Cognition* 2010;**114**:356–71.
- Conway CM, Christiansen MH. Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J Exp Psychol Learn Mem Cogn* 2005;**31**:24–39.
- Corbetta M, Shulman GL, Miezin FM et al. Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. *Science* 1995;**270**:802.
- Curran T, Keele SW. Attentional and nonattentional forms of sequence learning. *J Exp Psychol Learn Mem Cogn* 1993;**19**:189.
- Curtis CE, D'Esposito M. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn Sci* 2003;**7**:415–23.
- Daltrozzo J, Emerson SN, Deocampo J et al. Visual statistical learning is related to natural language ability in adults: an ERP study. *Brain Lang* 2017;**166**:40–51.
- Daltrozzo J, Conway C. Neurocognitive mechanisms of statistical-sequential learning: what do event-related potentials tell us? *Front Human Neurosci* 2014;**8**:437.
- Davachi L, Wagner AD. Hippocampal contributions to episodic encoding: insights from relational and item-based learning. *J Neurophysiol* 2002;**88**:982–90.
- Davidson DJ, Indefrey P. An inverse relation between event-related and time-frequency violation responses in sentence processing. *Brain Res* 2007;**1158**:81–92.
- Dehaene S, Changeux JP, Naccache L et al. Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn Sci* 2006;**10**:204–11.
- Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Meth* 2004;**134**:9–21.
- Dien J, Santuzzi AM. Application of repeated measures ANOVA to high-density ERP datasets: a review and tutorial. In: Handy TC (ed.), *Event-Related Potentials. A Methods Handbook*. Cambridge, MA: MIT Press, 2005, 57–82.
- Emberson LL, Liu R, Zevin JD. Is statistical learning constrained by lower level perceptual organization? *Cognition* 2013;**128**:82–102.

- Eriksen BA, Eriksen CW. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept Psychophys* 1974;16:143–9.
- Fan J, McCandliss BD, Fossella J et al. The activation of attentional networks. *Neuroimage* 2005;26:471–9.
- Fernández-Miranda JC, Wang Y, Pathak S et al. Asymmetry, connectivity, and segmentation of the arcuate fascicle in the human brain. *Brain Struct Funct* 2015;220:1665–80.
- Fiser J, Aslin RN. Statistical learning of higher-order temporal structure from visual shape sequences. *J Exp Psychol Learn Mem Cogn* 2002;28:458–67.
- Friederici AD, Alter K. Lateralization of auditory language functions: a dynamic dual pathway model. *Brain Lang* 2004;89:267–76.
- Fletcher P, Büchel C, Josephs O et al. Learning-related neuronal responses in pre-frontal cortex studied with functional neuroimaging. *Cerebral Cortex* 1999;9:168–78.
- Fox J. *Effect Displays for Linear, Generalized Linear, and Other Models*. 2003. <http://www.r-project.org>, <http://socserv.socsci.mcmaster.ca/jfox/> (15 June 2017, date last accessed).
- Frost R, Armstrong BC, Siegelman N et al. Domain generality vs. modality specificity: the paradox of statistical learning. *Trends Cogn Sci* 2015;19:117–25.
- Goldberg IL, Harel M, Malach R. When the brain loses its self: prefrontal inactivation during sensorimotor processing. *Neuron* 2006;50:329–39.
- Goschke T. *Implicit Learning of Perceptual and Motor Sequences: Evidence for Independent Learning Systems*. In: Stadler MA and Frensch PA (eds.), *Handbook of implicit learning*. Thousand Oaks, CA: Sage, 1998, 401–44.
- Grill-Spector K, Kourtzi Z, Kanwisher N. The lateral occipital complex and its role in object recognition. *Vision Res* 2001;41:1409–22.
- Gramfort A, Papadopoulos T, Olivi E et al. OpenMEEG: opensource software for quasistatic bioelectromagnetics. *Biomed Eng Online* 2010;9:1.
- Halberstadt J, Winkielman P, Niedenthal PM et al. Emotional conception: how embodied emotion concepts guide perception and facial action. *Psychol Sci* 2009;20:1254–61.
- Hendricks MA, Conway CM, Kellogg RT. Using dual-task methodology to dissociate automatic from nonautomatic processes involved in artificial grammar learning. *J Exp Psychol Learn Mem Cogn* 2013;39:1491.
- Itabashi R, Nishio Y, Kataoka Y et al. Damage to the left precentral gyrus is associated with apraxia of speech in acute stroke. *Stroke* 2016;47:31–6.
- Jacoby LL. A process dissociation framework: separating automatic from intentional use of memory. *J Mem Lang* 1991;30:513–41.
- Jeste SS, Kirkham N, Senturk D et al. Electrophysiological evidence of heterogeneity in visual statistical learning in young children with ASD. *Dev Sci* 2015;18:90–105.
- Jost E, Conway CM, Purdy JD et al. Exploring the neurodevelopment of visual statistical learning using event-related brain potentials. *Brain Res* 2015;1597:95–107.
- Klein A, Mensh B, Ghosh S et al. Mindboggle: automated brain labeling with multiple atlases. *BMC Med Imag* 2005;5:1.
- Kirkham NZ, Slemmer JA, Johnson SP. Visual statistical learning in infancy: evidence for a domain-general learning mechanism. *Cognition* 2002;83:B35–42.
- Kybic J, Clerc M, Abboud T et al. A common formalism for the integral formulations of the forward EEG problem. *IEEE Trans Med Imaging* 2005;24:12–28.
- Lamme VA. Why visual attention and awareness are different. *Trends Cogn Sci* 2003;7:12–8.
- Lamme VA, Roelfsema PR. The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci* 2000;23:571–9.
- Lee TW, Girolami M, Sejnowski TJ. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural computation* 1999;11:417–41.
- Li L, Miller EK, Desimone R. The representation of stimulus familiarity in anterior inferior temporal cortex. *J Neurophysiol* 1993;69:1918–29.
- Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Meth* 2007;164:177–90.
- McIntosh AR, Rajah MN, Lobaugh NJ. Interactions of prefrontal cortex in relation to awareness in sensory learning. *Science* 1999;284:1531–3.
- McIntosh AR, Rajah MN, Lobaugh NJ. Functional connectivity of the medial temporal lobe relates to learning and awareness. *J Neurosci* 2003;23:6520–8.
- McCarthy G, Wood CC. Scalp distribution of event-related potentials: an ambiguity associated with analysis of variance models. *Electroencephalogr Clin Neurophysiol* 1985;62:203–8.
- Misyak JB, Christiansen MH, Tomblin JB. Sequential expectations: the role of prediction-based learning in language. *Topics Cogn Sci* 2010;2:138–53.
- Moratti S, Clementz BA, Gao Y et al. Neural mechanisms of evoked oscillations: stability and interaction with transient events. *Hum Brain Mapp* 2007;28:1318–33.
- Newman AJ, Tremblay A, Nichols ES et al. The influence of language proficiency on lexical semantic processing in native and late learners of English. *J Cogn Neurosci* 2012;24:1205–23.
- Perruchet P, Pacton S. Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn Sci* 2006;10:233–8.
- Preston AR, Gabrieli JD. Dissociation between explicit memory and configural memory in the human medial temporal lobe. *Cerebral Cortex* 2008;18:2192–207.
- Preuschoff K, Quartz SR, Bossaerts P. Human insula activation reflects risk prediction errors as well as risk. *J Neurosci* 2008;28:2745–52.
- Pritchett S, Zilberg E, Xu ZM et al. Peak and averaged bicoherence for different EEG patterns during general anaesthesia. *Biomed Eng Online* 2010;9:1.
- Reber AS. Implicit learning and tacit knowledge. *J Exp Psychol Gen* 1989;118:219–35.
- Reber PJ, Squire LR. Parallel brain systems for learning with and without awareness. *Learn Mem* 1994;1:217–29.
- Reber PJ. The neural basis of implicit learning and memory: a review of neuropsychological and neuroimaging research. *Neuropsychologia* 2013;51:2026–42.
- Ro T, Breitmeyer B, Burton P et al. Feedback contributions to visual awareness in human occipital cortex. *Curr Biol* 2003;13:1038–41.
- Rüsseler J, Hennighausen E, Münte TF et al. Differences in incidental and intentional learning of sensorimotor sequences as revealed by event-related brain potentials. *Brain Res Cogn* 2003;15:116–26.
- Rypma B, Prabhakaran V, Desmond JE et al. Load-dependent roles of frontal brain regions in the maintenance of working memory. *Neuroimage* 1999;9:216–26.
- Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science* 1996;274:1926–8.
- Saffran JR, Newport EL, Aslin RN et al. Incidental language learning: listening (and learning) out of the corner of your ear. *Psychol Sci* 1997;8:101–5.

- Schapiro AC, Kustner LV, Turk-Browne NB. Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr Biol* 2012;**22**:1622–7.
- Skosnik PD, Mirza F, Gitelman DR et al. Neural correlates of artificial grammar learning. *Neuroimage* 2002;**17**:1306–14.
- Song S, Howard JH, Howard DV. Implicit probabilistic sequence learning is independent of explicit awareness. *Learn Mem* 2007;**14**:167–76.
- Tabullo Á, Sevilla Y, Segura E et al. An ERP study of structural anomalies in native and semantic free artificial grammar: evidence for shared processing mechanisms. *Brain Res* 2013;**1527**:149–60.
- Tadel F, Baillet S, Mosher JC et al. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput Intell Neurosci* 2011;**2011**:8.
- Tadel F, Bock E, Mosher JC et al. Tutorial 23: Scouts. 2016. <http://neuroimage.usc.edu/brainstorm/Tutorials/Scouts> (1 March 2017, date last accessed).
- Tallon-Baudry C. On the neural mechanisms subserving consciousness and attention. *Front Psychol* 2012;**2**:397.
- Toro JM, Sinnott S, Soto-Faraco S. Speech segmentation by statistical learning depends on attention. *Cognition* 2005;**97**:B25–34.
- Travers BG, Klinger MR, Mussey JL et al. Motor-linked implicit learning in persons with autism spectrum disorders. *Autism Res* 2010;**3**:68–77.
- Tremblay A, Ransijn J. *Model Selection and Post-hoc Analysis for (G)LMER Models*. 2015. <https://cran.rproject.org/web/packages/LMERConvenienceFunctions/LMERConvenienceFunctions.pdf> (14 September 2015, date last accessed).
- Turk-Browne NB, Junge JA, Scholl BJ. The automaticity of visual statistical learning. *J Exp Psychol Gen* 2005;**134**:522–64.
- Turk-Browne NB, Scholl BJ, Chun MM et al. Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J Cogn Neurosci* 2009;**21**:1934–45.
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 2002;**15**:273–89.
- Ullsperger M, Harsay HA, Wessel JR et al. Conscious perception of errors and its relation to the anterior insula. *Brain Struct Funct* 2010;**214**:629–43.
- Urbach TP, Kutas M. The intractability of scaling scalp distributions to infer neuroelectric sources. *Psychophysiology* 2002;**39**:791–808.
- van Gaal S, Ridderinkhof KR, Fahrenfort JJ et al. Frontal cortex mediates unconsciously triggered inhibitory control. *J Neurosci* 2008;**28**:8053–62.
- van Zuijen TL, Simoens VL, Paavilainen P et al. Implicit, intuitive, and explicit knowledge of abstract regularities in a sound sequence: an event-related brain potential study. *J Cogn Neurosci* 2006;**18**:1292–303.
- Veroude K, Norris DG, Shumskaya E et al. Functional connectivity between brain regions involved in learning words of a new language. *Brain Lang* 2010;**113**:21–7.
- West BT, Welch KB, Galecki AT. *Linear Mixed Models: A Practical Guide Using Statistical Software, Second Edition*. Boca Raton, FL: Taylor & Francis, 2014.
- Wessel JR, Haider H, Rose M. The transition from implicit to explicit representations in incidental learning situations: more evidence from high-frequency EEG coupling. *Exp Brain Res* 2012;**217**:153–62.
- Wierda SM, Van Rijn H, Taatgen NA et al. Distracting the mind improves performance: an ERP study. *PLoS One* 2010;**5**:e15024.