

Visual statistical learning is related to natural language ability in adults: An ERP study



Jerome Daltrozzo^{a,*}, Samantha N. Emerson^a, Joanne Deocampo^a, Sonia Singh^a, Marjorie Freggens^a, Lee Branum-Martin^a, Christopher M. Conway^{a,b,*}

^a Department of Psychology, Georgia State University, Atlanta, GA, USA

^b Neuroscience Institute, Georgia State University, Atlanta, GA, USA

ARTICLE INFO

Article history:

Received 26 August 2015

Revised 6 October 2016

Accepted 17 December 2016

Keywords:

Implicit learning
Statistical learning
Sequential learning
P300
P600

ABSTRACT

Statistical learning (SL) is believed to enable language acquisition by allowing individuals to learn regularities within linguistic input. However, neural evidence supporting a direct relationship between SL and language ability is scarce. We investigated whether there are associations between event-related potential (ERP) correlates of SL and language abilities while controlling for the general level of selective attention. Seventeen adults completed tests of visual SL, receptive vocabulary, grammatical ability, and sentence completion. Response times and ERPs showed that SL is related to receptive vocabulary and grammatical ability. ERPs indicated that the relationship between SL and grammatical ability was independent of attention while the association between SL and receptive vocabulary depended on attention. The implications of these dissociative relationships in terms of underlying mechanisms of SL and language are discussed. These results further elucidate the cognitive nature of the links between SL mechanisms and language abilities.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Historically, two areas of controversy in the study of language are: (1) to what extent language is innately specified versus learned through experience and (2) to what extent the mechanisms activated during language acquisition are specific to language processing (i.e., domain-specificity, Chomsky, 2011; Gallistel, 2011; Newport, 2011). Increasingly, research findings and theory have suggested that language is largely experience-dependent and mediated by domain-general processes that allow organisms to perceive and learn environmental patterns (Aslin & Newport, 2012; Christiansen & Chater, 2015; Erickson & Thiessen, 2015; Saffran, Aslin, & Newport, 1996). This type of implicit learning or statistical learning (SL) can be defined as the ability to extract probabilities from a series of discrete elements and to form expectations about upcoming events based on that information (Fiser & Aslin, 2001; Kirkham, Slemmer, & Johnson, 2002; Krogh, Vlach, & Johnson, 2013; Saffran, 2003; Siegelman & Frost, 2015; Turk-Browne, Scholl, Johnson, & Chun, 2010). SL can

be construed as arising from general principles of cortical plasticity, in which experience with a specific type of input leads to improvements in the processing and efficiency of neural networks that are devoted to processing that type of input (Reber, 2013).

SL is particularly important for the development of language (Arciuli & Torkildsen, 2012; Conway & Pisoni, 2008; Gervain & Mehler, 2010; Gogate & Hollich, 2010; Kuhl, 2004; Reber, 1967) including the acquisition of word boundaries (Saffran et al., 1996), syntax (Ullman, 2004), word order (Conway, Bauernschmidt, Huang, & Pisoni, 2010), and receptive vocabulary (Ellis, Robledo, & Deák, 2014). The link between SL and language ability is further supported by research showing that impairments to SL may contribute to some of the observed language and learning difficulties (Nicolson & Fawcett, 2007; Ullman & Pierpont, 2005) such as in dyslexia (Du & Kelly, 2013; Howard, Howard, Japikse, & Eden, 2006), specific language impairment (Evans, Saffran, & Robe-Torres, 2009), autism (Jeste et al., 2015), and populations with language delays caused by a period of deafness early in development (Conway, Pisoni, Anaya, Karpicke, & Henning, 2011).

Early in the study of SL, most research focused on understanding its underlying mechanisms, processes, and constraints. More recently, research endeavors have taken an individual-differences approach to determine whether SL abilities are associated with

* Corresponding authors at: Department of Psychology, P.O. Box 5010, Georgia State University, Atlanta, GA 30302-5010, USA.

E-mail addresses: jerome.daltrozzo@gmail.com (J. Daltrozzo), cconway@gsu.edu (C.M. Conway).

language skill as a way to provide a more direct empirical connection between the two constructs. In these studies, performance on SL tasks has been found to be significantly correlated with measures of language performance (e.g., Arciuli & Simpson, 2012; Conway et al., 2010; Ellis et al., 2014; Misyak, Christiansen, & Tomblin, 2010; Spencer, Kaschak, Jones, & Lonigan, 2014). However, with few exceptions such as Morgan-Short, Faretta-Stutenberg, Brill, Carpenter, and Wong (2014), Morgan-Short et al. (2015) who reported correlations between SL performance measured with the Weather Prediction Task (Foerde, Knowlton, & Poldrack, 2006) and neural activation of second language learning in adults, most of this research has been based only on behavioral data. This represents a potential limitation because behavioral measures can only assess the final result of a complex chain of multiple operations that may include perceptual/cognitive mechanisms, response selection, and motor preparation and execution. In contrast, neural measures such as event-related potentials (ERPs) follow more directly (in real-time) the processing of information of the cognitive mechanisms of interest (e.g. Coch & Holcomb, 2003; Schwartz, Federmeier, van Petten, Salmon, & Kutas, 2003). Additionally such measures do not necessarily require an overt motor response.

Therefore the nature of the linkage between SL and language ability could be further explored with a neurophysiological approach. This approach remains to date very limited (Christiansen, Conway, & Onnis, 2012; Tabullo, Sevilla, Segura, Zanutto, & Wainseboim, 2013). For example, Christiansen et al. (2012) measured the ERP correlates of SL using an artificial language paradigm that required participants to learn the mapping between a visual scene of unfamiliar shapes and non-word sequences. The visual scenes and non-word sequences were generated according to an artificial grammar that specified the structural relations among the stimuli. After the artificial grammar had been learned, violations of the structural regularities elicited ERP correlates of SL that were very similar in amplitude and scalp topography to a modulation of the P600 component, an ERP correlate of syntactic violations in natural language. This similarity was further confirmed statistically by the lack of significant difference between the P600 effects (ERP to syntactic violations minus ERP to syntactically correct trials) observed with both artificial and natural language grammars and a significant correlation between the ERP effects on the two tasks. The authors argued that the P600 generated by the two types of syntactic violations were indistinguishable and hence that SL (observed with the artificial grammar paradigm) and language processing (observed with syntactic processing within a natural grammar paradigm) tap into the same underlying neural mechanisms providing evidence for a domain-general processing mechanism underlying both artificial and natural grammar processing, albeit still within a language-like context.

However, to further test the extent to which domain-general SL is related to natural language processing, it is necessary to employ SL tasks that are even less language-like than the one used by Christiansen et al. (2012). For instance, that study incorporated non-word sequences that are likely to tap into phonological coding processes. Christiansen et al. (2012) also incorporated an artificial language and grammar that was meant to (at least superficially) reflect basic aspects of natural language, with both a referent system (or vocabulary) and a simple form of syntax. Thus, the SL task used in this experiment incorporated elements of language (phonology, visual referent system, etc.) and therefore may have been limited in its ability to fully examine the domain-generality of SL.

One experimental task that might be useful in this regard is that used by Jost, Conway, Purdy, Walk, and Hendricks (2015). In this variation of the standard oddball paradigm (Squires, Squires, & Hillyard, 1975), participants were presented with sequences of

visual, non-linguistic stimuli (colored circles; randomly assigned between participants). Some stimuli were “standards” (i.e., that were presented repeatedly). The remaining “deviant” stimuli belonged to one of two different categories: predictor or target. The participants were asked to respond to the target stimuli by pressing a button. Participants were not told that the other type of deviant stimuli—the predictor stimuli—would precede the target with one of three different levels of fixed contingent probability in relation to the target (i.e., “high”, “low”, and “null” predictability with 90%, 20%, and 0% probabilities of the target following each predictor type, respectively). With exposure to the three predictor-target probabilities, participants learned these statistical associations as revealed by a late positive ERP component that increased with predictor-target probability. The ERP component displayed the greatest amplitude for the “high” compared to the “low” or “null” predictors and hence was interpreted as an ERP index of SL.

Jost et al. (2015) proposed that the observed late positivity was a P300, an ERP component reflecting stimulus evaluation-based decision or categorization processes as well as memory updating of contextual representations (for a review, see Polich, 2007). However, given the late latency of the ERP effect reported by Jost et al. (2015), an alternative interpretation could be that they found a modulation of the P600 component. (Note that Coulson, King, and Kutas (1998) have suggested that the P300 and P600 may actually be the same ERP component, but see Frisch, Kotz, von Cramon, & Friederici, 2003). Interestingly, there is extensive evidence that the P600 is a correlate of syntactic violation for both artificial grammar paradigms (Christiansen et al., 2012; Friederici, Steinhauer, & Pfeifer, 2002) and natural grammar paradigms (Gunter, Stowe, & Mulder, 1997; Hagoort, Brown, & Groothusen, 1993; Neville, Nicol, Barss, Forster, & Garrett, 1991). Therefore, it is possible that the component observed by Jost et al. (2015) was in fact a P600 reflecting statistical rule-based predictor-target contingency variation. These results may thus provide further, albeit somewhat indirect, support for a link between SL and language processing. To provide a more direct test of this relationship, here we examine whether ERP correlates elicited in a task similar to Jost et al. (2015) are associated with participants' language ability.

In the present study, the statistical structure of input sequences provided to participants was similar to the one used by Jost et al. (2015). Therefore, we expected to find a similar late positivity that would increase with target predictability, i.e. with predictor-target statistical contingency. Hence, our operational definition of SL at the neurophysiological level was a centro-parietal late positive ERP effect (between experimental “Predictability” conditions with varying predictor-target statistical contingency). In addition, similar to most behavioral studies of SL such as serial reaction time tasks (Nissen & Bullemer, 1987), our operational definition of SL was also defined by participants' response times (RTs) on the SL task with faster RT to targets as predictability (predictor-target statistical contingency) increases i.e. fastest RT when the target follows the “high predictability” (HP) rather than the “low predictability” (LP) stimuli.

To test the relationship between SL and language ability, we used three measures of language performance. Given that SL has been implicated in vocabulary acquisition (Ellis et al., 2014; Howard, Howard, Japikse, & Eden, 2006), receptive vocabulary was measured with the “Peabody Picture Vocabulary Test IV” (PPVT-IV; Dunn & Dunn, 2007). Because previous research has suggested a link between SL and performance on natural language grammar (e.g., Christiansen et al., 2012; Tabullo et al., 2013), the second was a measure of grammatical ability, the “Grammaticality Judgment” subtest of the Comprehensive Assessment of Spoken Language (CASL; Carrow-Woolfolk, 1999). The last measure assessed the ability to predict words within a sentence: the “Sen-

tence Completion” subtest of the CASL (Carrow-Woolfolk, 1999). This was included because Conway et al. (2010) observed a behavioral correlation between SL and the ability to predict the final words in spoken sentences.

One additional contribution of the present study is that it also examines whether the ERP correlates of SL are related to language ability while controlling for each individual's general level of selective attention. This control is important to demonstrate the specificity of the relationship between SL and language ability. Indeed, general selective attention could, in principle, affect performance on any task (e.g. Schneider & Shiffrin, 1977) including SL (for a review, see Daltrozzo & Conway, 2014) and language assessments. Thus, it is possible that inter-individual variability in attention may become a confounding variable if not controlled. That is, a given participant may devote a high level of attention to an SL task and also to language assessments thereby scoring better on all measures compared to a participant who devotes less attention and thereby scores lower overall. This general cognitive effect, if not controlled, could drive any observed interactions between SL and language ability. Therefore, in the present study we attempted to control for attention by requiring participants to complete an additional task measuring general selective attention, and then treating it as a covariate in our analyses.

We predicted that, consistent with previous evidence of a link between SL and language ability, our behavioral (i.e. RT effect between two experimental conditions of target “Predictability”: HP and LP) and neurophysiological (i.e. centro-parietal late positive ERP effect between HP and LP) operational definitions of visual SL would be associated with measures of language ability in adults independently of variations of attention.

2. Materials and methods

In this within-subject design, participants were first assessed for their general level of selective attention with a Flanker task (Eriksen & Eriksen, 1974) and then completed a visual SL task while ERPs were recorded. Finally, participants were administered three standardized measures of language ability.

2.1. Participants

Seventeen participants (11 females, $M = 23.94$ years, $SD = 8.61$, 18–49 years) from Georgia State University participated in the study to receive class credit. Participants reported no language, cognitive, neurological, or psychological deficits. All were native English speakers and all except two were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). All participants provided written informed consent to the study, which was conducted in accordance with the guidelines of the Declaration of Helsinki and approved by the local Ethics Committee (the Institutional Review Board of Georgia State University).

2.2. Reliability of our measures

The reliability of our behavioral and neurophysiological measures were computed from the data of our sample of participants using the standard Cronbach's α coefficient (Cronbach, 1951) as well as the “greatest lower bound” (*glb*, Woodhouse & Jackson, 1977) because the *glb* is a better lower bound estimate of the reliability than α (Sijtsma, 2009) with: $\alpha < glb < \text{true reliability} < 1$. α and *glb* were computed using the *cronbach()* function from the *psy* library and the *glb()* function from the *psych* library in R version 3.2.0.

2.3. Language measures

Participants were administered three standardized measures of language: the Peabody Picture Vocabulary Test, 4th Edition (PPVT-IV; Dunn & Dunn, 2007); and the Grammaticality Judgment (GJ) and Sentence Completion (SC) subtests of the Comprehensive Assessment of Spoken Language (CASL; Carrow-Woolfolk, 1999). While the CASL is only normed through age 21 years and 11 months, some of our participants exceeded that age. Consequently, we chose to analyze raw scores for all measures instead of standardized scores that depend on the CASL norms. Although the CASL subtests were not intended for use with participants above the age of 21, the raw scores still give an absolute measure of grammatical and sentence prediction abilities, which is appropriate for the purpose of the present study that is focused on these abilities in each of our participants in relation to the other participants of the study rather than in relation to an age-peer population norm.

2.3.1. Peabody picture vocabulary test-IV

The PPVT-IV is a widely used assessment of standard American English receptive vocabulary in which examinees are asked to select from four full-color drawings the one that best depicts the meaning of a spoken word. There are a total of 228 items representing various parts of speech on the test, each of which is scored as correct or incorrect, for a maximum possible raw score of 228. However, items are arranged into sets of 12, and administration discontinues when a participant responds incorrectly to eight items within one set. Dunn and Dunn (2007) provide Cronbach's α by age range. Our participants spanned 6 age ranges: 17–18 $\alpha = 0.96$, 19–21 $\alpha = 0.95$, 22–24 $\alpha = 0.93$, 25–30 $\alpha = 0.95$, 31–40 $\alpha = 0.94$, and 41–50 $\alpha = 0.97$. Based on the data of our sample of participants (from the 45 items that were completed by all participants and where the standard deviation across participants was not null) we computed the lower bound of the reliability of the PPVT-IV measure according to α and *glb* (see “Reliability of our Measures” section above): $\alpha = 0.734$; *glb* = 0.849.

2.3.2. Grammaticality Judgment

The GJ subtest is part of the Syntactic Language Category of the CASL. It measures an individual's ability to identify and correct grammatical errors within a sentence. Errors are assessed in ten syntactic areas: noun-verb agreement, noun number, verb tense, pronouns, negatives, prepositions, irregular forms, direct/indirect objects, active/passive voice coordination, and embedding. Sentences both with and without errors are read aloud to the examinee, and he or she is asked to determine if “the sentence is the correct way to speak in class.” If not, the participant is asked to change a single word to correct the sentence. Practice sentences include, “The boy are happy” (for which the correction could be “The **boys** are happy” or “The boy **is** happy”) and “My mother is at home” (which is correct as is). High performance on this test requires both grammar ability and metalinguistic skill that allows for separately attending to the meaning and form of a sentence. According to Carrow-Woolfolk (1999), the ability to judge grammaticality is different from the ability to choose appropriate morphemes and requires that basic sentence patterns of the native language, English in this case, be stored at a deep level. The GJ subtest measures how deep an examinee's knowledge is by scoring ability to recognize an error (1 point) and the ability to correct it (1 point) separately. The maximum raw score is 103. Carrow-Woolfolk provides Spearman-Brown adjusted Rasch split-half internal reliability calculations by age band. Our participants spanned two age bands and beyond: 16–19 $r = 0.90$ and 19–22 $r = 0.85$. Based on the data of our sample of participants (from the 13 items that were completed by all participants and where

the standard deviation across participants was not null) we computed the lower bound of the reliability of the GJ measure according to α and glb (see “Reliability of our Measures” section above): $\alpha = 0.696$; $glb = 0.905$.

2.3.3. Sentence completion

The SC subtest is part of the Lexical/Semantic Language Category of the CASL. It measures an individual's ability to retrieve and express one of a handful of appropriate words that will accurately complete a spoken sentence. The examiner reads a sentence aloud leaving out the last word. The examinee is tasked with completing the sentence using a single grammatically and semantically correct word that is not already contained in the sentence. Appropriate responses include nouns, verbs, adjectives, and adverbs. An example of a practice item is, “In order to start the car, Dad must turn the ____”. Acceptable responses include “key” or “ignition”. As the test advances, sentences become much longer and more complex with more complicated vocabulary and concepts. Performance on this assessment requires understanding of sentence context and use of it to make a prediction, and retrieve and orally express an appropriate word. Understanding of vocabulary, grammar, and event knowledge and memory are critical. The maximum raw score for this subtest is 61. Spearman-Brown split-half reliability for 16–19 years is reported as 0.76 and for 19–22 years as 0.82. Based on the data of our sample of participants (from the 26 items that were completed by all participants and where the standard deviation across participants was not null; since the glb computation on 26 items resulted in a non-positive definite matrix, the computation was performed on only 25 items) we computed the lower bound of the reliability of the SC measure according to α and glb (see “Reliability of our Measures” section above): $\alpha = 0.487$; $glb = 0.696$.

2.4. Flanker task: control of the general level of selective attention

To control for possible variations in the general level of selective attention across participants, we measured the performance on a Flanker task (Eriksen & Eriksen, 1974; Shalev & Tsai, 2003), which was collected immediately before the SL task. The Flanker task is a visual search task measuring a person's ability to detect relevant information in the midst of irrelevant information (Eriksen & Eriksen, 1974). This task is commonly used to test response inhibition and selective attention (Casey et al., 2000; Fenske & Eastwood, 2003; Hübner, Steinhauser, & Lehle, 2010; Lavie, Hirst, de Fockert, & Viding, 2004). It comprises a central target (e.g., ‘<’) flanked by non-target stimuli, which either match the direction of the target in half of the trials (congruent stimuli; e.g., ‘<<<<<’) or are in the opposite direction of the target in the rest of the trials (incongruent stimuli; e.g., ‘>>>>>’). A directional response (left or right) is assigned to the target stimulus. Response times for incongruent trials are typically longer than for congruent trials—a difference known as the Flanker effect (Eriksen & Eriksen, 1974). The Flanker task allows the testing of how well attention is restricted to a particular object or location (Fenske & Eastwood, 2003) and thus can be used as a measure of selective attention (Lavie et al., 2004). Participants were required to give a speeded left- or right-hand response to indicate the direction of the target. The number of correct button-press responses with the left and right hand was counterbalanced across congruent and incongruent experimental conditions. A new trial was presented immediately after the participant's response. Immediately after the Flanker task, participants were administered the SL task. Flanker effect scores were computed for accurate trials only by subtracting mean response times for incongruent trials from mean response times for congruent trials. Wöstmann et al. (2013) report a Cronbach α of 0.68 for this measure. Based on the data of our sample of participants we computed the lower bound of the reliability of the Flanker Effect

measure according to α and glb (see “Reliability of our Measures” section above): $\alpha = 0.472$, $glb = 0.910$.

2.5. Sequential learning task

In the SL task (Fig. 1A) colored circles appeared in the center of the computer screen on a dark background similar to those used in Jost et al. (2015). Stimuli were displayed for 500 ms followed by a dark screen displayed for an additional 500 ms. Thus, the visual stimuli were presented with a 1000 ms stimulus onset asynchrony.

Participants were instructed to press a button as quickly as possible when a given “target” stimulus was presented. Each participant was presented with a target selected pseudo-randomly from among a set of six colored circles stimuli. Unbeknownst to the participants, the sequence of stimuli followed a set of fixed statistical regularities (see Fig. 1A). For each participant, in addition to the target stimulus, one of the stimuli was pseudo-randomly chosen as a “standard” stimulus, one as a “high predictability” (HP) predictor, and one as a “low predictability” (LP) predictor. During each trial the standard stimulus was repeated a pseudorandom number of times. Next, one of the two (HP or LP) predictors was presented, each with a 50% probability of occurrence. The HP predictor was followed by the target in 90% of the trials and by the standard in 10% of the trials. The LP predictor was followed by the target in 20% of the trials and by the standard in 80% of the trials. Each trial concluded with a second series of standards of a random length.

For each predictability condition (HP and LP) there were 50 trials for a total of 100 trials divided among five blocks of 20 trials. All trials were randomly ordered across the two predictability conditions (HP and LP) in a continuous fashion such that the participant was unable to distinguish one trial from another. A break lasting a minimum of 30 s was given between each block. Stimuli were presented on a Dell Optiplex 755 computer running E-Prime version 2.0.8.90.

2.6. Electroencephalography acquisition

While the participant performed the SL task, the electroencephalograph (EEG) was recorded from 256 scalp electrodes using an Electrical Geodesic Inc. sensor net (Fig. 1B) and was pre-processed using Net Station version 4.3.1 with subsequent processing using custom scripts written in MATLAB (version R2012b 8.0.0783, MathWorks) and the EEGLAB toolbox (version 10.2.2.2.4a; Delorme & Makeig, 2004). Active electrode impedances were kept below 50 k Ω . The EEG was acquired with a 0.1–100 Hz band-pass filter at 250 Hz with vertex reference and then re-referenced to the average reference, resampled to 256 Hz (to allow filtering with Fast Fourier Transformation), and low-pass filtered at 30 Hz. Participants were instructed to refrain from blinking throughout the experiment. Eye blink artifacts were corrected by independent component analysis. Other remaining artifacts were removed manually. This procedure removed 18.2% of the trials [HP: $M = 17.5\%$, $SD = 12.5$; LP: $M = 18.8\%$, $SD = 14.0$; two-tailed paired Student: $t(16) = 0.723$, $p = 0.480$, $N = 17$; Wilcoxon matched-pairs signed-ranks test: $W = -11$, $p = 0.762$, n pairs = 15, two pairs excluded because values in each sample were equal]. The continuous EEG was segmented into epochs –200 ms to +1000 ms with respect to the predictor onset. ERPs were baseline-corrected at 200 ms prestimulus. Separate ERPs were computed for each participant, predictability condition, and electrode. All experimental sessions were conducted in a 132 square foot double-walled, sound-deadened acoustic chamber.

2.7. Statistical procedures

Individual mean response times (RT) of the SL task to target stimuli following the HP and LP predictors and mean amplitude

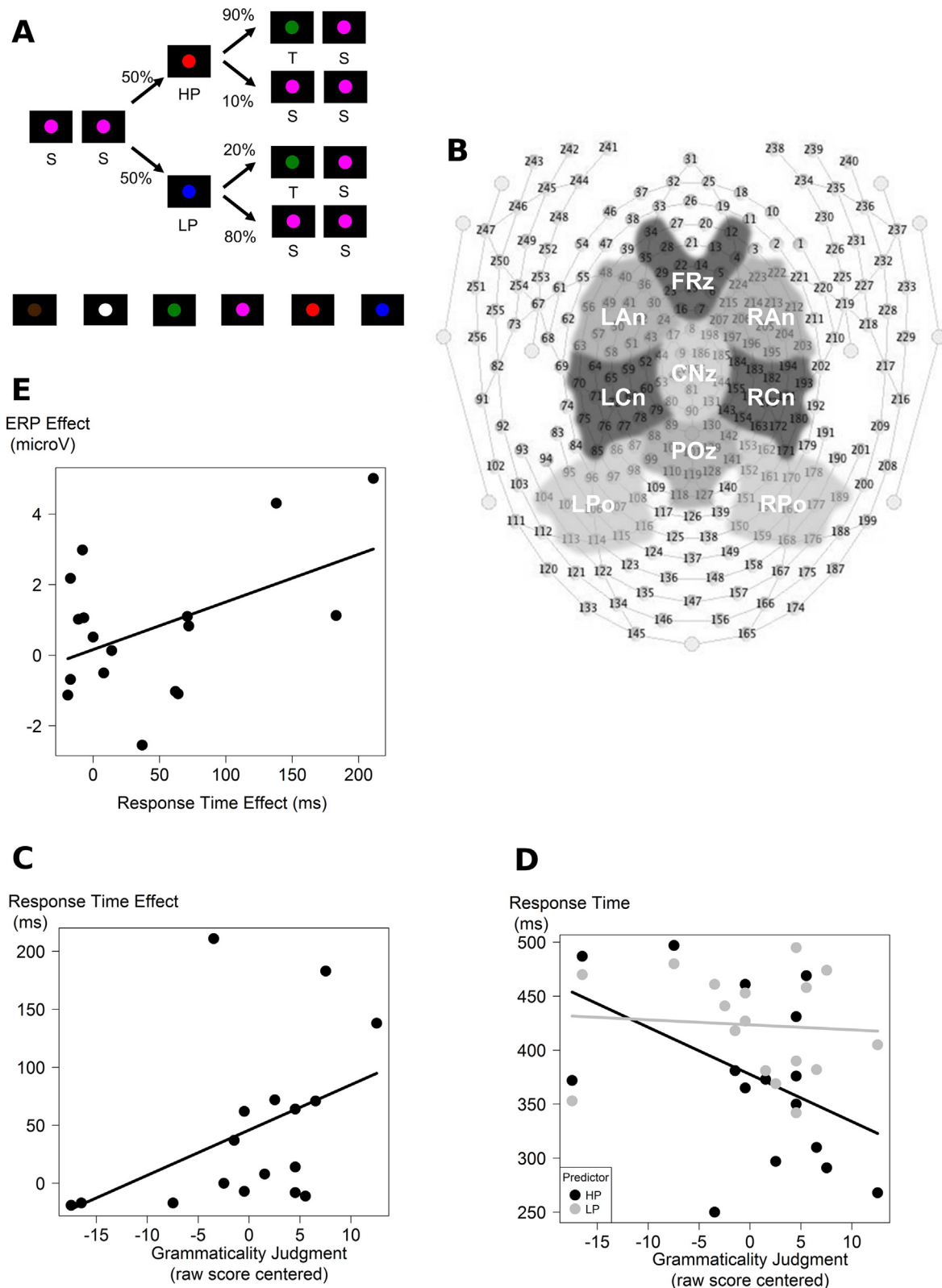


Fig. 1. (A) Sample sequence of colored circles. HP: "High Predictability" predictor, LP: "Low Predictability" predictor, T: Target, S: Standard. For each participant, HP, LP, T, and S were pseudo-randomly assigned to one of six colored circles (brown, white, green, pink, red, and blue; see bottom of the figure). (B) 256 Electrical Geodesic Inc. sensor net for EEG recordings and the nine regions of interest. 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 topographic analyses. (C) Scatter plot between the SL Response Time (RT) effect and the Grammaticality Judgment (GJ) raw score centered. RT effect: RT in the Low Predictability condition (LP) minus RT in the High Predictability condition (HP). Vertical unit: millisecond. (D) Scatter plot between the RT in the high and low predictability conditions and the GJ raw score centered. This figure indicates reduced RT as GJ raw score increased in only the HP condition (dark line), i.e. higher SL with higher GJ raw score was related to learning the predictor-target statistical contingency in the HP condition. The gray line indicates RT as a function of GJ raw score in the LP condition. Vertical unit: millisecond. (E) Scatter plot between the neurophysiological and behavioral operational definitions of sequential learning. SL ERP effect (ERP in the HP condition minus ERP in the LP condition) over the centro-parietal (POz) region of interest between 500 ms and 1000 ms (Unit: microvolts) - SL RT effect (RT in the LP condition minus RT in the HP condition) (Unit: ms). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ERPs time-locked to the predictors between 500 ms and 1000 ms (time-window based on the grand average; Fig. 2) were analyzed with separate linear mixed models (LMM, see [Supplementary Materials and Methods](#)), one for each language score.

To estimate the lower bound of the reliability of our behavioral and our neurophysiological operational definitions of SL based on the data of our sample of participants, we computed the SL RT effect (RT in the LP condition minus RT in the HP condition) and the SL ERP effect (mean ERP between 500 ms and 1000 ms over the centro-parietal ROI (POz) in the HP condition minus corresponding mean ERP in the LP condition). The lower bounds of the corresponding reliabilities are: $\alpha = 0.892$ and $glb = 0.974$ for the SL RT effect; and $\alpha = 0.897$ and $glb = 0.944$ for the SL ERP effect.

3. Results

3.1. Language measures mean and correlation

The mean raw scores on the three language measures were: GJ: $M = 78.5$, $SD = 7.9$; PPVT-IV: $M = 205.9$, $SD = 7.3$; and SC: $M = 49.1$, $SD = 3.2$. These scores were uncorrelated [GJ vs. PPVT-IV: $r_s = 0.095$, $p = 0.717$; GJ vs. SC: $r_s = -0.150$, $p = 0.565$; PPVT-IV vs. SC: $r_s = 0.440$, $p = 0.077$; $N = 17$; two-tailed].¹

3.2. Sequential learning task response times

The applied LMMs for the analysis of the RTs are described in [Tables 1 and 2](#) (See [Supplementary Material](#)). The LMM on individual mean RTs showed a main effect of Predictability [$F(1, 15) = 8.04$; $p = 0.013$] indicating faster responses to the target following the HP predictor ($M = 379$ ms, $SD = 71$) compared to the LP predictor ($M = 424$ ms, $SD = 51$) but no interaction that reached significance between Predictability and any of the language scores, although the interaction with GJ was marginally significant [PPVT-IV raw score: $F(1, 15) = 2.08$, $p = 0.169$; GJ raw score: $F(1, 15) = 3.46$, $p = 0.083$; SC raw score: $F(1, 15) = 0.65$, $p = 0.432$]. Because the 2-way interactions were not significant, Attention was not included in any of the models.

Given the marginal significance of the interactions between Predictability and GJ raw score ($p = 0.083$), we tested the correlations between the SL RT effect (RT in the LP condition minus RT in the HP condition) and the GJ raw score, which revealed a positive relationship between the GJ raw score and the SL RT effect [$r_s = 0.522$, $p = 0.031$; $N = 17$; two-tailed] (see [Fig. 1C](#)). [Fig. 1D](#) additionally shows the regression lines for each Predictability condition separately, indicating that RTs to the HP condition were faster as the GJ score increase.

In sum, the behavioral results indicate that with exposure, participants were able (at least in the HP condition) to extract the statistical structure embedded within the sequences as reflected by quicker responses to the target when following a HP predictor than when following a LP predictor. This behavioral measure of SL appeared to be related to the GJ raw score with larger SL RT effect associated with higher performance on this language measure.

3.3. SL Task: event-related potentials

[Fig. 2](#) displays the grand averaged ERPs across all participants time-locked to the two predictors (HP and LP) at the nine regions of interest used for topographic analyses. Visual inspection suggests a larger centro-parietal late positivity between

approximately 500 ms and 1000 ms for the HP predictor compared to the LP predictor, replicating the general effect observed by [Jost et al. \(2015\)](#).

The applied LMMs for the analysis of the ERPs are described in [Tables 1 and 3–5](#) (See [Supplementary Material](#)). The LMM on the mean amplitude ERPs time-locked to the predictors between 500 ms and 1000 ms (time-window based on the grand average, [Fig. 2](#)) revealed a main effect of Predictability [$F(1, 4522.0) = 51.16$; $p < 0.001$] indicating SL had occurred according to our neurophysiological operational definition. To test the correlation between our behavioral and our neurophysiological operational definition of SL, we computed the mean SL ERP effect (i.e. the mean amplitude of the ERP in the HP condition minus the mean amplitude of the ERP in the LP condition) over the centro-parietal (POz) ROI between 500 ms and 1000 ms. The trend for a positive correlation between SL RT effect (RT in the LP condition minus RT in the HP condition) and SL ERP effect did not reach significance with the Spearman test [$r_s = 0.306$, $p = 0.232$; $N = 17$; two-tailed] ([Fig. 1E](#)).

We also found an interaction between Predictability, GJ raw score, and ROI [$F(8, 4522.0) = 2.06$; $p = 0.036$] together with an interaction between Predictability, GJ raw score, ROI, and general level of selective attention (Flanker effect centered) [$F(18, 208.7) = 5.94$; $p < 0.001$] as depicted on [Figs. 3 and 4](#). These interactions remained significant after [McCarthy and Wood \(1985\)](#) correction [$F(8, 4522.0) = 3.56$; $p < 0.001$ and $F(18, 208.7) = 3.21$; $p < 0.001$; respectively]. [Figs. 3 and 4](#) indicate that the relationship between the centro-parietal SL ERP effects and GJ raw score was similar between high and low attention. Specifically, the figures reveal that in the centro-parietal ROI (POz), higher amplitudes for the HP condition (but not the LP condition) are associated with higher language (GJ) scores, and this was true for high and low general selective attention. As indicated in [Figs. 3 and 4](#), the interaction with the general level of selective attention stems mostly from ERP effects at frontal and central ROIs (i.e. LAN, FRz, RAN, LCn, and CNz) with larger (absolute) ERP effects (across GJ scores) under high compared to low attention while the ERP effects at POz, (our operational definition of SL ERP effect, see Introduction) remains similar between high and low attention.

In addition, an interaction was found between Predictability, PPVT-IV raw score centered, and ROI [$F(8, 4522.0) = 3.08$; $p = 0.002$] together with an interaction between Predictability, PPVT-IV raw score, ROI, and general level of selective attention (Flanker effect centered) [$F(18, 208.7) = 4.74$; $p < 0.001$] as depicted on [Figs. 5 and 6](#). These interactions remained significant after [McCarthy and Wood \(1985\)](#) correction [$F(8, 4522.0) = 3.97$; $p < 0.001$ and $F(18, 208.7) = 3.74$; $p < 0.001$; respectively]. [Figs. 5 and 6](#) indicate that the relationship between SL ERP effects and PPVT-IV raw score differed between high and low attention. Specifically, the figures reveal that in the centro-parietal (POz) ROI, higher amplitudes for the HP condition (but not the LP condition) are associated with higher language (PPVT-IV) scores, but this was true only for high general selective attention. As indicated in [Figs. 5 and 6](#), the interaction with the general level of selective attention stems not only from the ERP effects at POz, (our operational definition of SL ERP effect, see Introduction) but also from ERP effects at other cortical sites (i.e. frontal and central ROIs) with large (absolute) ERP effects (across PPVT-IV scores) under high compared to low attention.

Finally, Predictability and SC raw score centered [$F(8, 4522.0) = 0.021$; $p = 0.884$] and Predictability, SC raw score centered, and ROI did not interact significantly [$F(8, 4522.0) = 0.655$; $p = 0.732$]. All the above-mentioned LMM are compared in [Table 1](#) (See [Supplementary Material](#)).

In sum, these neurophysiological results indicate that participants were able to extract the statistical structure embedded

¹ Spearman's correlation r_s was used throughout the manuscript because this correlation coefficient is less sensitive to outliers than Pearson's and thus more appropriate for small sample sizes.

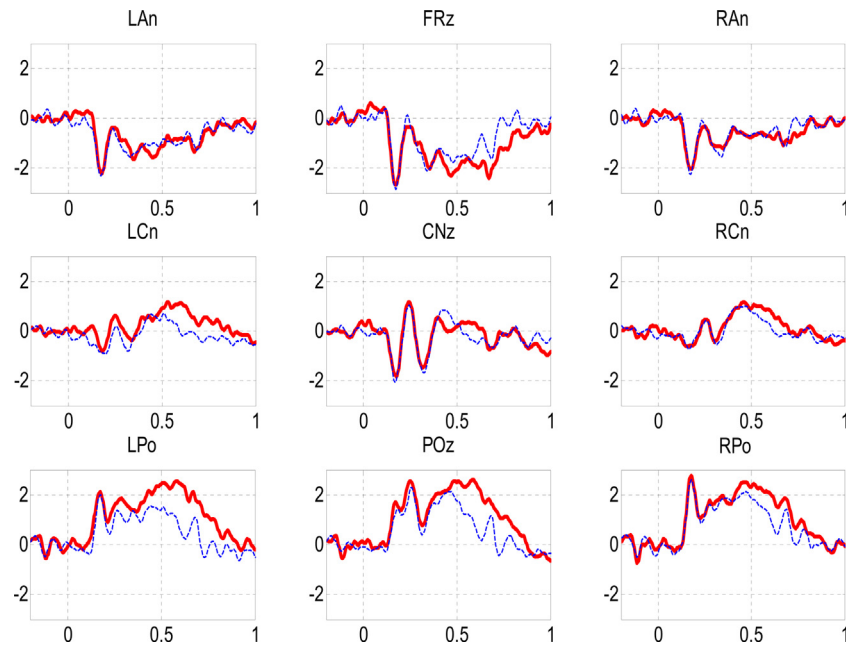


Fig. 2. Grand averaged ERPs. N = 17 participants; Nine regions of interest (see Fig. 1B); Grand averaged ERPs in response to the HP (solid thick lines) and LP (dotted thin lines) predictors (positivity upward in microvolts; time in seconds).

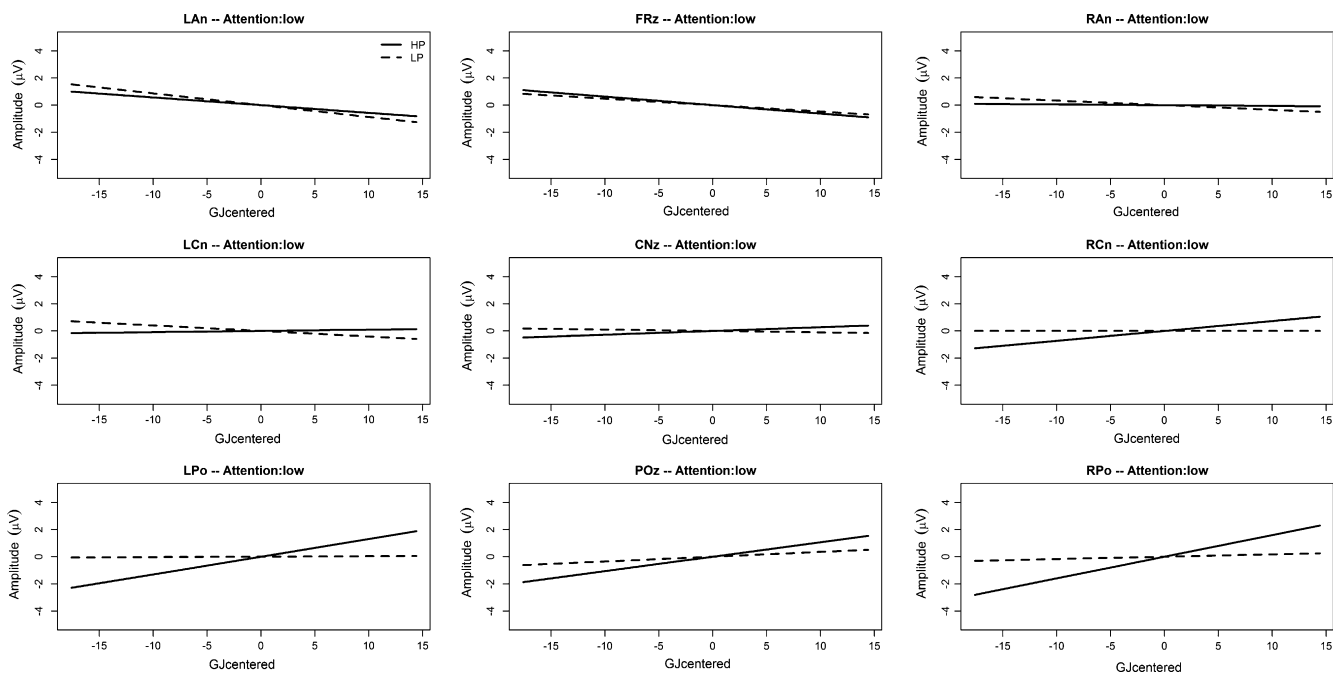


Fig. 3. Predictability Condition \times GJ raw score interaction for each ROI under low attention. The x-axis is the mean-centered GJ raw score, and the y-axis is the LMM estimated ERP mean amplitude between 500 ms and 1000 ms post predictor onset (positivity upward in microvolts). The solid gray line is the simple effect of the mean-centered GJ raw score in the “High Predictability” condition (HP, predictor followed in 90% of the trials by the target), and the broken gray line is the simple effect of the mean-centered GJ raw score in the “Low Predictability” condition (LP, predictor followed in 20% of the trials by the target).

within the sequences. Importantly, the centro-parietal SL ERP effects appeared to be related to the GJ raw score independently of the general level of selective attention, with larger ERP effects being associated with higher performance on this measure of grammatical ability. On the other hand, larger centro-parietal SL ERP effects were associated with higher performance on receptive vocabulary according to the PPVT-IV raw score but only for participants who scored high on general selective attention according to the Flanker task. Finally, no ERP effects were associated with the measure of sentence completion (SC).

4. Discussion

The aim of this study was to investigate whether visual SL is related to language ability. Visual SL was assessed with a predictor-target task that contained statistical contingencies (modeled after Jost et al., 2015). Language ability was assessed with three tasks that explored receptive vocabulary (the PPVT-IV; Dunn & Dunn, 2007), grammatical ability (the GJ subtest of the CASL; Carrow-Woolfolk, 1999), receptive vocabulary (the PPVT-IV), and sentence completion (the SC subtest of the CASL;

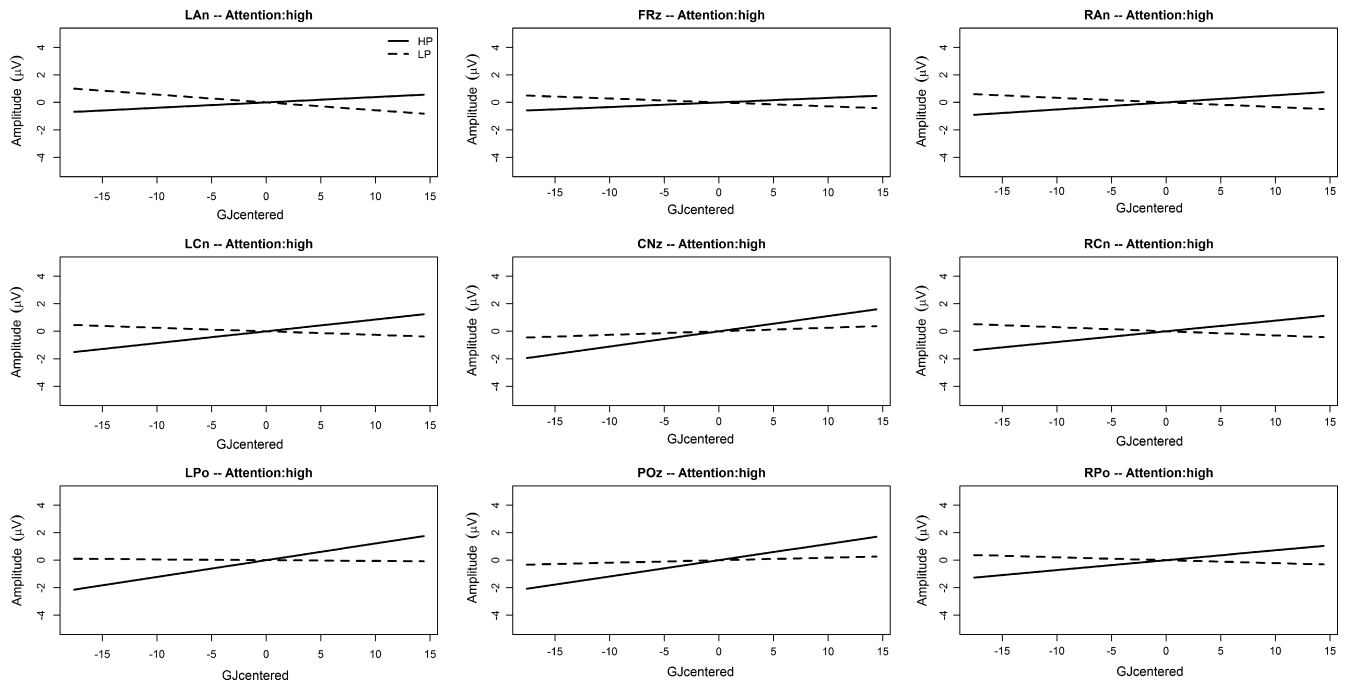


Fig. 4. Predictability Condition \times GJ raw score interaction for each ROI under high attention. The x-axis is the mean-centered GJ raw score, and the y-axis is the LMM estimated ERP mean amplitude between 500 ms and 1000 ms post predictor onset (positivity upward in microvolts). The solid gray line is the simple effect of the mean-centered GJ raw score in the “High Predictability” condition (HP, predictor followed in 90% of the trials by the target), and the broken gray line is the simple effect of the mean-centered GJ raw score in the “Low Predictability” condition (LP, predictor followed in 20% of the trials by the target).

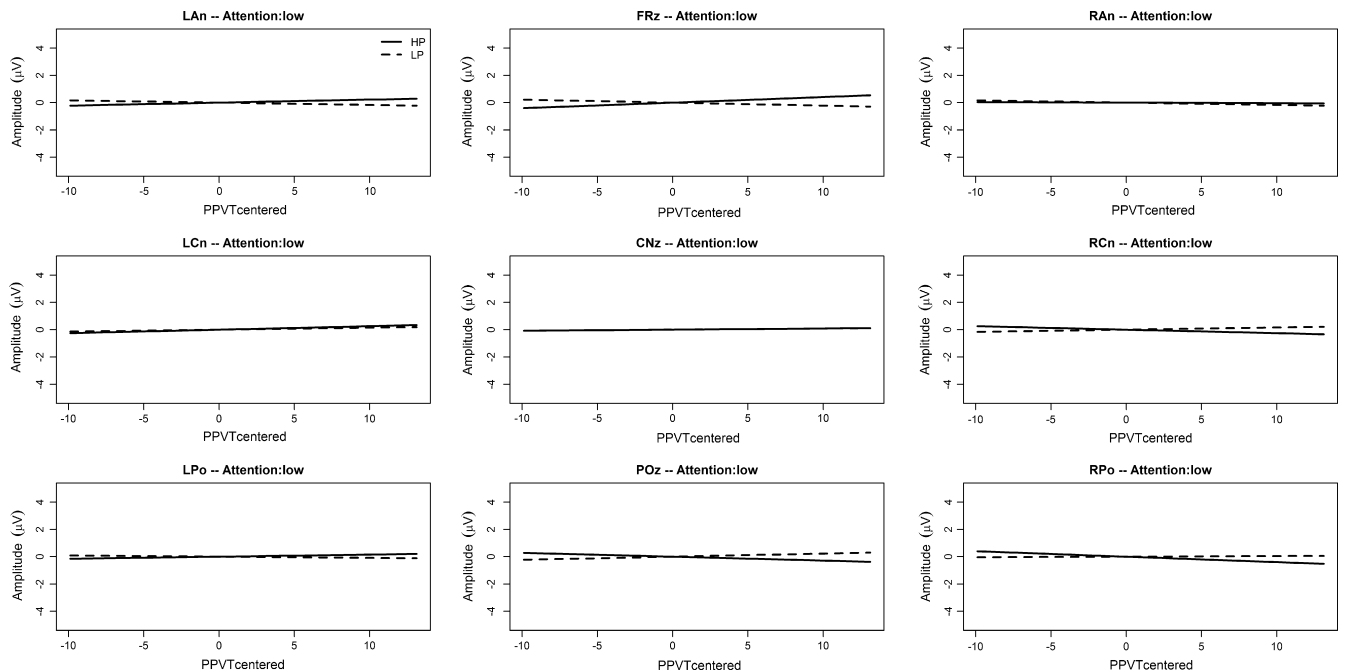


Fig. 5. Predictability Condition \times PPVT-IV raw score interaction for each ROI under low attention. The x-axis is the mean-centered PPVT-IV raw score, and the y-axis is the LMM estimated ERP mean amplitude between 500 ms and 1000 ms post predictor onset (positivity upward in microvolts). The solid gray line is the simple effect of the mean-centered PPVT-IV raw score in the “High Predictability” condition (HP, predictor followed in 90% of the trials by the target), and the broken gray line is the simple effect of the mean-centered PPVT-IV raw score in the “Low Predictability” condition (LP, predictor followed in 20% of the trials by the target).

Carrow-Woolfolk, 1999). Consistent with our prediction, the results indicated—both at a behavioral and neurophysiological level—that visual SL is related to language performance. Importantly, late centro-parietal ERPs indicated: (1) a relationship between SL and grammatical ability in both low and high general selective attention groups, and (2) a relationship between SL and

receptive vocabulary only in the high general selective attention group. Next, we discuss these findings in more detail and consider the implications of this dissociation in terms of the underlying memory systems and mechanisms (e.g., procedural vs. declarative and implicit vs. explicit) that are involved in language processing.

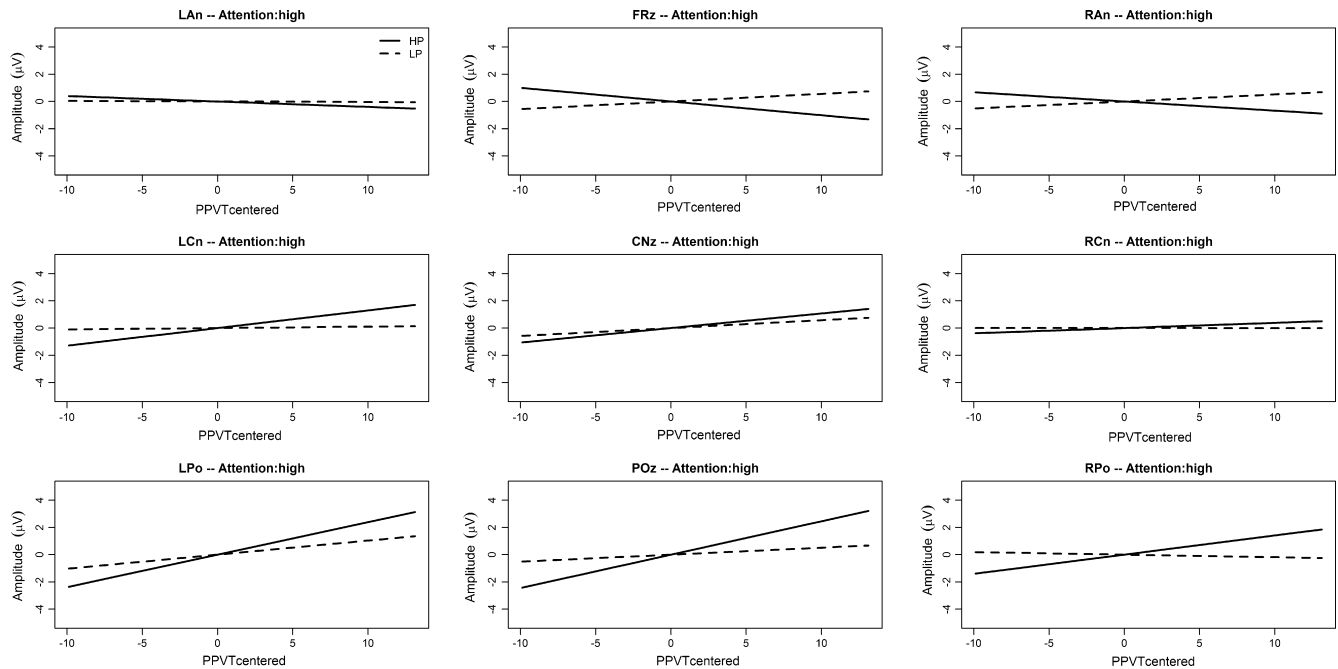


Fig. 6. Predictability Condition \times PPVT-IV raw score interaction for each ROI under high attention. The x-axis is the mean-centered PPVT-IV raw score, and the y-axis is the LMM estimated ERP mean amplitude between 500 ms and 1000 ms post predictor onset (positivity upward in microvolts). The solid gray line is the simple effect of the mean-centered PPVT-IV raw score in the “High Predictability” condition (HP, predictor followed in 90% of the trials by the target), and the broken gray line is the simple effect of the mean-centered PPVT-IV raw score in the “Low Predictability” condition (LP, predictor followed in 20% of the trials by the target).

4.1. Behavioral and neurophysiological correlates of statistical learning

The behavioral data demonstrated that SL occurred as indicated by a SL RT effect: faster RTs to the target when it was highly predictable by the preceding stimulus (i.e. when the target followed the predictor with a 90% probability in the “HP condition”) compared to RTs to the target when it was less predictable by the preceding stimulus (i.e. when the target followed the predictor with a 20% probability in the “LP condition”). The size of this SL RT effect was 45 ms, which is in the typical range of RT effects previously reported in the SL literature (e.g. Baldwin & Kutas, 1997; Eimer, Goschke, Schlaghecken, & Stürmer, 1996; Ferdinand, Mecklinger, & Kray, 2008).

SL was further confirmed with the ERP findings. Visual inspection of grand averages (Fig. 2) in the present study suggests a larger centro-parietal positivity between 500 ms and 1000 ms in response to the HP compared to the LP predictor. This SL ERP effect showed a trend for a positive correlation with the SL RT effect (Fig. 1E) in line with the assumption that the late centro-parietal ERP effect between HP and LP conditions reflects an index of SL. Furthermore, the SL ERP effect observed in the present study closely replicates the findings of Jost et al. (2015), who found a larger centro-parietal positivity in a similar time-window in response to the HP predictor (followed in 90% of the trials by the target, as in our paradigm) compared to the LP predictor (followed in 20% of the trials by the target, also like in our paradigm). Given the similar design to the study of Jost et al. (2015) this replication was expected. Jost et al. (2015) interpreted the occurrence of the P300 to the HP predictor stimuli as an indication that participants formed an association or mental “chunk” between the HP predictor and the target. This interpretation is consistent with previous suggestions that this ERP component reflects chunking processes (Verleger, 1988) and that chunking processes are an important aspect of implicit learning and SL (Perruchet & Pacton, 2006). Perruchet and Pacton proposed that through tracking statistical contingencies, participants form chunks between highly associated

stimuli. In the present study, this means that participants essentially treat the HP predictor at some level to be equivalent to the target itself. This interpretation converges with one view of the P300 wherein this component reflects target detection and evaluation processes during sequence processing (Van Zuijen, Simoens, Paavilainen, Näätänen, & Tervaniemi, 2006). An alternative explanation is that the positive ERP component elicited in the present study could be a P600, a component known to be elicited by syntactic violations from grammars of natural languages (Hagoort et al., 1993) and from artificial grammars (Christiansen et al., 2012). However, the P600 component is typically elicited for violations of syntactic or grammatical structure. Since the current task did not manipulate syntactic violation per se, the observed component here may differ from a typical P600. The possible link between SL and natural language ability is explored further in Section 4.2.

4.2. Relations between statistical learning and language ability

Results from the LMM and correlational analyses revealed that the SL RT effect was related to our measure of grammatical ability (i.e. the GJ raw score). While the LMM analysis suggested a relationship with our measure of receptive vocabulary (i.e., the PPVT-IV raw score), correlational analyses were not able to confirm this result. At a neurophysiological level, our ERP measure of SL—the centro-parietal SL ERP effect, which may be interpreted as a P300 (or possibly a P600 effect, see discussion above)—was related to grammatical ability as evidenced by an interaction between predictor type (HP versus LP), GJ raw score, and ROI between 500 ms and 1000 ms (with a larger centro-parietal positive amplitude in response to the HP predictor compared to the LP predictor for participants with higher GJ raw score). That is, greater visual SL was related to a higher GJ raw score. This relationship was found to be similar (at the centro-parietal cortical area, the cortical site used to define our operational definition of SL ERP effect) in both the high and low general selective attention groups. In addition, a

similar interaction was found with another language measure assessing receptive vocabulary (i.e. the PPVT-IV raw score). However, this later interaction was observed only in the high general selective attention group.

From an information processing point of view, the link between SL and grammatical ability is likely due to SL mechanisms that are used to encode structural and grammatical regularities in spoken language (Christiansen et al., 2012). These results are consistent with previous research demonstrating links between SL and grammatical ability (Christiansen et al., 2012; Conway et al., 2011; Kidd, 2012). One contribution of the current study is that the association between visual SL (according to our ERP operational definition of SL) and grammatical ability of the natural language did not appear to depend on the participants' general level of selective attention (as measured by a Flanker task). Thus, this result suggests that visual SL (the ability to encode statistical regularities and to use this information to predict upcoming units in the visual modality) and grammatical ability of the natural language share implicit mechanisms (i.e. mechanisms that do not require attention to be activated, Shiffrin & Schneider, 1977) and might depend more upon procedural and sequential memory. The same cannot be said of the relationship between visual SL and receptive vocabulary. Indeed, the centro-parietal SL ERP effect interaction with the PPVT-IV raw score varied dramatically with attention, being present only in the high general selective attention group. Possibly, SL and the processing of vocabulary would share explicit mechanisms (i.e. mechanisms that do require attention to be activated; Shiffrin & Schneider, 1977) and relies more on declarative memory. In fact, the dissociation between the SL–grammatical ability and the SL–receptive vocabulary relationships is in line with the Ullman's dual-system model of language (Ullman, 2004), which posits exclusive associations between procedural/sequential memory (a type of implicit or automatic learning) and grammatical processing on one end and declarative memory (a type of explicit memory) and semantic/vocabulary processing on the other end.

The third measure of language, which assessed an individual's ability to retrieve and express one of the few appropriate words that would accurately complete a spoken sentence (the SC raw score) did not interact with SL. Indeed, this measure depends not only on a correct syntactic analysis of the to-be-completed sentences and on single-word vocabulary processing (which may explain the low–though non-significant–correlation between the PPVT-IV and the SC score) of each words of a sentence, but also on other mechanisms required to: (1) analyze the semantic content at the sentential level, and (2) retrieve an appropriate word to complete the sentence. These other mechanisms likely rely on explicit cognitive control processes and require a higher level of attention (e.g., Daltrozzo, Wioland, & Kotchoubey, 2012; January, Trueswell, & Tompson-Schill, 2009). Because of the likely involvement of these more attention-demanding-mechanisms, it is possible that the SC task is less sensitive to variations in SL ability but more sensitive to variations in aspects of cognitive control or executive function (January et al., 2009). Alternatively or as a complementary explanation, the lack of interaction between SL and SC while interactions were found between SL and PPVT-IV and GJ could stem from differences of reliability between these language measures (higher than 0.849 and 0.905 for the PPVT-IV and GJ scores, respectively, but only higher than 0.696 for the SC score).

The above-mentioned difference in results between our three measures of language is also consistent with our correlational results, which showed that these language measures were essentially unrelated to each other in our sample of adults (with a very low correlation coefficient between the GJ score and the two other language scores and a low correlation between the PPVT-IV and the SC score). This lack of correlation further supports the idea that these three measures are not assessing the performance of a

general language ability, but instead measure separate, somewhat independent skills.

The above-mentioned difference in results between our three measures of language is also consistent with our correlational results, which showed that these language measures were essentially unrelated to each other in our sample of adults. This lack of correlation further supports the idea that these three measures are not measuring the performance of a general language ability, but instead separate, somewhat independent skills.

Importantly, not only was the link between SL and grammatical ability not due to individual variations in general selective attention, but the link between SL and language performance overall also did not appear to be simply due to individual variations in perceptual processing. Indeed, visual processing performance was controlled for in our task because our measures of SL were not absolute measures but relative measures. Our tests of interactions with language performance were with the two levels of the Predictor condition (i.e. HP and LP). That is, we tested how the RT effect (RT differences between HP and LP conditions) and ERP effects (ERP differences between HP and LP conditions) varied with language performance. Therefore, assuming that visual processing performance was similar between the HP and the LP conditions, none of the reported interactions or correlation between language skills and the SL RT effect or SL ERP effect could be explained by a variation of visual processing performance across participants.

While general selective attention and perceptual processing can hardly explain the reported relationships between SL and language abilities, the fact that our measures of SL were not absolute measures but relative measures (between our HP and LP experimental conditions) cannot fully rule out that other cognitive abilities (e.g., intelligence or working memory) may underlie these relationships (Misyak & Christiansen, 2012). Although several studies indicate that SL is weakly correlated with these other cognitive abilities (e.g. McGeorge, Crawford, & Kelly, 1997; Reber, Walkenfeld, & Hernstadt, 1991), further studies will be required to fully address these issues.

Our discussion of ERP effects reported in this study focused mostly on those observed at the centro-parietal cortical site because our ERP operational definition of SL was at this scalp location (see Introduction), however ERP effects between HP and LP found also at other locations including the frontal area could have driven our reported interaction between the SL ERP effect, GJ, and attention. With a very similar paradigm, Jost et al. (2015) reported also large frontal ERP effects between various target predictability conditions. There are several possible (competing or complementary) interpretations for the source(s) of these frontal ERP effects, e.g. the modulations of a Contingent Negative Variation (Walter, Cooper, Aldridge, McCallum, & Winter, 1964) or a Slow Negative Wave (Lang & Kotchoubey, 2000) or the opposite scalp projection of a dipole source (oriented along the frontal to centro-parietal direction) of the SL ERP effect. These issues are discussed more extensively in Daltrozzo and Conway (2014).

While there is a clear positive trend between the SL (RT effect) and a language (GJ raw score) behavioral measure (Fig. 1C), the linear correlation is not highly significant $p = 0.03$. Indeed, the sample size of $N = 17$ limits the statistical power of the analysis. This issue limits also the statistical power of the tested interactions between the SL and the language behavioral measures. However, even under these conditions of statistical power, we were still able to find a marginally significant interaction between these behavioral measures.

4.3. General discussion

The fact that the observed relationship between SL and language ability was with a visual non-linguistic SL task suggests that, at least to some extent, the processes involved in SL are

domain-general (Kirkham et al., 2002). This stands in contrast to other research findings suggesting that SL may be mediated by a distributed set of modality-specific mechanisms, each tied to a particular type of input (Conway & Christiansen, 2005; Frost, Armstrong, Siegelman, & Christiansen, 2015). Conway and Pisoni (2008) suggested that there may be a combination of modality-specific and domain-general processes involved in SL, each relying on different neural substrates, which could help account for both the presence of stimulus-specificity in SL tasks but also findings revealing associations between non-linguistic SL and language measures (Shafra, Conway, Field, & Houston, 2012). Future research is needed to disentangle the role of domain-general and domain-specificity in SL and how each is used in the service of language processing.

It is important to note that, despite using advanced analyses, the present approach cannot determine directional causation, as we relied notably upon interactions between pre-existing language ability and SL effects rather than using an experimental manipulation to modify SL or language ability. Therefore, even though our results provide evidence of a link between SL and language ability, and even though evidence of this link extends beyond our study, being strengthened by previous behavioral (Conway et al., 2010) and neurophysiological studies (Christiansen et al., 2012; Tabullo et al., 2013), the correlational nature of this approach does not allow us to make a causal conclusion about the relationship between SL and language performance. As most of the previous SL research has emphasized, it seems likely that the association between SL and language is due to SL playing a causal role in the acquisition and processing of language. On the other hand, it is at least possible that the causal direction goes in the opposite direction, with better language abilities somehow causing better learning in SL tasks. Future research could fruitfully attempt to disentangle the causal nature of this relationship. This may be achieved for instance through training studies that attempt to increase SL ability and then test if language performance improves via cognitive transfer (Conway, Gremp, Walk, Bauernschmidt, & Pisoni, 2012; Smith, Conway, Bauernschmidt, & Pisoni, 2015).

Finally, as per Newman, Tremblay, Nichols, Neville, and Ullman (2012), we modeled ROIs with an unstructured random effect (see [Supplementary Materials and Methods](#)). Thus theoretically, our model allowed ROIs to freely correlate. However, it is in principle possible to model the spatial correlation within and across ROIs. However, this approach could in practice be problematic because it corresponds to more complex models, requiring more parameters to estimate, hence a higher risk of lack of estimation convergence.

In sum, the current findings suggest that language ability is based in part on neural mechanisms that are devoted to learning statistical contingencies in the environment, whether the input is linguistic or not. ERPs indicated that the relationship between SL and grammatical ability was independent of attention, while the association between SL and receptive vocabulary depended on attention. The relationships between SL mechanisms and these two language abilities in turn do not appear to be mediated by differences in perceptual processing. These results contribute to the debate about the nature of language, suggesting that it is both experience-dependent and based on domain-general pattern learning mechanisms. This finding in turn provides the theoretical basis for future exploration of the efficacy of cognitive training approaches to enhance SL and language, which could be a boon for individuals with a learning or communication disorder, such as dyslexia, specific language impairment, autism, or other language delays.

Acknowledgments

This research was funded by the National Institute of Health (grant R01DC012037 to CMC) and the Georgia State University's

Language and Literacy Initiative. We are grateful to A. Tremblay, A.J. Newman, S. Sims, J. Trapani, A. Ghali, G. Signiski, Lauterbach A., and Raley P. for their help with data acquisition and analysis.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.bandl.2016.12.005>.

References

- Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognit. Sci.*, 36, 286–304.
- Arciuli, J., & Torkildsen, J. V. K. (2012). Advancing our understanding of the link between statistical learning and language acquisition: The need for longitudinal data. *Frontiers in Psychology*, 3, 324.
- Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, 2, 170–176.
- Baldwin, K. B., & Kutas, M. (1997). An ERP analysis of implicit structured sequence learning. *Psychophysiology*, 34, 74–86.
- Carrow-Woolfolk, E. (1999). *Comprehensive assessment of spoken language (CASL)*. Circle Pines, MN: American Guidance Service.
- Casey, B. J., Thomas, K. M., Welsh, T. F., Badgaiyan, R. D., Eccard, C. H., Jennings, J. R., & Crone, E. A. (2000). Dissociation of response conflict, attentional selection, and expectancy with functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 8728–8733.
- Chomsky, N. (2011). Language and other cognitive systems. What is special about language? *Language Learning and Development*, 7, 263–278.
- Christiansen, M. H., & Chater, N. (2015). The language faculty that wasn't: A usage-based account of natural language recursion. *Frontiers in Psychology*, 6, 1182. <http://dx.doi.org/10.3389/fpsyg.2015.01182>.
- Christiansen, M. H., Conway, C. M., & Onnis, L. (2012). Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Language and Cognitive Processes*, 27, 231–256.
- Coch, D., & Holcomb, P. J. (2003). The N400 in beginning readers. *Developmental Psychobiology*, 43, 146–166.
- Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114, 356–371.
- Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 24.
- Conway, C. M., Gremp, M. A., Walk, A. D., Bauernschmidt, A., & Pisoni, D. B. (2012). Can we enhance domain-general learning abilities to improve language function? In P. Rebuschat & J. Williams (Eds.), *Statistical learning and language acquisition* (pp. 305–336). Boston: Walter de Gruyter.
- Conway, C. M., & Pisoni, D. B. (2008). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Annals of the New York Academy of Sciences*, 1145, 113–131.
- Conway, C. M., Pisoni, D. B., Anaya, E. M., Karpicke, J., & Henning, S. C. (2011). Implicit sequence learning in deaf children with cochlear implants. *Developmental Science*, 14, 69–82.
- Coulson, S., King, J., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13, 21–58.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- Daltrozzo, J., & Conway, C. (2014). Neurocognitive mechanisms of statistical-sequential learning: What do event-related potentials tell us? *Frontiers in Human Neuroscience*, 8, 437.
- Daltrozzo, J., Wioland, N., & Kotchoubey, B. (2012). The N400 and Late Positive Complex (LPC) effects reflect controlled rather than automatic mechanisms of sentence processing. *Brain Sciences*, 2, 267–297.
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134, 9–21.
- Du, W., & Kelly, S. W. (2013). Implicit sequence learning in dyslexia: A within-sequence comparison of first- and higher-order information. *Annals of Dyslexia*, 63(2), 154–170.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test-fourth edition*. Bloomington (MN): Pearson.
- Eimer, M., Goshke, T., Schlaghecken, F., & Stürmer, B. (1996). Explicit and implicit learning of event sequences: evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 970–987.
- Ellis, E., Robledo, M., & Deák, G. O. (2014). Contingency learning in infancy and its relation to later vocabulary. *Language Learning and Development*, 10, 36–50.
- Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, 37, 66–108.

- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics*, 16, 143–149.
- Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 52, 321–335.
- Fenske, M. J., & Eastwood, J. D. (2003). Modulation of focused attention by faces expressing emotion: Evidence from flanker tasks. *Emotion*, 3, 327.
- Ferdinand, N. K., Mecklinger, A., & Kray, J. (2008). Error and deviance processing in implicit and explicit sequence learning. *Journal of Cognitive Neuroscience*, 20, 629–642.
- Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12, 499–504.
- Foerle, K., Knowlton, B. J., & Poldrack, R. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences*, 103, 11778–11783.
- Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 529–534.
- Frisch, S., Kotz, S. A., von Cramon, D. Y., & Friederici, A. D. (2003). Why the P600 is not just a P300: The role of the basal ganglia. *Clinical Neurophysiology*, 114, 336–340.
- Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, 19, 117–125.
- Gallistel, C. R. (2011). Prelinguistic thought. *Language Learning and Development*, 7, 253–262.
- Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology*, 61, 191–218.
- Gogate, L. J., & Hollich, G. (2010). Invariance detection within an interactive system: A perceptual gateway to language development. *Psychological Review*, 117, 496–516.
- Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, 34, 660–676.
- Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP-measure of syntactic processing. *Language and Cognitive Processes*, 8, 439–483.
- Howard, J. H., Jr., Howard, D. V., Japikse, K. C., & Eden, G. F. (2006). Dyslexics are impaired on implicit higher-order sequence learning, but not on implicit spatial context learning. *Neuropsychologia*, 44, 1131–1144.
- Hübner, R., Steinhauser, M., & Lehle, C. (2010). A dual-stage two-phase model of selective attention. *Psychological Review*, 117, 759–784.
- January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience*, 21, 2434–2444.
- Jeste, S. S., Kirkham, N., Senturk, D., Hasenstab, K., Sugar, C., Kupelian, C., ... Johnson, S. P. (2015). Electrophysiological evidence of heterogeneity in visual statistical learning in young children with ASD. *Developmental Science*, 18, 90–105.
- Jost, E., Conway, C. M., Purdy, J. D., Walk, A. M., & Hendricks, M. A. (2015). Exploring the neurodevelopment of visual statistical learning using event-related brain potentials. *Brain Research*, 1597, 95–107. <http://dx.doi.org/10.1016/j.brainres.2014.10.017>.
- Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, 48, 171–184.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83, B35–B42.
- Krogh, L., Vlach, H. A., & Johnson, S. P. (2013). Statistical learning across development: Flexible yet constrained. *Frontiers in Psychology*, 3, 598.
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews. Neuroscience*, 5, 831–843.
- Lang, S., & Kotchoubey, B. (2000). Learning effects on event-related brain potentials. *NeuroReport*, 11, 3327–3331.
- Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology. General*, 133, 339.
- McCarthy, G., & Wood, C. C. (1985). Scalp distribution of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, 62, 203–208.
- McGeorge, P., Crawford, J. R., & Kelly, S. W. (1997). The relationships between psychometric intelligence and learning in an explicit and an implicit task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 239–245.
- Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, 62(1), 302–331.
- Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Online individual differences in statistical learning predict language processing. *Frontiers in Psychology*, 1, 31.
- Morgan-Short, K., Deng, Z., Brill-Schuetz, K. A., Faretta-Stutenberg, M., Wong, P. C. M., & Wong, F. C. K. (2015). A view of the neural representation of second language syntax through artificial language learning under implicit contexts of exposure. *Studies in Second Language Acquisition*, 37(2), 383–419. <http://dx.doi.org/10.1017/S0272263115000030>.
- Morgan-Short, K., Faretta-Stutenberg, M., Brill, K. A., Carpenter, H., & Wong, P. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition*, 17(1), 56–72.
- Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 3, 151–165.
- Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., & Ullman, M. T. (2012). The influence of language proficiency on lexical semantic processing in native and late learners of English. *Journal of Cognitive Neuroscience*, 24, 1205–1223.
- Newport, E. (2011). The modularity issue in language acquisition: A rapprochement? Comments on Gallistel and Chomsky. *Language Learning and Development*, 7, 279–286.
- Nicolson, R. I., & Fawcett, A. J. (2007). Procedural learning difficulties: Reuniting the developmental disorders? *Trends in Neurosciences*, 30, 135–141.
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, 19, 1–32.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113.
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10, 233–238.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118, 2128–2148.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855–863.
- Reber, P. J. (2013). The neural basis of implicit learning and memory: A review of neuropsychological and neuroimaging research. *Neuropsychologia*, 51(10), 2026–2042.
- Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 888–896.
- Saffran, J. R. (2003). Statistical learning learning: Mechanisms and constraints. *Current Directions in Psychological Science*, 12, 110–114.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1–66.
- Schwartz, T. J., Federmeier, K. D., van Petten, C., Salmon, D., & Kutas, M. (2003). Electrophysiological analysis of context effects in Alzheimer's disease. *Neuropsychologia*, 17, 187–201.
- Shafra, C. L., Conway, C. M., Field, S. L., & Houston, D. M. (2012). Visual sequence learning in infancy: Domain-general and domain-specific associations with language. *Infancy*, 17, 247–271.
- Shalev, L., & Tsai, Y. (2003). The wide attentional window a major deficit of children with attention difficulties. *Journal of Learning Disabilities*, 36, 517–527.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84, 127.
- Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, 81, 105–120.
- Sijsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74, 107–120.
- Smith, G. N. L., Conway, C. M., Bauernschmidt, A., & Pisoni, D. B. (2015). Can we improve structured sequence processing? Exploring the direct and indirect effects of computerized training using a mediational model. *PLoS ONE*, 10, e0127148. <http://dx.doi.org/10.1371/journal.pone.0127148>.
- Spencer, M., Kaschak, M. P., Jones, J. L., & Lonigan, C. J. (2014). Statistical learning is related to early literacy-related skills. *Reading and Writing*, 28, 467–490.
- Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, 38, 387–401.
- Tabullo, Á., Sevilla, Y., Segura, E., Zanotto, S., & Wainseboim, A. (2013). An ERP study of structural anomalies in native and semantic free artificial grammar: Evidence for shared processing mechanisms. *Brain Research*, 1527, 149–160.
- Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. V. (2010). Implicit perceptual anticipation triggered by statistical learning. *The Journal of Neuroscience*, 30, 11177–11187.
- Ullman, M. T. (2004). Contribution of memory circuits to language: The declarative/procedural model. *Cognition*, 92, 231–270.
- Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, 41, 399–433.
- Van Zuijen, T., Simons, V., Paavilainen, P., Näätänen, R., & Tervaniemi, M. (2006). Implicit, intuitive, and explicit knowledge of abstract regularities in a sound sequence: An event-related brain potential study. *Journal of Cognitive Neuroscience*, 18, 1292–1303.
- Verleger, R. (1988). Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3. *Behavioral and Brain Sciences*, 11, 343–437.
- Walter, W. G., Cooper, R., Aldridge, V. J., Mccallum, W. C., & Winter, A. L. (1964). Contingent Negative Variation: An electric sign of sensorimotor association and expectancy in the human brain. *Nature*, 203, 380–384.
- Woodhouse, B., & Jackson, P. H. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: II: A search procedure to locate the greatest lower bound. *Psychometrika*, 42, 579–591.
- Wöstmann, N. M., Aichert, D. S., Costa, A., Rubia, K., Möller, H. J., & Ettinger, U. (2013). Reliability and plasticity of response inhibition and interference control. *Brain and Cognition*, 81(1), 82–94.