

Distinct neural networks for detecting violations of adjacent versus nonadjacent sequential dependencies: An fMRI study

Christopher M. Conway^{a,b,c,*}, Leyla Eghbalzad^a, Joanne A. Deocampo^a, Gretchen N.L. Smith^a, Sabrina Na^{a,b}, Tricia Z. King^{a,b}

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

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

^c Center for Childhood Deafness, Language, and Learning, Boys Town National Research Hospital, Omaha, NE, USA¹

ARTICLE INFO

Keywords:

Sequential processing
Statistical learning
Artificial grammar learning
Nonadjacent dependencies
Anterior cingulate cortex
Angular gyrus

ABSTRACT

The ability to learn and process sequential dependencies is essential for language acquisition and other cognitive domains. Recent studies suggest that the learning of adjacent (e.g., “A-B”) versus nonadjacent (e.g., “A-X-B”) dependencies have different cognitive demands, but the neural correlates accompanying such processing are currently underspecified. We developed a sequential learning task in which sequences of printed nonsense syllables containing both adjacent and nonadjacent dependencies were presented. After incidentally learning these grammatical sequences, twenty-one healthy adults (age $M = 22.1$, 12 females) made familiarity judgments about novel grammatical sequences and ungrammatical sequences containing violations of the adjacent or nonadjacent structure while in a 3T MRI scanner. Violations of adjacent dependencies were associated with increased BOLD activation in both posterior (lateral occipital and angular gyrus) as well as frontal regions (e.g., medial frontal gyrus, inferior frontal gyrus). Initial results indicated no regions showing significant BOLD activations for the violations of nonadjacent dependencies. However, when using a less stringent cluster threshold, exploratory analyses revealed that violations of nonadjacent dependencies were associated with increased activation in subcallosal cortex, paracingulate cortex, and anterior cingulate cortex (ACC). Finally, when directly comparing the adjacent condition to the nonadjacent condition, we found significantly greater levels of activation for the right superior lateral occipital cortex (BA 19) for the adjacent relative to nonadjacent condition. In sum, the detection of violations of adjacent and nonadjacent dependencies appear to involve distinct neural networks, with perceptual brain regions mediating the processing of adjacent but not nonadjacent dependencies. These results are consistent with recent proposals that statistical-sequential learning is not a unified construct but depends on the interaction of multiple neurocognitive mechanisms acting together.

1. Introduction

Many domains with which humans and other complex organisms engage involve the encoding, representation, manipulation, and production of items in sequence. Some examples include music perception and production, language acquisition and processing, event segmentation, and skill learning. Across these examples, sequential learning and processing likely involves a plethora of cognitive and neural operations all working in concert, including low-level perceptual processes, attention and memory, and possibly even higher-level cognitive control or executive functions (Arciuli, 2017; Conway, in preparation; Conway,

Deocampo, Smith, & Eghbalzad, 2016; Daltrozzo & Conway, 2014; Frost, Armstrong, Siegelman, & Christiansen, 2015; Sawi & Rueckl, 2018; Thiessen, Kronstein, & Huftnagle, 2013).

One important distinction to be made with sequential processing is the difference between adjacent and nonadjacent dependencies (e.g., Creel, Newport, & Aslin, 2004; Gómez, 2002; Lany & Gómez, 2008; Lany, Gómez, & Gerken, 2007; Newport & Aslin, 2004; Onnis, Monaghan, Richmond, & Chater, 2005; Peña, Bonatti, Nespor, & Mehler, 2002; Perruchet, Tyler, Galland, & Peereman, 2004; Romberg & Saffran, 2013; Van den Bos, Christiansen, & Misyak, 2012; Vuong, Meyer, & Christiansen, 2016). An adjacent dependency is one in which

* Corresponding author at: Brain, Learning, and Language Lab, Center for Childhood Deafness, Language, and Learning, Boys Town National Research Hospital, Lied Learning and Technology Center, 555 N. 30th St., Omaha, NE 68131, USA.

E-mail address: christopher.conway@boystown.org (C.M. Conway).

¹ Present address.

<https://doi.org/10.1016/j.nlm.2020.107175>

Received 31 October 2018; Received in revised form 3 January 2020; Accepted 24 January 2020

Available online 01 February 2020

1074-7427/ © 2020 Elsevier Inc. All rights reserved.

two items in the sequence occur in immediate succession. In natural language, learning which adjacent syllables co-occur can provide a statistical cue for segmenting continuous speech into the relevant word units (Saffran, Aslin, & Newport, 1996) (e.g., the syllables “ro-” and “-bot” co-occur regularly in natural speech, forming the word “robot”). On the other hand, a nonadjacent dependency is one in which one or more items intervene between the dependency of interest (Gómez, 2002). In natural language, nonadjacent dependencies are found in certain aspects of grammar (e.g., in the phrases “the robot **is** walking”, and “the robot **is** jumping” the auxiliary “is” and the inflectional morpheme “ing” are dependent upon one another, independent of the intervening verb stem). Crucially, such nonadjacent or long-distance dependencies are thought to constitute an essential part of what makes human language different from other communication systems (e.g., Christiansen & Chater, 2015).

Perhaps not surprisingly, there is good reason to believe that the learning of adjacent versus nonadjacent dependencies involves distinct neurocognitive mechanisms. The learning of nonadjacent dependencies appears to be more difficult than learning adjacent dependencies, only occurring, for instance, when the nonadjacent stimuli are highlighted perceptually or when the nonadjacent regularities are much more predictive than the adjacent regularities (Creel et al., 2004; Gómez, 2002; Newport & Aslin, 2004). Recently, De Diego-Balaguer, Martínez-Alvarez, and Pons (2016) suggested that learning adjacent and nonadjacent regularities rely on distinct attentional mechanisms: exogenous attention for the former and endogenous attention for the latter. Endogenous attentional control is needed to learn a nonadjacent dependency because it requires top-down cognitive control mechanisms to inhibit attention to intervening items and direct attention to the nonadjacent dependencies of interest. On the other hand, learning adjacent dependencies does not require top-down selective attention. This dual-view of sequential pattern learning is consistent with the proposals put forth by for instance Daltrozzo and Conway (2014) and Conway, under review, who suggested that the learning of relatively simple sequential regularities is a bottom-up, automatic learning process relying on sensory cortical networks while the learning of more complex patterns such as long-distance dependencies requires top-down cognitive control mediated by anterior cortical regions.

The goal of this study was to examine the neural correlates associated with processing adjacent versus nonadjacent dependencies in order to test the idea that separate neurocognitive mechanisms are involved in the two types of learning. Specifically, we postulated the involvement of at least three neural networks that would be differentially activated during the processing of adjacent versus nonadjacent dependencies:

Perceptual processing networks: the learning of statistical regularities is known to rely in part on modality-specific perceptual processes (Conway & Christiansen, 2005; Conway, 2005; Frost et al., 2015), in which perceptual networks show increased facilitation with experience for input with similar structural regularities (Reber, Stark, & Squire, 1998). Indeed, the involvement of perceptual processing networks such as visual processing areas for visual pattern learning is commonly observed in neuroimaging studies of sequential learning (e.g., Forkstam, Hagoort, Fernandez, Ingvar, & Petersson, 2006; Lieberman, Chang, Chiao, Bookheimer, & Knowlton, 2004; Turk-Browne, Scholl, Chun, & Johnson, 2009; see Conway & Pisoni, 2008 for a review). It is likely that the learning of adjacent dependencies engages perceptual learning mechanisms that automatically encode environmental regularities (Conway, under review; Daltrozzo & Conway, 2014); however, these same mechanisms may be insufficient for learning nonadjacent dependencies (De Diego-Balaguer et al., 2016).

Prefrontal cortex: Importantly, it appears that processing of regularities over a temporal sequence cannot be achieved by low-level perceptual processing regions alone, but must also rely upon

prefrontal cortex (PFC) and other downstream cortical regions that have larger temporal receptive fields and thus can process information occurring over longer-time scales (Fuster, 2001; Hasson, Chen, & Honey, 2015). The PFC is part of a larger frontoparietal network that underlies working memory and attention related processes (Leung, Gore, & Goldman-Rakic, 2002; Rämä et al., 2001; Sarthain, Petsche, Rappelsberger, Shaw, & von Stein, 1998; Xang, Leung, & Johnson, 2003). The PFC has been frequently observed to be active in studies of sequence learning and implicit learning (Conway & Pisoni, 2008; Fletcher, Büchel, Josephs, Friston, & Dolan, 1999; Folia & Petersson, 2014; Skosnik et al., 2002). In addition, the PFC, in conjunction with connections to the basal ganglia, is thought to underlie procedural learning and memory, important for learning and processing sequential input (Ullman, 2004).

Cognitive control networks: As mentioned above, it has been proposed that learning nonadjacent dependencies requires additional processing over and above what is required to learn adjacent dependencies. It is likely that to process nonadjacent dependencies requires cognitive control to inhibit attention to the intervening items in a sequence and direct attention to the nonadjacent stimuli containing the dependency in question (De Diego-Balaguer et al., 2016). One set of brain regions involved in such cognitive inhibition and control abilities are the paracingulate and anterior cingulate cortex (ACC) (e.g., Carter, Mintun, & Cohen, 1995; Coderre, Filippi, Newhouse, & Dumas, 2008; Kemmotsu, Villalobos, Gaffrey, Courchesne, & Müller, 2005; Woodward, Ruff, & Ngan, 2006). These regions are also thought to be important for conflict monitoring and thus would be expected to be more active for more difficult cognitive processing operations that require cognitive control (Botvinick, Cohen, & Carter, 2004; Kerns et al., 2004). Thus, it is possible that cognitive control and conflict monitoring networks such as the ACC will be specifically associated with the learning and processing of nonadjacent dependencies.

In summary, we predict that we will observe fMRI activation in distinct sets of networks for the processing of adjacent versus nonadjacent dependencies in an incidental perceptual sequence learning task. We predict that, based on findings from other statistical-sequential learning studies (e.g., as reviewed by Conway & Pisoni, 2008), both perceptual (i.e., occipital) regions and the PFC will show greater activation for violations of sequential dependencies, relative to sequences without violations. However, we predict an important dissociation: sensory regions (posterior / occipital cortex) will primarily show activation for processing adjacent dependencies whereas cognitive control regions such as the ACC will show greater activation for violations of nonadjacent dependencies.

2. Materials and methods

This study reports data from participants who took part in an fMRI experiment that was part of a larger two-session study involving multiple measures and additional participants who did not participate in the fMRI portion of the study. Here we focus mainly on measures collected during session 2 (which included the fMRI task). We also describe aspects of session 1 that are relevant to the current paper.

We used an artificial grammar learning paradigm incorporating written verbal nonsense syllable sequences that spanned across two sessions and involved both behavioral and fMRI measurements. In the first session of the learning task, participants viewed sequences of nonsense syllables and then for each sequence, were required to replicate the sequence using buttons that corresponded to each written syllable (similar to methodology used for instance by Conway, Bauernschmidt, Huang, & Pisoni, 2010; Karpicke & Pisoni, 2004). Unbeknownst to the participants, the sequences contained both adjacent and nonadjacent dependencies. After exposure to a subset of sequences

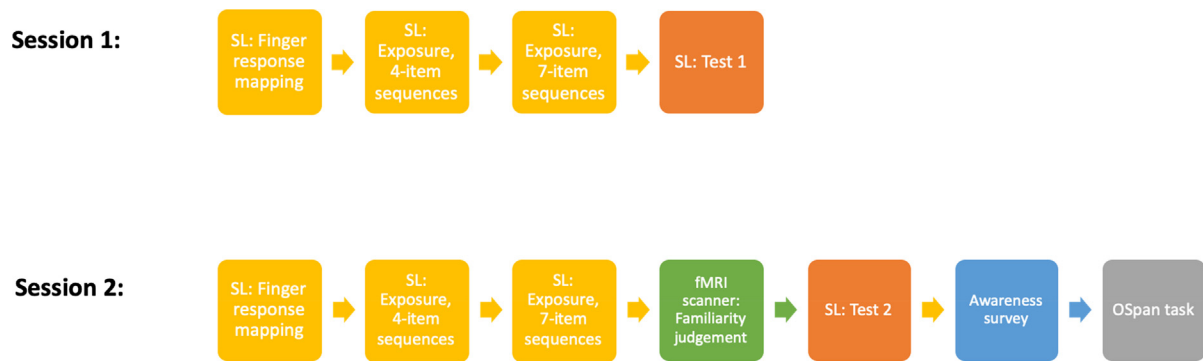


Fig. 1. Graphical depiction of the order of experimental procedures for both sessions.

containing the adjacent and nonadjacent dependencies, a test phase was given in which participants continued to do the sequence reproduction task. However, the test sequences were all novel and either were consistent with the dependencies to which they had been familiarized or contained violations of either the adjacent or nonadjacent dependencies. Learning was assessed behaviorally by examining sequence reproduction accuracy for grammatical versus ungrammatical sequences.

At the start of session 2, participants were exposed a second time to the adjacent and nonadjacent dependencies using the same sequence replication procedure described above. Next, participants were given a familiarity task inside the fMRI scanner. In the familiarity task, participants were told that some sequences were “new” and others were “old” and they were required to decide whether each sequence was familiar. In fact, all sequences were test sequences that they had seen previously during session 1, but some were grammatical while others contained violations of the adjacent or nonadjacent dependencies. In this way, participants were assessed on their processing of grammatical and ungrammatical sequences that either were consistent with or inconsistent with the dependencies without explicitly instructing them about the presence of rules and without asking them to make explicit judgments about the dependencies. Following the in-scanner task, participants were given the replication sequence task one more time using the grammatical and ungrammatical test sequences in order to obtain a final measure of learning.

It is important to note that in our paradigm, participants’ neural activity was measured while they viewed sequences that were either consistent or inconsistent with the adjacent and nonadjacent dependencies to which they had previously been exposed. Thus, the results of this study do not necessarily indicate which areas are involved in *learning* per se, but they can tell us which brain regions are involved in the *detection of violations* of the regularities after learning has occurred.

2.1. Participants

Informed consent was obtained from all participants prior to their participation. In session 1, participants who passed an initial pre-screening were either given 3 course credits or were paid \$25 for participation. This pre-screening ensured that all participants were monolingual English speakers and had no known cognitive, neurological, sensory, or motor impairment. Following participation in session 1, twenty-five participants were identified who met the following additional criteria: (1) demonstrated learning (more grammatical than ungrammatical test items reproduced correctly) on the sequential learning task administered in session 1 (see session 1 procedures below); (2) met standard health and safety criteria for receiving an MRI scan; (3) agreed to participate in session 2, which included the fMRI scan; and (4) received an additional screening to ensure that none were using medications or had a history of drug abuse, head trauma, or

neurological/psychiatric illness. These participants took part in session 2 and were compensated \$50 for their time and given a disk with images of their brain. Out of these 25 participants, 4 were excluded from further analysis for the following reasons: 1 was excluded due to significant head motion during imaging, 1 participant fell asleep in the scanner, brain images from 1 participant showed evidence of brain lesions, and 1 participant had tattooed eyeliner which did not meet the standard safety criteria for MRI scanning.

Thus, data are reported from 21 healthy adults (age $M = 22.1$, 12 females) who completed both sessions 1 and 2. All but one of the participants were right-handed. All were monolingual English speakers and had normal or corrected-to-normal vision.

2.2. Session 1

In session 1, a larger set of participants was recruited and completed a number of experimental measures and assessments, only some of which are reported in the current manuscript. The full set of measurements and data are reported in [Deocampo, King, and Conway \(2019\)](#). Based on performance on the sequential learning task in session 1, a subset of participants was then invited to participate in the fMRI part of the study in session 2. A graphical depiction of the order of experimental procedures for each session are outlined in [Fig. 1](#).

2.2.1. Materials

We designed a sequential learning task that consisted of printed nonsense syllables that were presented sequentially on the screen (see [Fig. 2A](#)). Four orthographic nonsense syllables were used to construct the sequences: ka, po, lu, di. Syllables appeared individually in the center of the screen to form the sequences.

An artificial grammar was created that dictated both adjacent and nonadjacent dependencies within each sequence. The grammar consisted of four pairs of nonadjacent dependencies and four pairs of adjacent dependencies (see [Fig. 2B](#)). Unlike most previous studies, adjacent and nonadjacent dependencies were composed of the same items (but following different rules) so as not to artificially highlight either type of dependency or to make the different types of dependencies more salient by using different types of stimuli for each. This also served to make dependencies more similar to real-world language and other types of dependencies in which adjacent and nonadjacent dependencies are often made up of the same items. To form sequences, the adjacent pairs were used as the intervening element (x) between the nonadjacent pair items (see [Fig. 2C](#) and [2D](#)). Each sequence was either composed of a single non-adjacent dependency pair and single adjacent dependency pair (4 total items) or two of each type of pair with the second item of the first nonadjacent pair becoming the first item of the second nonadjacent pair (7 total items). All dependencies were deterministic. Thus, in all sequences, the first item of an adjacent or nonadjacent pair regardless of position in the sequence was always 100% predictive of the second item in the pair. These made up the grammatical sequences, 16

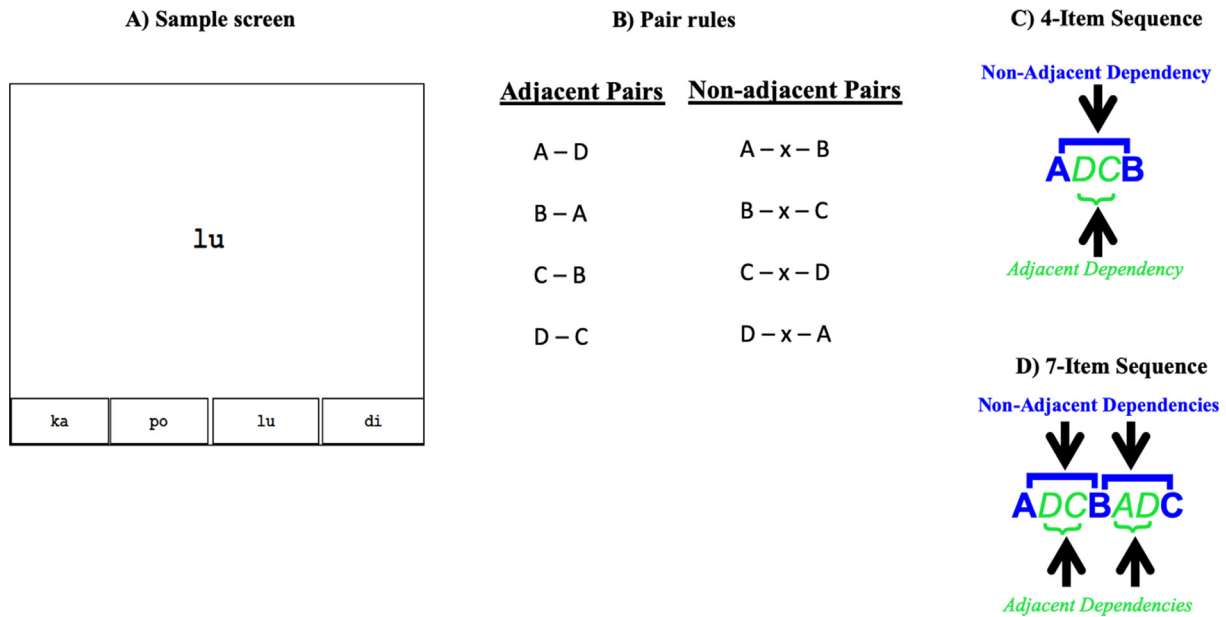


Fig. 2. Sample screen from the sequential learning task (A). List of adjacent and nonadjacent pair rules (B). The intervening element in the nonadjacent pairs (x) consisted of any of the adjacent pairs and therefore consisted of two items. Depiction of example 4-item (C) and 7-item (D) grammatical sequences with adjacent and non-adjacent pairs marked (1 of each in the 4-item sequence, 2 of each in the 7-item sequence, figure adapted from [Deocampo et al., 2019](#)). Color is for illustrative purposes only. Each element of the grammar (A, B, etc.) was randomly mapped to a syllable (“lu”, “ka”, etc.).

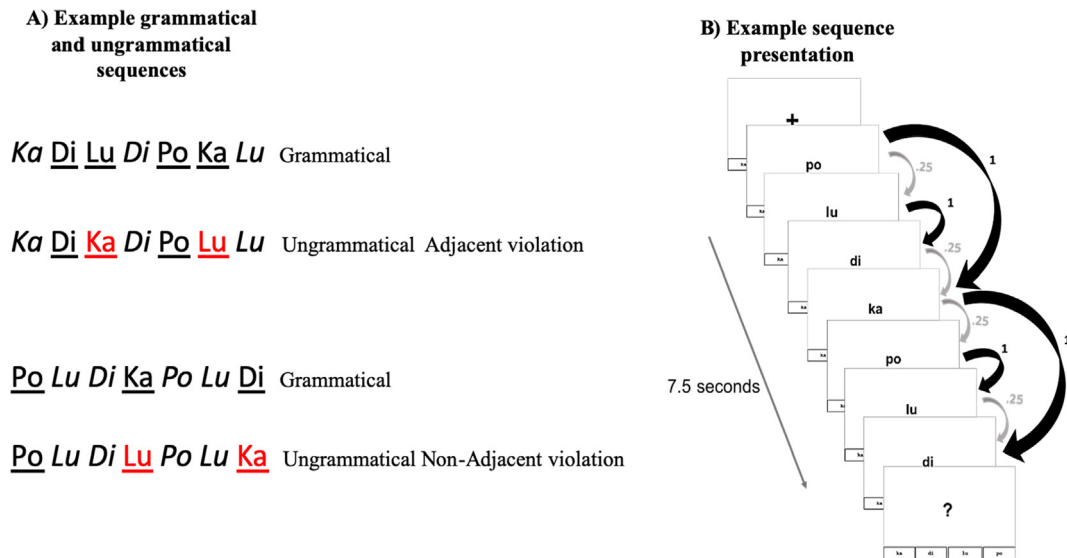


Fig. 3. Examples of two 7-item grammatical sequences and two 7-item ungrammatical sequences after mapping to syllables (A). One ungrammatical sequence contains adjacent violations and the other contains nonadjacent violations. Bold, italics, and color are for illustrative purposes only. Example grammatical sequence presentation trial during the in-scanner familiarity task (B). Each trial was preceded by a fixation cross. Transition probabilities (TPs) of each adjacent and non-adjacent pair are presented next to the arrows. Black arrows represent nonadjacent and adjacent dependency pairs, respectively. Grey arrows represent TPs of 0.25 between adjacent and nonadjacent items.

4-item sequences and 64 7-item sequences (32 for exposure and 32 for test). The 32 7-item grammatical test sequences were also used to create 32 7-item ungrammatical sequences. Half of the grammatical sequences had violations introduced into both adjacent pairs by replacing one member of each pair with an incorrect item. This created 16 “adjacent ungrammatical” sequences. The other half of the grammatical 7-item sequences were given similar violations in both nonadjacent pairs to make 16 “nonadjacent ungrammatical” sequences (see [Fig. 3A](#)). Although all of the grammatical sequences were made with the same adjacent and nonadjacent pairs and thus did not differ in adjacency, we will call those grammatical sequences used to make the adjacent ungrammatical sequences “adjacent grammatical” and those used to

create the nonadjacent ungrammatical sequences “nonadjacent grammatical”. In the analyses to assess learning-related effects, we used adjacent grammatical sequences for comparison with adjacent ungrammatical sequences and nonadjacent grammatical sequences for comparison with nonadjacent ungrammatical sequences since each grammatical sequence only differs from its ungrammatical pair by the two violations. Note that although all participants were exposed to the same grammatical and ungrammatical items, the mapping between the elements of the grammar (i.e., “A”, “B”, etc.) to the syllable tokens (i.e., “po”, “ka”, etc.) was randomly determined for each participant and remained stable for each participant for all phases of the experiment. A sample sequence presentation is shown in [Fig. 3B](#).

2.2.2. Procedure

All participants followed the same procedure and completed the same tasks and assessments in the same order. Although presentation of sequences within the sequential learning task was randomized, task order was not counterbalanced. The decision not to counterbalance was made because it was intended that a subset of the participants would subsequently participate in the fMRI portion of the study for which participation in this session would serve as exposure to the sequences used in the fMRI session. We wished to remove as much variability in brain activation as possible due to such extraneous variables as order of task exposure. All participants completed a set of experimental and standardized assessments first (reported in [Deocampo et al., 2019](#)) and the sequential learning task at the end.

After informed consent was completed, participants were led to a private, sound-attenuated room and seated in front of a Dell Optiplex 990 personal computer running Windows 7 Enterprise with a standard keyboard and a 17 in. ELO touchscreen monitor. An introduction to the sequential learning task was given verbally by the experimenter and the task was presented with Eprime 2.0 psychology experiment presentation software. The entire task took approximately 20 min.

2.2.2.1. Finger response mapping. The purpose of the mapping portion of the task was for participants to learn which keyboard buttons (numbers 1 through 4) were associated with which syllables on the screen when responding. The syllables ka, po, lu, di were printed at the bottom of the screen to remind the participant that the 1 key went with ka, 2 with po, 3 with lu, and 4 with di (see [Fig. 1A](#)). Participants were presented 16 randomly ordered trials, with a single syllable presented on each trial, 4 trials for each of the 4 syllables. On each trial, the participant was instructed to press the corresponding key each time a syllable was presented. All participants reached 80% or higher accuracy on the mapping task.

2.2.2.2. Exposure: 4-item sequences. Participants were told that they would now be replicating sequences, first a set of 4-item sequences and then a set of 7-item sequences. Ordering sequences in this way was meant to help participants learn the adjacent and nonadjacent pairs by presenting shorter and simpler sequences before longer sequences. Such a “starting small” strategy has been shown to increase learning effectiveness in artificial grammar learning paradigms (e.g., [Poletiek et al., 2018](#)). Participants were reminded to use only their right hand to respond and keep the keyboard in a comfortable position. Participants were told that they would see sequences of four syllables and that after a sequence was presented, they were to reproduce the sequence in the same order. Participants completed 32 4-item trials composed of 2 presentations each of 16 4-item sequences in random order. All sequences followed the grammar by always containing a single “legal” adjacent pair and a single “legal” nonadjacent pair in the prescribed format ([Fig. 2C](#)). Participants were not told that there was an underlying grammar or that there were adjacent and nonadjacent dependencies embedded in the sequences.

Throughout sequence presentation and response, a representation of the location of the response button (keyboard 1 through 4) for each syllable remained at the bottom of the screen to simplify the participant’s task of mapping the syllables to response buttons while reproducing the sequences (see [Fig. 1A](#)). Individual syllables were presented for 400 ms continuously with an ISI of 200 ms. When the sequence presentation in a trial was complete, a blank white screen was shown for 200 ms. The cue for the participant to respond was the re-appearance of the syllable mapping at the bottom of the screen. The participant used the 1–4 keys on a standard keyboard to reproduce the sequence. As the participant pressed the buttons, the corresponding syllables showed on the screen. If the participant pressed at least one key in response, the program waited for 3 s after each press to allow for further responses. When there was no further response within 3 s of the last key press, the next sequence was presented. If the participant did

not begin to respond within 13 s, the next sequence was presented.

2.2.2.3. Exposure: 7-item sequences. Upon completion of 4-item set of sequences, participants took a self-paced break during which they read instructions telling them that in the next section they would see 7-item sequences and that they should replicate them in the same way as in the previous section. When the participant pressed a key to start the next session, he or she was presented with 64 trials of grammatical 7-item sequences, two each of 32 7-item sequences that consisted of two “legal” adjacent pairs and two “legal” non-adjacent pairs (as per [Fig. 2D](#)). Sequences were presented in random order. As before, participants were not instructed about the nature of the embedded dependencies. The timing characteristics of the sequences were identical to that described in [Section 2.2.2.2](#) for the 4-item sequences.

2.2.2.4. Test. When the 7-item exposure phase was complete, participants took another self-paced break during which they read instructions that were exactly the same as those for the exposure sections. For this test section, they were randomly presented with 128 trials (two each of the 64 test sequences). All of the sequences were new to the participants, with half of them following the same grammar as presented during exposure and the other half containing grammar violations as described previously ([Fig. 3A](#)). Half of the ungrammatical sequences were adjacent ungrammatical, meaning that they had an incorrect item within each adjacent pair. The other half were nonadjacent ungrammatical with incorrect items in both nonadjacent pairs. Participants were not given any indication that this section served as a test or that some sequences contained sequential violations. In all other ways including time of sequences this portion of the experiment was identical to the 7-item exposure phase.

2.3. Session 2

Session 2 was conducted at the Georgia State / Georgia Institute of Technology Joint Center for Advanced Brain Imaging.

2.3.1. Procedure

After completing informed consent and a second fMRI screening for session 2, participants were taken to a private room and seated in front of a 12.5 in. Lenovo Thinkpad laptop running Eprime 2.0. See [Fig. 1](#) for a depiction of session 2 procedures. Participants first completed the sequential learning task described in session 1: they were given the same instructions and completed the same finger response mapping and exposure phases described in [Section 2.2.2](#). However, they did not complete the test phase of the sequential learning task at this point in the study. Sequences in exposure phases were again presented in random order but were the same sequences from session 1. Participants then participated in the fMRI portion of the session (described below). After their fMRI scan, participants completed the test phase of the sequential learning task. They were then interviewed about their awareness of patterns in the sequences (see [Table 1](#) for list of questions).

Finally, participants completed the Shortened Operation Span (called the OSpan) task ([Foster, Harrison, & Redick, 2015](#)), which is an assessment of verbal working memory. Participants viewed sequences of 3–7 letters to be remembered and correctly sequenced later. After each letter was presented and before the next one was presented, the participant was required to complete a math problem while holding the current sequence in mind.

2.3.2. Functional MRI experimental task

For the in-scanner task, participants were instructed to judge whether or not each sequence presented to them was familiar by pressing a button on the keyboard that corresponded to ‘yes’ or ‘no’. The sequences that were used for this in-scanner judgement task were the same test sequences used in session 1, in which half of the sequences

Table 1
Post-scan Pattern Awareness Interview Questions.

Questions	Response Options
<u>Pattern Awareness:</u> Did you notice a pattern in the sequences presented? If so, try to describe it.	1. There was absolutely no pattern at all. 2. The items occurred somewhat randomly. 3. There may have been a pattern. 4. There was a pattern at certain times. 5. There was definitely a pattern.
<u>Sequence Reproduction Confidence:</u> On a scale of 1 to 10, how confident were you in your sequence imitation?	1 through 10 ratings (1 = least confident)
<u>Awareness of Violations:</u> Did it ever seem like there were mistakes in the sequences? Explain.	1. No 2. Yes

were grammatical and the other half consisted of violations to the adjacent or nonadjacent dependencies. All of the sequences had been seen by participants previously (in the session 1 test phase), and thus, strictly speaking, were equally familiar. However, it was expected that sequences following the learned dependencies of the grammar would seem more familiar than those violating the grammar (e.g., Wan, Dienes, & Fu, 2008). However, behavioral responses were not of primary interest, rather we wanted a task that would maximize their attention to the stimuli while minimizing motor movement responses.

The familiarity judgment task consisted of a block design with a total of four runs. Each run was preceded by a screen containing written instructions for the task followed by a fixation crosshair for 10 s, after which occurred 6 blocks of trials (Fig. 3). Subsequent blocks were separated by a fixation crosshair lasting 1.5 s. Each run contained one block each of grammatical adjacent trials, grammatical nonadjacent trials, ungrammatical adjacent trials, ungrammatical nonadjacent trials, and control trials (consisting of a single syllable repeated 7 times). Each block consisted of 4 trials (separated by 500-ms fixation crosses) and each trial was made up of a 7-item test sequence (described previously), with individual syllables presented on the screen one at a time for 400 ms each with an inter-stimulus interval of 200 ms. Runs and trials followed a set order shown in Fig. 4. Each run lasted approximately 3 min and 47 s, yielding a total task time of 15 min and 8 s for all four runs. Eprime (2.0.8) was used for stimuli presentation and collection of behavioral responses. A depiction of a grammatical test sequence trial is shown in Fig. 3B, with the adjacent and nonadjacent transitional probabilities indicated.

2.3.3. Functional MRI acquisition and analyses

Functional MRI data were acquired using a 3T Siemens Trio MRI scanner using a 12-channel head coil. Cushions and forehead straps were placed around the participant's heads to minimize head movement. Behavioral data were collected by placing a scanner-safe keyboard with 5 keys right next to the participants' hand. Task dependent

image series were collected using a gradient-recalled T_2^* -weighted echo-planar-imaging sequence (EPI) sequence based on blood oxygenated level-dependent (BOLD) contrast. The primary imaging parameters for the BOLD contrast included: 37 slices, 3-mm slice thickness and 0-mm slice gap, repetition time (TR) = 2000 ms, echo time (TE) = 30 ms, flip angle = 90° , nominal resolution = $3 \times 3 \times 3 \text{ mm}^3$. For anatomical registration, high-resolution T_1 -weighted structural images were acquired with a multi-echo magnetization prepared rapid gradient echo (ME-MPRAGE) sequence using the following parameters: 176 sagittal slices, field of view = $256 \text{ mm} \times 256 \text{ mm}$, 2 mm^3 voxel size, TR = 2530 ms, TE = 1.74 ms, 3.6 ms, 5.46 ms, 7.32 ms, inversion time TI = 1260 ms, flip angle = 7 degrees.

fMRI data analysis was conducted using FEAT (fMRI Expert Analysis Tool) Version 6.01, which is part of FSL (fMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). Data from all four runs was preprocessed in the following sequence: motion correction with the MCFLIRT tool of FSL (Jenkinson, Bannister, Brady, & Smith, 2002); brain extraction using the BET tool of FSL (Smith, 2002), slice timing correction; spatial smoothing with Gaussian kernel (FWHM = 5 mm); high pass temporal filtering. The preprocessed data was then registered to its corresponding high resolution T_1 ME-MPRAGE images and subsequently to the standard brain template (MNI152 T1 2 mm) using FNIRT nonlinear registration (Andersson et al., 2007a, 2007b).

General linear modeling was performed using FILM with local autocorrelation correction on individual data. A total of 5 regressors were entered into the GLM setup: control, adjacent grammatical (AG), adjacent ungrammatical (AU), nonadjacent grammatical (NG), and nonadjacent ungrammatical (NU). We created 4 contrasts to compare BOLD signal during violations in adjacent and nonadjacent conditions. The [AU – AG] contrast represents the difference in activation associated with judgement of familiarity for sequences containing violations of the adjacent dependencies relative to sequences containing grammatical adjacent dependencies. The [NU – NG] contrast represents the difference in activation associated with judgement of familiarity for

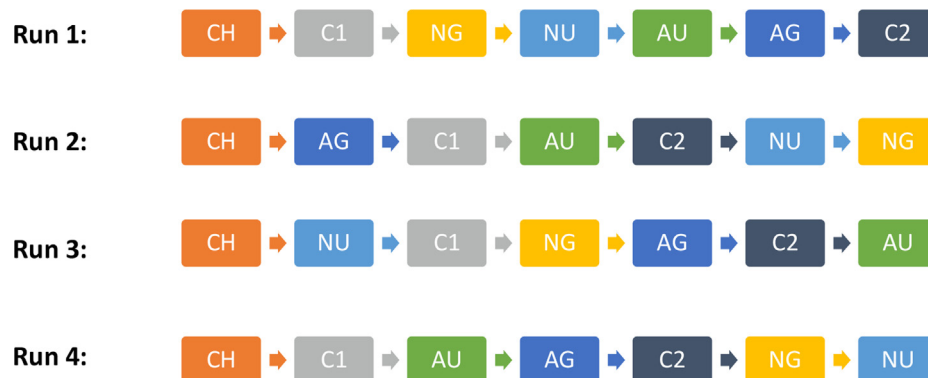


Fig. 4. Order of blocks in each run of the in-scanner task. CH = crosshair baseline block, C1 = 1st control block, C2 = 2nd control block, NG = grammatical (nonadjacent) block, AG = grammatical (adjacent) block, NU = ungrammatical nonadjacent block, AU = ungrammatical adjacent block.

sequences containing violations of the nonadjacent dependencies relative to sequences containing grammatical nonadjacent dependencies. In addition to the previous 2 contrasts, we created the [(AU-AG) – (NU-NG)] and [(NU-NG) – (AU-AG)] contrasts to compare the difference in activation in detecting violations between adjacent and nonadjacent conditions.

Higher-level within-subject statistical analyses for all contrasts were carried out using a fixed effects model by forcing the random effects variance to zero in FLAME (fMRIB's Local Analysis of Mixed Effects; Beckmann, Jenkinson, & Smith, 2003; Woolrich, 2008).

Higher-level between-subject statistical analyses for all contrasts were carried out using FLAME (fMRIB's Local Analysis of Mixed Effects) stage 1 with automatic outlier detection (Woolrich, 2008). Z statistic images were thresholded non-parametrically using clusters determined by $Z > 3.1$ and a corrected cluster significance threshold of $p = 0.05$ for all contrasts (Worsley, 2001). Importantly, due to the low statistical power resulting from our small sample size, it may be difficult to detect significant differences in activation in contrasts of interest; therefore, as an exploratory strategy, we conducted higher-level between-subject analyses with lower cluster thresholds of $Z > 2.3$ and $Z > 1.8$ using corrected cluster significance threshold of $p = 0.05$ for any contrasts that did not obtain significant differences at the $Z > 3.1$ threshold. Finally, we created binary masks based on significant clusters of interest for each group-level contrast. These masks were used to extract percent of the BOLD signal change (PSC) for each subject using the 'featquery' tool (Part of FSL). We used these subject-wise PSC values to correlate behavioral measures using nonparametric correlation analyses of Kendall's τ (continuous variables) and Point-biserial correlation (dichotomous variables).

3. Results

3.1. Behavioral results

3.1.1. Session 1 sequential learning test performance

As in previous studies using similar artificial grammar learning designs (e.g., Conway et al., 2010; Karpicke & Pisoni, 2004), we operationalized behavioral learning as higher reproduction performance for grammatical compared to ungrammatical sequences. To test for behavioral learning, paired samples t -tests were performed to compare total number of items reproduced correctly (correct item in correct serial position) out of 224 (7 items per sequence times 32 sequences) for grammatical versus ungrammatical sequences, separately for sequences with adjacent and nonadjacent violations. Results indicated significant learning during session 1 for both adjacent ($t(20) = 4.35$, $p < .001$, $d = 0.949$; AG $M = 158.29$ total correct, $SD = 39.00$; AU $M = 141.71$ total correct, $SD = 35.27$) and nonadjacent dependencies ($t(20) = 3.16$, $p = .005$, $d = 0.690$; NG $M = 155.14$ total correct, $SD = 38.79$; NU $M = 143.81$ total correct, $SD = 35.42$).

An alternative way to assess learning is to compare total number of correct sequences for grammatical versus ungrammatical sequences, rather than scoring total number of items correct within all sequences. We used paired samples t -tests to compare number of error-free reproduced sequences out of 32 per condition as the dependent variable. Using this alternative method, results again indicated learning – i.e., greater levels of performance for grammatical versus ungrammatical sequences – for both adjacent, $t(20) = 5.86$, $p < .001$, $d = 1.28$; AG $M = 10.48$ sequences correct, $SD = 7.43$; AU $M = 5.67$ sequences correct, $SD = 5.00$, and nonadjacent dependencies, $t(20) = 4.68$, $p < .001$, $d = 1.03$; NG $M = 10.38$ sequences correct, $SD = 8.03$; NU $M = 6.95$ sequences correct, $SD = 6.05$.

Because both outcome measures, total number of individual items correct across sequences and total number of error-free reproduced sequences resulted in similar findings, total number of items correct was used for further behavioral analysis because the greater number of items allowed for more variability across participants.

3.1.1.1. Adjacency. To determine the effect of adjacency on learning, we constructed percent change scores to represent learning for adjacent dependencies (as demonstrated by percent change in performance between AG and AU sequences) and nonadjacent dependencies (as demonstrated by percent change in performance between NG and NU sequences). We compared them using a paired samples t -test. Percent change for each condition (adjacent sequences or nonadjacent sequences) was calculated as

$$\frac{(G - U)}{G} * 100$$

in which G is the total number of grammatical items correct and U is the total number of ungrammatical items correct. This represents the extent to which performance was facilitated for grammatical sequences compared to ungrammatical sequences, relative to baseline performance on grammatical sequences. The paired samples t -test indicated that both types of adjacency were learned to the same extent ($t(20) = 0.856$, $p = .402$; A $M = 9.75\%$ change, $SD = 12.96$; N $M = 6.42\%$ change, $SD = 11.56$).

3.1.2. Session 2 in-scanner familiarity test performance

To determine whether participants endorsed a sequence as familiar differentially based on grammaticality and adjacency, we conducted a 2 (grammaticality: grammatical or ungrammatical) \times 2 (adjacency: adjacent or nonadjacent) repeated measures analysis of variance (ANOVA) on percent of trials endorsed. Results indicated a trend toward a significant main effect of grammaticality ($F(1, 20) = 2.24$, $p = .15$, $\eta_p^2 = 0.10$) in which participants endorsed familiarity for grammatical sequences at a higher rate ($M = 61.16\%$ endorsed, $SD = 15.69$) than ungrammatical sequences ($M = 56.99\%$, $SD = 16.56$). There was no significant main effect of adjacency ($F(1, 20) = 0.961$, $p = .339$) and no significant interaction ($F(1, 20) = 0.162$, $p = .692$). These results suggest that participants showed some indication of discriminating behaviorally between sequences that did and did not follow the learned grammar (though this was not a significant effect), and there was no difference in familiarity ratings for adjacent and nonadjacent dependencies.

3.1.3. Session 2 sequential learning post-fMRI scan test performance

As with session 1, a paired samples t -test was conducted comparing grammatical and ungrammatical total correct separately for adjacent and nonadjacent dependencies for session 2 to determine whether there was still evidence of learning. Participants continued to show significant evidence of learning for adjacent dependencies ($t(18^2) = 4.50$, $p < .001$, $d = 1.04$; AG $M = 162.42$ total correct, $SD = 34.84$; AU $M = 144.26$ total correct, $SD = 34.50$), but not for nonadjacent dependencies ($t(18) = -0.068$, $p = .947$; NG $M = 156.47$ total correct, $SD = 37.87$; NU $M = 156.68$, $SD = 34.43$). The lack of learning displayed for nonadjacent dependencies was due to an increase in performance on ungrammatical sequences. Increased performance on ungrammatical sequences suggests that participants were no longer using the underlying grammatical structure to help recall sequences and thus, they no longer preferentially recalled sequences conforming to that structure. An alternative way to think of this pattern of results is that attempting to recall sequences with a violation of the grammar carries with it a cost to processing and interferes with recall. This is only the case if adequate learning of the grammatical sequences has occurred. Therefore, these results suggest that participants' learning levels for nonadjacent grammatical structure is lower in session 2 (see Deocampo et al., 2019, for additional discussion on possible causes for this).

² Two participants did not have behavioral post-scan SL data due to data loss during a computer crash, leaving 19 participants for analyses with post-scan SL data.

3.1.3.1. Adjacency. For session 2, a paired samples *t*-test comparing percent change learning scores indicated that participants showed significantly more learning ($t(18^2) = 3.14$, $p = .006$, $d = 1.09$) of adjacent ($M = 10.97\%$ change, $SD = 11.30$) compared to non-adjacent dependencies ($M = -1.51\%$ change, $SD = 11.32$).

3.1.4. Session 2 OSpan performance

We scored the OSpan using the partial OSpan scores (Foster et al., 2015). The mean³ OSpan partial score was 18.22 with a standard deviation of 5.89.

3.1.5. Session 2 pattern awareness interview results

Participants reported moderate levels of pattern awareness with a mean verbal pattern level score of 3.19 ($SD = 0.98$, range = 1–5, Table 1, question 1) out of 5. Thus, participants' level of awareness on average fell somewhere between "there may have been a pattern" and "there was a pattern at certain times". Participants were somewhat confident of their reproduction of sequences with a mean rating of 5.41 ($SD = 2.12$) out of 10 (Table 1, question 2). Finally, only 19% (4 participants) endorsed that sometimes there were mistakes in the sequences.

3.2. Functional MRI results

3.2.1. Cluster threshold of $Z > 3.1$

Whole brain analysis for the adjacent contrast [AU-AG] revealed significantly greater activation for ungrammatical sequences (violations) relative to grammatical sequences in right frontal pole, right medial frontal gyrus (MFG; BA 44 & 45), right superior lateral occipital cortex, and right angular gyrus (BA 19 & 39; Table 2A) as shown in Fig. 5a and b.

Whole brain analysis comparing the difference in activation between adjacent and nonadjacent conditions [(AU-AG) - (NU-NG)] revealed significantly greater activation for the adjacent condition compared to the nonadjacent condition in the superior division of right lateral occipital cortex (BA 19; Table 2B; Fig. 5c). On the other hand, whole brain analyses with contrasts [NU-NG] and [(NU-NG) - (AU-AG)] did not show any significant differences in activation.

3.2.2. Cluster threshold of $Z > 1.8$

As an exploratory strategy, for the contrasts [NU-NG] and [(NU-NG) - (AU-AG)] which did not show any significant results with cluster threshold of $Z > 3.1$, we conducted higher-level analyses with lower cluster thresholds of $Z > 2.3$ and $Z > 1.8$ with corrected cluster significance threshold of $p = 0.05$. Using the cluster threshold of $Z > 2.3$ did not produce any significant results for either of the contrasts; however, analyses with a less conservative cluster threshold of $Z > 1.8$ for [NU-NG] revealed significantly greater activation in left subcallosal cortex, left paracingulate gyrus, and left ACC (BA 32; Table 2C) for ungrammatical sequences compared to grammatical sequences in the nonadjacent condition (Fig. 5d).

The cluster threshold of $Z > 1.8$ for [(NU-NG) - (AU-AG)] contrast revealed significantly greater difference in activation during detection of violations in the nonadjacent condition compared to the adjacent condition in the following regions: right frontal pole, right paracingulate gyrus, right subcallosal cortex, and right frontal medial cortex (BA 9 & 10; Table 2D; Fig. 5e). As recommended by a reviewer, we explored the individual contribution of each experimental condition (NU, NG, AU, & AG) in the activation results of [(NU-NG) - (AU-AG)] contrast by calculating mean beta values for each condition using the 'featquery' tool in FSL. The results revealed that in the nonadjacent

condition, ungrammatical sequences showed greater activation compared to grammatical sequences (NU $\beta = -20.9 > NG \beta = -21.4$). In the adjacent condition, grammatical sequences showed greater activation compared to ungrammatical sequences (AU $\beta = -24.8$, AG $\beta = -21.8$). Paired-sample *t*-tests did not reveal any significant differences between mean beta values for these 4 conditions in this contrast.

3.2.3. Correlation analyses

Exploratory correlation analyses were used to examine associations between activation levels for each contrast and the behavioral measures of learning and awareness questionnaire results. After correcting for multiple comparisons, no significant correlations were observed. Correlation results are reported in Table 3.

4. Discussion

After two sessions of participation in an incidental perceptual sequence learning task involving the reproduction of printed nonsense syllables containing both adjacent and nonadjacent regularities, participants' brain activation during a familiarity task was measured. Two primary contrasts were examined: sequences containing violations of the adjacent regularities (compared to sequences without violations) and sequences containing violations of the nonadjacent regularities (compared to sequences without violations). In addition, two additional contrasts were examined to explore significant differences between the adjacent and nonadjacent conditions. Consistent with our hypotheses, distinct sets of brain areas showed significant activation for the two types of sequential dependencies.

For the adjacent dependency contrast, significant activation was observed with the more stringent threshold of $Z > 3.1$ in two clusters: posterior regions including right superior lateral occipital cortex and angular gyrus (BA 19, 39) and frontal regions including right frontal pole, middle frontal gyrus, and inferior frontal gyrus (BA 44, 45). Considering the posterior activation first, lateral occipital cortex is known to mediate visual processing. Interestingly, implicit learning of sequential dependencies is believed to recruit modality-specific perceptual regions (e.g., Conway & Pisoni, 2008; Frost et al., 2015; Turk-Browne et al., 2009). In fact, it has been argued that much of implicit pattern learning can be thought of as a type of perceptual learning (Chang & Knowlton, 2004; Conway & Christiansen, 2005; Conway, 2005; Conway, Goldstone, & Christiansen, 2007), in which sequences consistent with the learned structure confer a perceptual facilitation effect and processing advantage. In the present case, this visual processing region showed lower levels of activation for sequences with consistent sequential structure relative to sequences containing violations of the (adjacent) sequential structure, similar to what has been observed in perceptual categorization studies (Reber et al., 1998). This cluster also included angular gyrus (BA 39). Left angular gyrus is generally considered to be part of the language network and perhaps specifically might be involved in transferring visual information to Wernicke's area to make meaning out of visually perceived words (Horwitz, Rumsey, & Donohue, 1998), which appears to be feasible considering the present task involved (nonsense) words. However, because the activation was right-lateralized, perhaps the more likely possibility is that the observed activity reflected the allocation of attention to salient features (i.e. letters) of the nonsense strings, a function attributed to the right angular gyrus (Seghier, 2013). Interestingly, greater levels of activation were observed for sequences containing violations of the adjacent dependencies. This could arise from a sort of "pop-out" effect where unexpected stimuli (or parts of stimuli) provoked greater levels of attention (Kristjánsson, Vuilleumier, Schwartz, Macaluso, & Driver, 2007). This is consistent with recent work showing that attention reorganizes as statistical structure in input sequences is learned (Hard, Meyer, & Baldwin, 2018; Zhao, Al-Aidroos, & Turk-Browne, 2013). Another function attributed to the angular gyrus is memory encoding

³ Three participants did not have OSpan data due to failure of the Eprime program used to present stimuli, leaving 18 participants for analyses with the OSpan.

Table 2

Whole brain results. (A) Regions showing greater activation for ungrammatical sequences compared to grammatical sequences in the adjacent condition with cluster threshold of $Z > 3.1$, $p < .05$ and (C) the nonadjacent condition with cluster threshold of $Z > 1.8$, $p < .05$; (B) Regions showing greater activation associated with familiarity judgement of violations in the adjacent condition compared to the nonadjacent condition with cluster threshold of $Z > 3.1$, $p < .05$ and (D) the nonadjacent condition compared to the adjacent condition cluster threshold of $Z > 1.8$, $p < .05$. Abbreviations: R = right hemisphere, L = left hemisphere, MFG = medial frontal gyrus, IFG = inferior frontal gyrus. Co-ordinates of peak-voxels are in MNI space.

	# Voxels in cluster	Region	Brodmann area	Max Z value	Co-ordinates of peak-voxel		
$Z > 3.1$		(A) Adjacent (AU > AG)					
	215	R superior lateral occipital cortex & angular gyrus	19, 39	4.4	32	-64	36
	108	R frontal pole, MFG, & IFG	44, 45	4.3	52	38	24
	165	(B) (AU-AG) > (NU-NG)					
		R superior lateral occipital cortex	19	4.2	34	-66	32
$Z > 1.8$	777	(C) Nonadjacent (NU > NG)					
		L subcallosal cortex, paracingulate gyrus, & anterior cingulate cortex	32	3.8	-2	30	-8
	1527	(D) (NU-NG) > (AU-AG)					
		R frontal pole, paracingulate gyrus, subcallosal cortex, & frontal medial cortex	9, 10	4.2	8	64	14

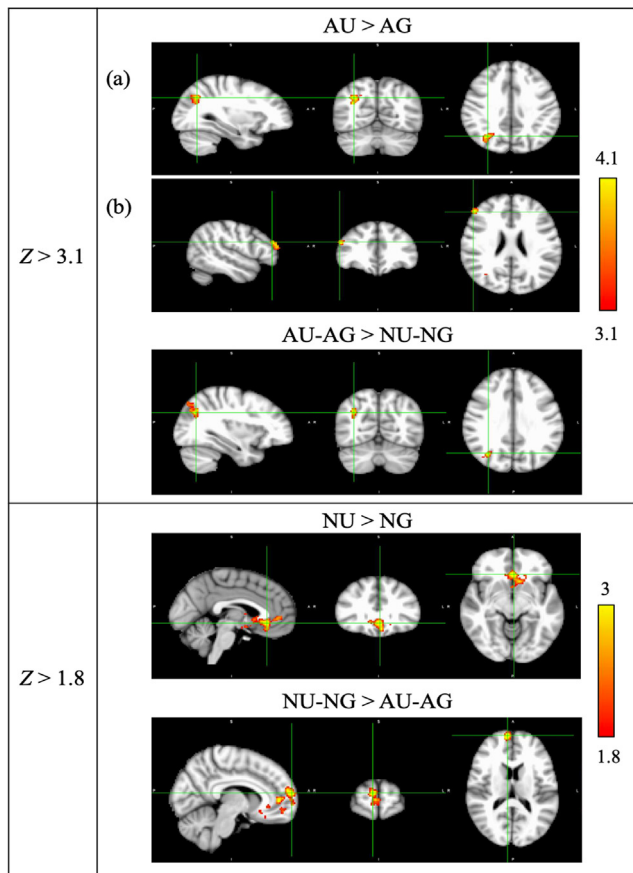


Fig. 5. Whole brain analyses, using cluster threshold of $Z > 3.1$, revealed greater activation for (a) ungrammatical sequences compared to grammatical sequences in the adjacent condition in a cluster consisting of right superior lateral occipital cortex and right angular gyrus as well as (b) a cluster consisting of right frontal pole, right MFG, and right IFG; (c) greater activation associated with familiarity judgement of violations in the adjacent condition compared to the nonadjacent condition was evident in a cluster consisting of right superior lateral occipital cortex. Exploratory group-level analyses, using a less stringent cluster threshold of $Z > 1.8$, revealed greater activation for (d) ungrammatical sequences compared to grammatical sequences in the nonadjacent condition in a cluster consisting of left subcallosal cortex, left paracingulate gyrus, and left anterior cingulate cortex; (e) greater activation associated with familiarity judgement of violations in the nonadjacent condition compared to the adjacent condition was evident in a cluster consisting of right frontal pole, right paracingulate gyrus, right subcallosal cortex, and frontal medial cortex.

and retrieval (Tibon, Fuhrmann, Levy, Simons, & Henson, 2019), the latter of which would likely be engaged while participants completed the familiarity task in the scanner.

In addition to the posterior activation, frontal regions also showed significant activation for the adjacent dependency contrast, specifically right frontal pole, middle frontal gyrus, and inferior frontal gyrus, all parts of the PFC. The PFC is known to be part of a larger frontoparietal network that underlies working memory (Sarnthein et al., 1998). Specifically, two separate frontoparietal networks have been proposed (Corbetta & Shulman, 2002): a dorsal frontoparietal network that consists of bilateral superior frontal, inferior parietal, and superior temporal cortices that is thought to mediate goal-directed orienting and maintenance of attention (Hopfinger, Buonocore, & Mangun, 2000) and a ventral frontoparietal network consisting of right-lateralized posterior parietal cortex and inferior and middle frontal gyri that is postulated to be important for stimulus-driven attention and detecting unexpected or infrequent stimuli (Corbetta, Kincade, Ollinger, McAvoy, & Shulman, 2000). The areas of activation observed in the current study appear to show some amount of overlap with the ventral frontoparietal network. Clearly, the familiarity task in which participants engaged while in the scanner likely recruits both working memory (to observe a sequence and compare to remembered sequences in long-term memory) and attention (to attend to the sequence and possibly notice violations, as discussed above). Another way to interpret the increased activity in PFC for ungrammatical sequences is that increased working memory resources might have been needed in order to process sequences containing violations (consistent with behavioral evidence that it is easier to process and remember sequences consistent with learned structure compared to sequences inconsistent with the structure, e.g., Conway et al., 2010; Karpicke & Pisoni, 2004; Page & Norris, 2009).

On the other hand, the processing of sequences containing violations of nonadjacent dependencies elicited increased activation in the ACC, paracingulate gyrus, and subcallosal gyrus. It should be noted that these areas were only found to be significantly active using the less stringent threshold of $Z > 1.8$ so these results must be interpreted with caution. When interpreting such findings, it is important to ascertain whether the areas of the brain implicated for this experimental condition are consistent with previous findings and/or a priori hypotheses. Recent theorizing has posited that the learning of nonadjacent dependencies is a more challenging process than learning adjacent dependencies (Creel et al., 2004; Gómez, 2002; Newport & Aslin, 2004) and specifically requires cognitive control and inhibition functions (De Diego-Balaguer et al., 2016). These are exactly the types of processing operations the implicated brain areas are thought to engage. Specifically, previous studies reported that activity in these regions was associated with cognitive inhibition and flexibility, error/conflict detection, as well as allocating attentional resources and selecting appropriate responses (Carter et al., 1995; Coderre et al., 2008;

Table 3
Correlation results for behavioral measures and active clusters with threshold of $Z > 3.1$ and $Z > 1.8$ using nonparametric analyses: Kendall's tau (continuous variables) and point-biserial (dichotomous variables). Abbreviations: R = right hemisphere, L = left hemisphere, MFG = medial frontal gyrus, IFG = inferior frontal gyrus. Refer to Table 1 for a list of post-scan interview questions.

	% AG endorsed as familiar	% NG endorsed as familiar	% AU endorsed as familiar	% NU endorsed as familiar	OSpan partial score	Post-scan sequence reproduction confidence	Post-scan pattern awareness	Post-scan awareness of violations	Learning adjacent dependencies (session 1)	Learning nonadjacent dependencies (session 1)	Learning adjacent dependencies (session 2)	Learning nonadjacent dependencies (session 2)
Adjacent (AU > AG), $Z > 3.1$												
R superior lateral occipital cortex & angular gyrus	-0.19	-0.24	-0.37*	-0.21	0.00	-0.23	-0.03	0.26	-0.34*	-0.21	0.30	-0.04
R frontal pole, MFG, & IFG	-0.45	-0.15	-0.18	0.05	-0.08	-0.04	0.01	0.18	-0.01	-0.17	0.01	0.32
(AU-AG) > (NU-NG), $Z > 3.1$												
R superior lateral occipital cortex	0.04	-0.13	-0.28	-0.05	-0.11	-0.37	-0.19	0.06	-0.17	-0.14	0.20	-0.39
Nonadjacent (NU > NG), $Z > 1.8$												
L subcallosal cortex, paracingulate gyrus, & anterior cingulate cortex	0.00	-0.01	-0.14	-0.15	0.12	-0.14	-0.26	0.08	0.11	0.09	0.46**	-0.14
(NU-NG) > (AU-AG), $Z > 1.8$												
R frontal pole, paracingulate gyrus, subcallosal cortex, & frontal medial cortex	-0.01	-0.13	-0.22	-0.20	0.04	-0.10	-0.22	0.001	0.19	0.17	0.38*	-0.09

* $p < .05$.** $p < .01$.

Gennari, Millman, Hymers, & Mattys, 2018; Kemmotsu et al., 2005; Nebel et al., 2005; Nobre et al., 1997; Woodward et al., 2006). Why might such cognitive operations be necessary for processing non-adjacent dependencies? In order to detect a violation of a nonadjacent dependency, one may need to utilize cognitive control to inhibit processing of intervening items occurring between the nonadjacent dependencies and focus on the long-distance dependency itself. A similar effect has been observed in natural language: cognitive control and inhibition appear to be important aspects of processing natural language syntax (January, Trueswell, & Thompson-Schill, 2009; Novick, Trueswell, & Thompson-Schill, 2005). We believe it is likely that this cluster of activity was involved in mediating cognitive control and inhibition when processing the syllable sequences, which was needed in order to detect violations of nonadjacent dependencies.

When directly comparing the adjacent condition to the nonadjacent condition, we found significantly greater levels of activation (using $Z > 3.1$) for the right superior lateral occipital cortex (BA 19) for the adjacent relative to nonadjacent condition. On the other hand, we observed significantly greater levels of activity (using $Z > 1.8$) in the right frontal pole, paracingulate gyrus, subcallosal cortex, and frontal medial cortex (BA 9, 10) for the nonadjacent relative to adjacent condition. Thus, the evidence suggests that there may be distinct neural networks that mediate the processing of violations of adjacent and nonadjacent sequential dependencies. Conway (under review) proposed two primary cortical mechanisms that mediate the learning and processing of sequential patterns. First, the general principle of cortical plasticity results in improved processing and perceptual facilitation of encountered stimuli in a modality-specific manner (Reber, 2013). This would account for activity in posterior regions such as the lateral occipital cortex. Second, an attention-dependent system, mediated by PFC and related networks involved in working memory and cognitive control, can modulate learning and is necessary to learn certain kinds of patterns including nonadjacent dependencies (De Diego-Balaguer et al., 2016). One reason to believe that low-level sensory regions are insufficient to detect violations of nonadjacent dependencies is that these areas of the brain appear only capable of processing stimuli over shorter time-scales, perhaps on the order of tens to hundreds of milliseconds, which would be insufficient to process information across a non-adjacent dependency that spans multiple items in the sequence. Due to the hierarchical arrangement of cortical networks, down-stream brain networks such as the PFC and superior parietal cortex are able to process information over longer periods of time (Fuster & Bressler, 2012; Hasson et al., 2015; Kiebel, Daunizeau, & Friston, 2008). For instance, Wacogne et al. (2011) found that violations of “local” (i.e. adjacent) learned dependencies activated modality-specific perceptual brain regions whereas “global” violations spanning over longer periods of time activated a distributed network that included more anterior brain regions. Although the $Z > 1.8$ contrasts should be taken with caution, the present findings are consistent with the notion that somewhat distinct networks are used to detect violations of different kinds of sequential dependencies.

5. Conclusions

In summary, consistent with dual-system or multi-component theories of implicit sequential pattern learning (Arciuli, 2017; Conway, under review; Conway & Pisoni, 2008; Daltrozzo & Conway, 2014; Frost et al., 2015; Thiessen & Erickson, 2013; Thiessen et al., 2013), and also consistent with previous neuroimaging studies of sequential learning showing distributed patterns of activity across many different neural regions (Folia & Petersson, 2014; Forkstam et al., 2006; Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006; Lieberman et al., 2004; Reber, 2013) the current evidence suggests that multiple complementary neural networks are involved in incidental sequential processing of adjacent and nonadjacent structures. Detecting violations of adjacent sequential dependencies involved a distributed network of

occipital and frontal regions that likely mediated perceptual, attention, and working memory operations. Crucially, posterior visual processing regions (i.e., lateral occipital cortex) appeared to be uniquely active for detecting violations of adjacent dependencies. We suggest that incidental sequential pattern processing is mediated by a hierarchy of neurocognitive mechanisms that include perceptual processing in modality-specific brain regions but also attention, memory and cognitive control in higher level executive networks (Conway, under review; Conway & Pisoni, 2008; Fuster, 2004). These findings are important for understanding the brain bases of sequential processing, a crucial aspect of human behavior.

Declaration of Competing Interest

The authors declared that there is no conflict of interest.

Acknowledgements

This work was supported by seed grants from the Georgia State University/Georgia Institute of Technology Joint Center for Advanced Brain Imaging and Georgia State's Center for Research on the Challenges of Acquiring Language and Literacy (CMC & TZK) as well as by the National Institute on Deafness and other Communication Disorders (R01DC012037 to CMC). The sponsors had no role in study design, data collection, analysis, or interpretation, or in the decision to submit the article for publication. We would like to thank Dr. Vish Ahluwalia at the Center for Advanced Brain Imaging at Georgia Institute of Technology for his valuable help with the fMRI analyses conducted in this study.

Appendix A. Supplementary material

Supplementary material to this article can be found online at <https://doi.org/10.1016/j.nlm.2020.107175>.

References

- Andersson, J. L. R., Jenkinson, M., & Smith, S. (2007a). Non-linear optimisation. FMRIB technical report TR07JA1. www.fmrib.ox.ac.uk/analysis/techrep.
- Andersson, J. L. R., Jenkinson, M., & Smith, S. (2007b). Non-linear registration, aka spatial normalisation. FMRIB technical report TR07JA2. www.fmrib.ox.ac.uk/analysis/techrep.
- Arciuli, J. (2017). The multi-component nature of statistical learning. *Philosophical Transactions of the Royal Society B*, 372, 20160058. <https://doi.org/10.1098/rstb.2016.0058>.
- Beckmann, C., Jenkinson, M., & Smith, S. M. (2003). General multi-level linear modelling for group analysis in fMRI. *NeuroImage*, 20, 1052–1063.
- Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, 8(12), 539–546.
- Carter, C. S., Mintun, M., & Cohen, J. D. (1995). Interference and facilitation effects during selective attention: An H2150 PET study of Stroop task performance. *NeuroImage*, 2(4), 264–272. <https://doi.org/10.1006/nimg.1995.1034>.
- Chang, G. Y., & Knowlton, B. G. (2004). Visual feature learning in artificial grammar classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 30, 714–722.
- 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. <https://doi.org/10.3389/fpsyg.2015.01182>.
- Coderre, E. L., Filippi, C. G., Newhouse, P. A., & Dumas, J. A. (2008). The Stroop effect in kana and kanji scripts in native Japanese speakers: An fMRI study. *Brain and Language*, 107(2), 124–132. <https://doi.org/10.1016/j.bandl.2008.01.011>.
- Conway, C.M., 2020. How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience and Biobehavioral Reviews* (2020). 10.1016/j.neubiorev.2020.01.032.
- 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. <https://doi.org/10.1016/j.cognition.2009.10.009>.
- Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, And Cognition*, 31(1), 24–39. <https://doi.org/10.1037/0278-7393.31.1.24>.
- Conway, C. M., Deocampo, J. A., Smith, G. N. L., & Eghbalzad, L. (2016). Multiple routes to implicit statistical learning? A dual-network perspective. Talk presented at the 57th annual meeting of the psychonomic society, Boston, MA, November, 2016.
- Conway, C. M., Goldstone, R. L., & Christiansen, M. H. (2007). Spatial constraints on visual

- statistical learning of multi-element scenes (pp. 185–190). Austin, TX: Cognitive Science Society.
- 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. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience and Biobehavioral Reviews*. (in preparation).
- Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience*, 3, 292–297.
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews: Neuroscience*, 3(201), 201–215.
- Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, And Cognition*, 30(5), 1119–1130. <https://doi.org/10.1037/0278-7393.30.5.1119>.
- Daltrozzo, J., & Conway, C. M. (2014). Neurocognitive mechanisms of statistical-sequential learning: What do event-related potentials tell us? *Frontiers in Human Neuroscience*, 8, 437. <https://doi.org/10.3389/fnhum.2014.00437>.
- De Diego-Balaguer, R., Martínez-Alvarez, A., & Pons, F. (2016). Temporal attention as a scaffold for language development. *Frontiers in Psychology*, 7, 44. <https://doi.org/10.3389/fpsyg.2016.00044>.
- Deocampo, J. A., King, T. Z., & Conway, C. M. (2019). Concurrent learning of adjacent and non-adjacent dependencies in visuo-spatial and visuo-verbal sequences. *Frontiers in Psychology*, 10, 1107. <https://doi.org/10.3389/fpsyg.2019.01107>.
- Fletcher, P., Büchel, C., Josephs, O., Friston, K., & Dolan, R. (1999). Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. *Cerebral Cortex*, 9, 168–178.
- Folia, V., & Petersson, K. M. (2014). Implicit structured sequence learning: An fMRI study of the structural mere-exposure effect. *Frontiers in Psychology*, 5, 41. <https://doi.org/10.3389/fpsyg.2014.00041>.
- Forkstam, C., Hagoort, P., Fernandez, G., Ingvar, M., & Petersson, K. M. (2006). Neural correlates of artificial syntactic structure classification. *NeuroImage*, 32, 956–967.
- Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, 43(2), 226–236. <https://doi.org/10.3758/s13421-014-0461-7>.
- Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences*, 103(7), 2458–2463.
- 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(3), 117–125.
- Fuster, J. M. (2001). The prefrontal cortex – An update: Time is of the essence. *Neuron*, 30, 319–333.
- Fuster, J. M. (2004). Upper processing stages of the perception-action cycle. *Trends in Cognitive Sciences*, 8(4), 143–145.
- Fuster, J. M., & Bressler, S. L. (2012). Cognit activation: A mechanism enabling temporal integration in working memory. *Trends in Cognitive Sciences*, 16(4), 207–218. <https://doi.org/10.1016/j.tics.2012.03.005>.
- Gennari, S. P., Millman, R. E., Hymers, M., & Mattys, S. L. (2018). Anterior paracingulate and cingulate cortex mediates the effects of cognitive load on speech sound discrimination. *NeuroImage*, 178, 735–743. <https://doi.org/10.1016/j.neuroimage.2018.06.035>.
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13(5), 431–436. <https://doi.org/10.1111/1467-9280.00476>.
- Hard, B. M., Meyer, M., & Baldwin, D. (2018). Attention reorganizes as structure is detected in dynamic action. *Memory & Cognition*. <https://doi.org/10.3758/s13421-018-0847-z>.
- Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences*, 19(6), 304–313.
- Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience*, 3(3), 284–291.
- Horwitz, B., Rumsey, J. M., & Donohue, B. C. (1998). Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proceedings of the National Academy of Sciences*, 95(15), 8939–8944. <https://doi.org/10.1073/pnas.95.15.8939>.
- 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.
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17, 825–841.
- Karipke, J. D., & Pisoni, D. B. (2004). Using immediate memory span to measure implicit learning. *Memory & Cognition*, 32, 956–964.
- Kemmotsu, N., Villalobos, M. E., Gaffrey, M. S., Courchesne, E., & Müller, R. (2005). Activity and functional connectivity of inferior frontal cortex associated with response conflict. *Cognitive Brain Research*, 24(2), 335–342.
- Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, 303(5660), 1023–1026.
- Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology*, 4(1), e1000209.
- Kristjánsson, A., Vuilleumier, P., Schwartz, S., Macaluso, E., & Driver, J. (2007). Neural basis for priming of pop-out during visual search revealed with fMRI. *Cerebral Cortex*, 17(7), 1612–1624.
- Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, 19(12), 1247–1252. <https://doi.org/10.1111/j.1467-9280.2008.02233.x>.
- Lany, J., Gómez, R. L., & Gerken, L. (2007). The role of prior experience in language acquisition. *Cognitive Science*, 31(3), 481–507. <https://doi.org/10.1080/15326900701326584>.
- Leung, H. C., Gore, J. C., & Goldman-Rakic, P. S. (2002). Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. *Journal of Cognitive Neuroscience*, 14(4), 659–671. <https://doi.org/10.1162/08989290260045882>.
- Lieberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S. Y., & Knowlton, B. J. (2004). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *Journal of Cognitive Neuroscience*, 16(3), 427–438.
- Nebel, K., Wiese, H., Stude, P., de Greiff, A., Diener, H., & Keidel, M. (2005). On the neural basis of focused and divided attention. *Cognitive Brain Research*, 25(3), 760–776. <https://doi.org/10.1016/j.cogbrainres.2005.09.011>.
- Newport, E. L., & Aslin, R. N. (2004). Learning at distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127–162. [https://doi.org/10.1016/S0010-0285\(03\)00128-2](https://doi.org/10.1016/S0010-0285(03)00128-2).
- Nobre, A. C., Sebestyen, G. N., Gitelman, D. R., Mesulam, M. M., Frackowiak, R. S., & Frith, C. D. (1997). Functional localization of the system for visuospatial attention using positron emission tomography. *Brain*, 120, 515–533.
- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, and Behavioral Neuroscience*, 5(3), 263–281.
- Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, 53(2), 225–237. <https://doi.org/10.1016/j.jml.2005.02.011>.
- Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 3737–3753. <https://doi.org/10.1098/rstb.2009.0173>.
- Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298, 604–607.
- Perruchet, P., Tyler, M. D., Galland, N., & Peereboom, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General*, 133(4), 573–583. <https://doi.org/10.1037/0096-3445.133.4.573>.
- Poletiek, F. H., Conway, C. M., Ellefson, M. R., Lai, J., Bocanegra, B. R., & Christiansen, M. H. (2018). Under what conditions can recursion be learned? Effects of starting small in artificial grammar learning of center embedded structure. *Cognitive Science*, 42(8), 2855–2889. <https://doi.org/10.1111/cogs.12685>.
- Rämä, P., Martinkauppi, S., Linnankoski, I., Koivisto, J., Aronen, H. J., & Carlson, S. (2001). Working memory of identification of emotional vocal expressions: An fMRI study. *NeuroImage*, 13(6), 1090–1101. <https://doi.org/10.1006/nimg.2001.0777>.
- Reber, P. J. (2013). The neural basis of implicit learning and memory: A review of neuropsychological and neuroimaging research. *Neuropsychologia*, 51, 2026–2042.
- Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 747–750.
- Romberg, A. R., & Saffran, J. R. (2013). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science*, 37(7), 1290–1318. <https://doi.org/10.1111/cogs.12050>.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928. <https://doi.org/10.1126/science.274.5294.1926>.
- Sarnthein, J., Petsche, H., Rappelsberger, P., Shaw, G. L., & von Stein, A. (1998). Synchronization between prefrontal and posterior association cortex during human working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 95(12), 7092–7096. <https://doi.org/10.1073/pnas.95.12.7092>.
- Sawi, O. M., & Rueckl, J. (2018). Reading and the neurocognitive bases of statistical learning. *Scientific Studies of Reading*. <https://doi.org/10.1080/10888438.2018.1457681>.
- Seghier, M. L. (2013). The angular gyrus: Multiple functions and multiple subdivisions. *The Neuroscientist*, 19(1), 43–61. <https://doi.org/10.1177/1073858412440596>.
- Skosnik, P. D., Mirza, F., Gitelman, D. R., Parrish, T. B., Mesulam, M.-M., & Reber, P. J. (2002). Neural correlates of artificial grammar learning. *NeuroImage*, 17(1306), 1314.
- Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17, 143–155.
- Thiessen, E. D., & Erickson, L. C. (2013). Beyond word segmentation: A two-process account of statistical learning. *Current Directions in Psychological Science*, 22(3), 239–243. <https://doi.org/10.1177/0963721413476035>.
- Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, 139(4), 792–814.
- Tibon, R., Fuhrmann, D., Levy, D. A., Simons, J. S., & Henson, R. N. (2019). Multimodal integration and vividness in the angular gyrus during episodic encoding and retrieval. *The Journal of Neuroscience*, 39(22), 4365–4374. <https://doi.org/10.1523/JNEUROSCI.2102-18.2018>.
- Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, 21, 1934–1945.
- Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92, 231–270.

- Van den Bos, E., Christiansen, M. H., & Misyak, J. B. (2012). Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. *Journal of Memory and Language*, 67, 507–520.
- Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (2016). Concurrent statistical learning of adjacent and nonadjacent dependencies. *Language Learning*, 66(1), 8–30. <https://doi.org/10.1111/lang.12137>.
- Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences*, 108(51), 20754–20759. <https://doi.org/10.1073/pnas.1117807108>.
- Wan, L., Dienes, Z., & Fu, X. (2008). Intentional control based on familiarity in artificial grammar learning. *Consciousness & Cognition*, 17(4), 1209–1218.
- Woodward, T. S., Ruff, C. C., & Ngan, E. C. (2006). Short- and long-term changes in anterior cingulate activation during resolution of task-set competition. *Brain Research*, 1068(1), 161–169.
- Woolrich, M. W. (2008). Robust group analysis using outlier inference. *NeuroImage*, 41(2), 286–301.
- Worsley, K. J. (2001). Statistical analysis of activation images (Ch. 14) In P. Jefferard, P. M. Matthews, & S. M. Smith (Eds.). *Functional MRI: An introduction to methods* Oxford Scholarship Online <https://doi.org/10.1093/acprof:oso/9780192630711.003.0014>.
- Xang, J. X., Leung, H. C., & Johnson, M. K. (2003). Frontal activations associated with accessing and evaluating information in working memory: An fMRI study. *Neuroimage*, 20(3), 1531–1539.
- Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science*, 24(5), 667–677. <https://doi.org/10.1177/0956797612460407>.