Journal of Visualized Experiments

Measuring statistical learning across modalities and domains in school-aged children via an online platform and neuroimaging techniques

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE61474R2
Full Title:	Measuring statistical learning across modalities and domains in school-aged children via an online platform and neuroimaging techniques
Section/Category:	JoVE Neuroscience
Keywords:	statistical learning; development; fMRI; neural mechanisms; domain; modality
Corresponding Author:	Julie M. Schneider University of Delaware Newark, Delaware UNITED STATES
Corresponding Author's Institution:	University of Delaware
Corresponding Author E-Mail:	juschnei@udel.edu
Order of Authors:	Julie M. Schneider
	Anqi Hu
	Jennifer Legault
	Zhenghan Qi
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$2,400)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Newark, DE, USA

Julie M. Schneider, Ph.D.
Department of Linguistics and Cognitive Science
University of Delaware
125 E. Main St.
Newark, DE 19716, United States
juschnei@udel.edu

May 27, 2020

Editorial Office
JOVE

Re-submission of our manuscript entitled: ***Measuring statistical learning across modalities and domains in school-aged children via an online platform and neuroimaging techniques*** **by Julie M. Schneider, Anqi Hu, Jennifer Legault and Zhenghan Qi.**

Dear Editor:

Please find attached our revised manuscript entitled "Measuring statistical learning across modalities and domains in school-aged children via an online platform and neuroimaging techniques", for review as a publication in JOVE.

We want to thank the reviewers and Editor for their thoughtful critiques of the current manuscript, which we believe to be improved by these revisions. We have provided a detailed point by point response to each of the reviewers' comments. We have also included both clean and 'marked up' versions of the manuscript. The 'marked up' version includes track changes denoting all revisions that were made. In both versions we have also addressed comments raised by the Editor, and revised the document to be in accordance with the JOVE formatting standards. We have also included a section related to participant recruitment, and ensured that highlighted steps make a cohesive story (despite not including a cohesive piece of text) and is no less than 1 page and no more than 2.75 pages including headings and spacings.

We believe the current protocol can promote research in statistical learning across a wide range of populations and allow for a deeper understanding of the neural mechanisms underlying this ability. Given that *JOVE* publishes research related to novel research protocols similar to our own, we believe that this manuscript will be of interest across the broad range of readership.

We hope that your Referees will find our work interesting and recommend it for publication.

Sincerely,

Julie M. Schneider

Julie M. Schneider (on behalf of all co-authors)

TITLE:

Measuring Statistical Learning Across Modalities and Domains in School-Aged Children Via an Online Platform and Neuroimaging Techniques

AUTHORS AND AFFILIATIONS:
Julie M. Schneider[1], Anqi Hu[1], Jennifer Legault[1], Zhenghan Qi[1]

[1]Department of Linguistics and Cognitive Science, University of Delaware, Newark, DE, USA

Corresponding author:
Julie M. Schneider (juschnei@udel.edu)
Zhenghan Qi (zqi@udel.edu)

Email addresses of co-authors:
Anqi Hu (anqihu@udel.edu)
Jennifer Legault (jlegault@udel.edu)

KEYWORDS:
statistical learning, web-based, fMRI, domain, modality, children

SUMMARY:
Presented here is a protocol introducing a set of child-friendly statistical learning tasks geared towards examining children's learning of temporal statistical patterns across domains and sensory modalities. The developed tasks collect behavioral data using the web-based platform and task-based functional magnetic resonance imaging (fMRI) data for examining neural engagement during statistical learning.

ABSTRACT:
Statistical learning, a fundamental skill to extract regularities in the environment, is often considered a core supporting mechanism of the first language development. While many studies of statistical learning are conducted within a single domain or modality, recent evidence suggests that this skill may differ based on the context in which the stimuli are presented. In addition, few studies investigate learning as it unfolds in real-time, rather focusing on the outcome of learning. In this protocol, we describe an approach for identifying the cognitive and neural basis of statistical learning, within an individual, across domains (linguistic vs. non-linguistic) and sensory modalities (visual and auditory). The tasks are designed to cast as little cognitive demand as possible on participants, making it ideal for young school-aged children and special populations. The web-based nature of the behavioral tasks offers a unique opportunity for us to reach more representative populations nationwide, to estimate effect sizes with greater precision, and to contribute to open and reproducible research. The neural measures provided by the functional magnetic resonance imaging (fMRI) task can inform researchers about the neural mechanisms engaged during statistical learning, and how these may differ across individuals on the basis of domain or modality. Finally, both tasks allow for the measurement of real-time learning, as changes in reaction time to a target stimulus is tracked across the exposure period. The main

limitation of using this protocol relates to the hour-long duration of the experiment. Children might need to complete all four statistical learning tasks in multiple sittings. Therefore, the web-based platform is designed with this limitation in mind so that tasks may be disseminated individually. This methodology will allow users to investigate how the process of statistical learning unfolds across and within domains and modalities in children from different developmental backgrounds.

INTRODUCTION:

Statistical learning is an elementary skill supporting the acquisition of rule-governed combinations in language inputs[1]. Successful statistical learning ability in infants predicts later language learning success[2,3]. Variability in statistical learning skills in school-aged children has also been associated with vocabulary[4] and reading[5,6]. Difficulty in statistical learning has been proposed as one etiological mechanism underlying language impairment[7]. Despite the association between statistical learning and language outcomes in both neurotypical and atypical populations, the cognitive and the neural mechanisms underlying statistical learning remain poorly understood. In addition, previous literature has revealed that, within an individual, statistical learning ability is not uniform but independent across domains and modalities[6,8,9]. The developmental trajectory of statistical learning abilities may further vary across domains and modalities[10]. These findings emphasize the importance of assessing individual differences in statistical learning across multiple tasks throughout the course of development. However, the field first requires a more systematic investigation of the relationship between statistical learning and first language development. To address these questions, we apply innovative methods including a web-based testing platform[11] that reaches a large number of children, and laboratory-based neuroimaging techniques (functional magnetic resonance imaging, or fMRI) that examine the real-time encoding of statistical information.

Standard measures of statistical learning begin with a familiarization phase and are followed by a two-alternative forced choice (2-AFC) task[12,13]. The familiarization phase introduces a continuous stream of stimuli embedded with statistical regularities, where some stimuli are more likely to co-occur than others. The presentation of these co-occurring stimuli follows a fixed temporal order. Participants are passively exposed to the stream during the familiarization phase, followed by a 2-AFC task that tests whether the participant successfully extracted the patterns. The 2-AFC accuracy task presents two consecutive sequences: one sequence has been presented to the participant during the familiarization phase, while the other is a novel sequence, or contains part of the sequence. Above-chance accuracy on the 2-AFC would indicate successful learning at the group level. Traditional behavioral tasks assessing statistical learning generally rely upon accuracy as the outcome measure of learning. However, accuracy fails to account for the natural learning of information as it unfolds in time. A measure of real-time learning is necessary to tap into the implicit learning process of statistical learning during which children are still encoding the regularities from the inputs[14,15,16]. Various adaptations across paradigms have been developed in an effort to move away from the 2-AFC measure, towards measures of on-line learning through behavioral responses during the exposure[16]. Studies utilizing these adaptations which measure reaction time during the exposure phase found they were related to post-learning accuracy[17] with better test-retest reliability compared to that of the accuracy in adult learners[18].

Neural measures are also foundational to our understanding of how learning unfolds over time, as the implicit process by which language learning occurs likely recruits different neural resources from those used once language is learned[19]. Neural measures also provide insights into differences in cognitive specializations underlying language ability across special populations[20]. How the condition contrast is designed in an fMRI study is crucial for how we interpret patterns of neural activation during learning. One common practice is to compare brain responses during the familiarization phase between sequences containing regular patterns versus those containing the same stimuli which are ordered randomly. However, previous research implementing such a random control condition found no evidence for learning in behavior, despite neural differences between structured and random sequences. This might be due to the interference of random sequences on learning of structured sequences, as both were constructed from the same stimuli[21,22]. Other fMRI studies which utilized backward speech or earlier learning blocks as the control condition confirmed learning took place behaviorally[19,23]. However, each of these paradigms introduced its own confounding factor, such as the effect of language processing for the former case and the effect of the experimental order for the latter case. Our paradigm uses the random sequence as the control condition but mitigates their interference on participants' learning of the structured sequences. Our fMRI paradigm also implements a mixed block/event-related design, which allows for the simultaneous modeling of transient trial-related and sustained task-related BOLD signals[24]. Lastly, and more broadly, neural measures allow for the measurement of learning in populations where eliciting an explicit behavioral response may be difficult (e.g., developmental and special populations)[25].

The current protocol adopts a response time measure, in addition to traditional accuracy measures, and examines brain activation during the familiarization phase. The combination of these methods aims to provide a rich dataset for the investigation of real-time learning processes. The web-based platform offers a set of learning measures by including both response time during the exposure phase and accuracy of the 2-AFC task during the test phase. The neuroimaging protocol allows for the investigation of the underlying neural mechanisms supporting statistical learning across domains and modalities. While it is optimal to measure statistical learning within an individual using both the web-based and fMRI protocols, the tasks are designed so that they may be disseminated independently, and therefore, as two independent measures of statistical learning. The fMRI experiments included in the current protocol can help clarify how stimulus encoding, pattern extraction, and other constituent components of statistical learning are represented by particular brain regions and networks.

PROTOCOL:

All participants gave written consent to participate and study was conducted in accordance with the Institutional Review Board.

1. Overview of the statistical learning paradigm utilized in the web-based protocol

1.1. Include four tasks in the current paradigm: image (visual-nonlinguistic), letter (visual-linguistic), tone (auditory-nonlinguistic), and syllable (auditory-linguistic).

1.1.1. Construct stimuli for visual tasks using 12 standalone alien cartoon images (image) and 12 letter images (letter; B, J, K, A, H, C, F, E, J, G, D, M) showing the same alien holding up 12 signs with capital letters written on them.

1.1.2. Construct auditory stimuli using 12 English syllables (syllable; pi,pu,pa,ti,tu.ta,di,du,da,bi,bu,ba) and 12 musical tones within the same octave (tone; F,G,D,G#,C#,B,C,F#,D#,E,A,A#). The syllable stimuli can be made using an artificial speech synthesizer, are recorded as separate files in Praat[26,27].

1.2. In the familiarization phase, present stimuli in a structured stream (see **Figure 1**). Feedback is not provided at any point during the familiarization or test phase.

NOTE: Within each task, a familiarization phase is immediately followed by a test phase.

1.2.1. For the Image (visual-nonlinguistic) task, structure 12 images into four target triplets. In the familiarization phase, repeat each of the four target triplets 24 times for a total of 96 triplets.

NOTE: The 96 triplets are randomly concatenated into a continuous stream, with the constraint that no triplet can be immediately repeated. Images are presented one at a time in the center of the screen. Each image is presented for 800 ms with 200 ms of inter-stimulus interval. The whole familiarization phase will last for 4 min 48 s.

1.2.2. Ensure the test phase always follows the familiarization phase and is composed of 32 two-alternative forced-choice (2AFC) questions. For each question, include 2 options: a target triplet from the familiarization phase and a triplet that was not included in the familiarization phase, referred to as a foil triplet.

NOTE: Foil triplets are constructed so that the relative position of each image in the foil triplet is the same as the target triplet. Each target and foil triplet are presented 8 times total in a test, and each foil-target pair is repeated. The test phase consists of 32 (4 target triplets x 4 foil triplets x 2 repetitions) randomly ordered trials.

1.2.3. For the letter (visual-linguistic) task include 12 images of capitalized letters that are organized into four target triplets (GJA, FKC, LBE, and MDH). For the test phase, create 4 foil triplets (GDE, FJH, LKA, and MBC) to pair to the target triplets to form the 32 2AFC test trials. No letter triplet contains any words, common acronyms, or initialisms.

1.2.4. For the tone (auditory-nonlinguistic) task include 12 musical pure tones within the same octave (a full chromatic scale starting from middle C) and concatenate them into four target triplets (F#DE, ABC, C#A#F, and GD#G#). Unlike in the visual tasks, presentation speed is faster due to differences in auditory perceptual preference[6,28,29].

NOTE: Each of the four target triplets is repeated 48 times for a total of 192 triplets (twice as much as the visual conditions). All triplets are concatenated into a sound stream with no triplet being repeated twice in a row. Pure tones are presented one at a time while participants view a blank screen. The duration of each tone is 460 ms with a 20 ms inter-stimulus interval. The whole stream lasts about 4 min and 36 s. As in the visual tasks, a test phase of 32 2AFC trials with pairs of target and foil triplets (F#BF, AA#G#, C#D#E, GDC) immediately follows the familiarization phase.

1.2.5. For the syllable (auditory-linguistic) task use 12 consonant vowel (CV) syllables created and grouped into four target triplets (pa-bi-ku, go-la-tu, da-ro-pi, and ti-bu-do). The duration of each syllable and the inter-stimulus interval is the same as the tone condition. Pair four foil triplets (pa-ro-do, go-bu-ku, da-bi-tu, and ti-la-pi) with the target triplets in the test phase.

1.3. Randomize the order of the four statistical learning tasks across participants.

2. Participant recruitment

NOTE: While the web-based protocol and the fMRI protocol are best implemented together within a single participant, here we outline the best practices for participant recruitment for each task independently.

2.1. Web based participant recruitment

2.1.1. Recruit participants age 6 years and above. Participants of any sex, race, and ethnicity may participate; however, study sample should be representative of the population.

2.1.2. Recruit participants who are a native English speaker and have been exposed to no languages besides English before the age of 5.

2.1.3. Ensure they report no known psychological (including ADD, depression, PTSD, and clinical anxiety) and/or neurological condition (including stroke, seizure, brain tumor, or closed head injury).

2.1.4. Ensure participants have normal or corrected-to-normal vision (glasses or contacts are okay), normal color vision and normal hearing (no hearing aid or cochlear implant devices).

2.2. Task based fMRI participant recruitment

2.2.1. Recruit participants age 6 years and above. Participants of any sex, race, and ethnicity may participate; however, study sample should be representative of the population.

2.2.2. To be eligible, recruit participants who are native English speaker and have never been exposed to any languages besides English before the age of 5.

2.2.3. Recruit right-handed individuals, with no known psychological (including ADD, depression, PTSD, and clinical anxiety) and neurological condition (including stroke, seizure, brain tumor, or closed head injury).

2.2.4. Exclude participants who are pregnant, claustrophobic, taking psychoactive drugs, or have any metal in the body (including pacemakers, neural implants, metal plates or joints, shrapnel, and surgical staples).

2.2.5. Ensure participants have normal or corrected-to-normal vision (glasses or contacts are okay), normal color vision and normal hearing (no hearing aid or cochlear implant devices).

2.2.6. Determine the eligibility to participate in the MRI by having participants (or parents if the participant is a minor) complete an MRI Safety Screening Form.

3. Web based protocol

NOTE: The web-based statistical learning paradigm is hosted on a secure website (https://www.cogscigame.co[11]) and developed using jsPsych, a JavaScript library for creating behavioral experiments online[30].

3.1. To reproduce tasks, go to DOI: 10.5281/zenodo.3820620. All scripts and materials are publicly available. Researchers can modify the scripts and run the experiments locally on any web browser as long as all the paths for the output files are set up appropriately.

3.2. Have participants complete a target detection cover task during all familiarization phases of each statistical learning task. Have participants press the space bar as soon as they see an image/letter or hear a tone/syllable during the familiarization phase.

3.3. Target stimulus assignment for each task

3.3.1. In the image, letter and syllable tasks, randomly choose one of the four triplets and assign the target to the third stimulus of the triplet. In the tone task, constrain the target stimulus to only the lowest or the highest tones of the third stimulus in the triplets, and assign the target to the third stimulus of the triplet. This is done because tone stimuli are relatively harder to discriminate than other types of stimuli.

3.3.2. In the syllable and tone tasks, introduce participants to an alien and the favorite word/note in its alien language/folk music. Tell participants that they will listen to the alien's language/music and to remember to press the spacebar whenever they hear the favorite word/ note. Give participants a practice trial where they must press the space bar as soon as the alien's favorite word/note is heard.

3.3.3. In the image task, tell participants to keep track of a special alien as a group of aliens line up to enter a spaceship. In the letter task, tell participants to keep track of the alien's favorite sign as the alien holds up signs for a parade. Give participant's a practice trial in both the image and letter tasks.

3.3.4. Do not provide explicit instructions about the presence of triplets.

3.3.5. Measure response time over the 24 trials in the visual tasks and over the 48 trials in the auditory tasks to assess online learning.

3.3.6. During the test phase, both a target (included in familiarization phase) and foil triplet (not included in familiarization phase) are presented to the participant. Instruct participants to then choose which one of the two is more similar to what they saw or heard in the familiarization phase. Each trial must end with a response.

3.4. Behavioral measures of statistical learning in the web-based protocol

3.4.1. Measure the real time learning during the familiarization phase via the linear slope of reaction time (change in reaction time throughout the familiarization phase).

3.4.2. To be considered a valid response to the target, check that the keypress must be in the time window of one stimulus before and one stimulus after the target stimulus. That is -480 ms to +960 ms relative to the onset of the target in the auditory tasks and -1000 ms to +2000 ms in the visual tasks. A keypress prior to the target is considered as anticipation and thus yields a negative reaction time.

3.4.3. To compare reaction times across conditions, transform the reaction times of each participant for each task into z scores. This normalizes the reaction times of an individual so they are less affected by the task differences and can be compared.

3.4.4. Calculate a reaction time slope of each participant for each condition using linear regression. Input the z-normed reaction times as the dependent variable and the target trial order as the independent variable (visual: 1 to 24; auditory: 1 to 48). The slope of the linear regression line is the reaction time slope (RT slope).

3.4.5. Measure offline accuracy of each participant for each condition by dividing the number of correct trials from the test phase by the total number of trials (32 trials).

4. Task based fMRI protocol

4.1. Modifications to the statistical learning paradigm (**Figure 2**).

4.1.1. For each task, present both a structured sequence (containing statistical regularities) and a random sequence (no statistical regularities).

NOTE: Structured sequences are identical to those described for the web-based protocol (see **Figure 1**). In contrast, random sequences contain the same 12 stimuli as presented in the structured sequences but are ordered pseudo-randomly. No combinations of any three stimuli are repeated more than once.

4.1.2. Divide each sequence into six smaller blocks of equal length (24 stimuli for the visual tasks and 48 stimuli for the auditory tasks).

4.1.3. Concatenate three structured blocks, 3 random blocks, and 6 resting blocks (silence with a blank screen) in a pseudorandom order to create four runs of auditory stimuli and four runs of visual stimuli. To maximize learning of the structured sequences, ensure that the random blocks in each run contain a different domain from the structured sequence (e.g., syllable structured sequences are presented together with tone random sequences in one run, and syllable random sequences are presented together with tone structured sequences in another run).

4.1.4. Include 288 images to be presented in each run for the visual task lasting approximately 4.77 min. Include 576 sounds to be presented in the auditory task which lasts approximately 4.42 min. At the beginning of each block, present a cue about the target with a verbal and visual probe: "Now listen/look for the [TARGET]".

4.1.5. Among the four runs of the visual task, ensure that two contain structured sequences of images and the other two contain structured sequences of letters. Among the four runs of the auditory task, ensure that two contain structured sequences of syllables and the other two contain structured sequences of tones.

4.2. fMRI statistical learning procedure

4.2.1. To help make participants, especially children get comfortable in the scanner, practice the MRI scanning session first using a mock scanner[31]. A mock scanner provides a naturalistic experience similar to the actual scanning session but is typically situated in a more child-friendly environment.

4.2.2. First introduce the child to the mock scanner, i.e., brain camera, and ensure they are comfortable before putting them in the scanner.

4.2.3. Introduce them to their "scan-buddy" and explain that the purpose of the scan buddy is to keep them accompanied and help them if they need anything. The scan buddy will gently remind the participant to keep still if too much motion is detected by the "camera".

4.2.4. Once they are in the scanner, play child-friendly videos to help them acclimate to the sound and video. When they are ready, play a few pre-recorded scanner sound clips to prepare them for the noises produced by the real MRI. During this time have them practice staying still and working with the scan buddy.

4.2.5. Introduce children to the statistical learning paradigm and have them practice outside of the scanner. This is done by having children complete a brief portion of the task on a computer, similar to the web-based protocol by performing steps 3.2.2 and 3.2.3 mentioned above.

NOTE: The practice stimuli are the same as those utilized in the task; however, children are only exposed to the random sequence and not the structured sequences, allowing for brief habituation to the stimuli and task demands without enabling learning of particular sequences.

4.2.6. Ensure the fMRI data collection protocol is appropriately set up on the MRI acquisition computer.

NOTE: The acquisition parameters follow the recommendations of the Adolescent Brain Cognitive Development (ABCD) Study[32].

4.2.7. Begin the scanning session with high resolution T1-weighted scans. Acquire these using a 176-slice 3D MPRAGE (Magnetization Prepared Rapid Gradient Echo) volume scan with TR (Repetition Time) = 2500 ms, TE (Echo Time) = 2.9 ms, flip angle = 8˚, FOV (Field of View) = 25.6 cm, 256 X 256 matrix size, and 1 mm slice thickness. This acquisition will last 7.2 min.

4.2.8. To acquire functional data, use T2*-weighted echo-planar imaging with simultaneous multi-slice scans acquisition with TR= 800 ms, TE = 32 ms, flip angle = 61˚, FOV = 21 cm, and matrix = 64 x 64. In this experiment, 60 adjacent slices are acquired in an interleaved sequence with 2.5 mm slice thickness, a 21 cm FOV, and a 64 X 64 matrix, resulting in an in-plane resolution of 2.5 mm x 2.5 mm x 2.5 mm.

4.2.9. Have participants lie comfortably on the bed of the fMRI scanner with headphones that protect their ears from the scanner noise and a response pad/button box in their hand (both headphones and button box must be scanner compatible).

4.2.10. Place additional padding around their head to ensure limited head motion during data collection. Give the button response box to the participant ahead of time to record responses and counterbalance whether the left or right hand is used to press buttons across participants.

4.2.11. Give every child an option of a scan buddy. For older, neurotypical children who are comfortable without a scan buddy, give them a squeeze ball to notify the experimenter if they are distressed or need to stop. Give younger children and special populations a squeeze ball but also provide them with a scan buddy to assist them (described in 4.2.3).

4.2.12. Place the head coil over the participant's head and align the patient's position in the bed.

4.2.13. On the acquisition computer register a new participant. Enter their participant ID, date of birth, weight and height. The participant may now be inserted into the bore of the MRI.

4.2.14. Acquire T1-weighted scan while showing participants a movie.

4.2.15. Before beginning the statistical learning paradigm, give participants the instructions of each task by speaking to them through an intercom system connected to their headphones.

4.2.16. In the auditory tasks, tell participants: “Now we’re going to play a button-pressing game. You will hear the aliens say words and play music. Remember to press the button in your LEFT/RIGHT hand whenever you hear the sound you are listening for. There will be 4 parts, and each part will last about 5 min.”

4.2.17. In the visual tasks, tell participants: “Now you are going to see the pictures of the aliens and the letters. Whenever you see the picture you are looking for, press the button in your LEFT/RIGHT hand. You will play this 4 times in a row. It will take about 5 minutes each time.”

4.2.18. Start the statistical learning paradigm on the presentation computer and acquire the task fMRI data.

4.2.19. Once the participant has completed the paradigm, stop the MRI, safely remove them from inside the scanner, and remove the head coil.

4.2.20. After data collection, transfer all MRI data from the acquisition computer to a secured server for further analyses.

4.3. fMRI data analyses

4.3.1. Analyze in-scanner reaction time during the fMRI task similarly to the web-based calculation of reaction time during the familiarization phase. Normalize reaction time to compare across conditions, and calculate a linear slope using the normalized reaction time for each condition of an individual.

4.3.2. When analyzing the fMRI data, first organize and convert data to Brain Imaging Data Structure[33] (BIDS) formatting using HeuDiConv[34] (https://github.com/nipy/heudiconv).

4.3.3. Preprocess these data using fMRIPrep[35,36]. This automated preprocessing pipeline combines methodology from AFNI[37], ANTs[38], Freesurfer[39], FSL[40], and Mindboggle[41] to provide scientifically rigorous and reproducible data for use in data analysis.

NOTE: The current study implements a mixed block/event-related design. The representative results (below) treat each mini block as an event (e.g., random sequence is an event, structured sequence is an event, etc.). However, the task is also designed so that one can model each stimulus as an event.

4.3.4. Include two task regressors for each run (“image” and “letter” for the visual condition, and “syllable” and “tone” for the auditory condition) in the first-level model design. Determine task regressors by convolving a vector of event onset times with their durations with a canonical hemodynamic response function. Compute differences and means between runs within each subject for higher-level model designs. This will result in a contrast between structured and random sequences within each type of stimuli.

4.3.5. Create a group mean of activation for structured blocks compared to random blocks within each modality/domain.

REPRESENTATIVE RESULTS:

Web-based Behavioral Results

Given the current protocol is designed for easy dissemination with developmental populations, we have included preliminary web-based results based on data from 22 developing school-aged children (Mean (M) age = 9.3 years, Standard Deviation (SD) age = 2.04 years, range = 6.2-12.6 years, 13 girls). In the web-based statistical learning task, children performed significantly better than 0.5 chance-level on all conditions, indicating successful statistical learning at the group level (see **Table 1** for statistics; **Figure 3**). Mean reaction time slope was negative and significantly below 0 in the syllable condition (M = -0.01, SD = 0.02, $t(14) = -2.36$, one-tailed $p = .02$) and marginally significant in the letter condition (M = -0.02, SD = 0.06, $t(15) = -1.52$, one-tailed $p = .07$, **Figure 4**), suggesting a faster acceleration of target detection during the familiarization phase in the linguistic tasks. Mean reaction time slope was not significantly different from zero in the image condition (M = 0.02, SD = 0.04, $t(17) = 1.54$, one-tailed $p > .1$) or the tone condition (M = 0.005, SD = 0.02, $t(15) = -5.7 \times 10^{-17}$, one-tailed $p > .1$), despite evidence of learning in the offline measures of accuracy. Cronbach's alpha was 0.75 for the Letter task, 0.09 for the Syllable task, 0.67 for the Tone task, and 0.86 for the Image task. Correlations between implicit measures (RT slope) and explicit measures (accuracy) of statistical learning identify a significant relationship for the Image task (R = -.48, $p = 0.04$) and Letter task (R = -.54, $p = 0.03$). Inter-task correlations further suggest that the four tasks may have a modest degree of overlapping learning mechanism (**Figure 5**). While accuracy on both visual tasks was highly correlated (R = .60, $p = 0.02$), they were also positively associated with accuracy on the Syllable task (Image R = .66, $p = 0.01$; Letter R = .85, $p < 0.001$).

fMRI Results

Preliminary fMRI results were based on data from 9 developing school-aged children. These 9 children were a subset of the 22 children included in the web-based behavioral results, as not all children came to the lab to complete the fMRI portion of the study. All nine completed the auditory statistical learning tasks (*M* age = 10.77 years, *SD* = 1.96 years, range = 7.7-13.8 years, 4 girls) and seven completed the visual statistical learning tasks (*M* age = 11.41 years, *SD* = 2.37 years, range = 7.7-13.8 years, 4 girls). When comparing structured blocks to random blocks, significant clusters were observed in all four conditions (**Figure 6**). In the syllable condition, greater activation was found at the left superior temporal gyrus, right insula/frontal operculum, and anterior cingulate gyrus. In the tone condition, greater activation was found at left middle temporal gyri, bilateral angular gyri, left frontal pole, right lateral occipital cortex, right insula, and right frontal operculum. In the letter condition, greater activation was found at the left planum temporal. In the image condition, greater activation was found at the right lateral occipital cortex. These preliminary findings suggest that children's neural activation patterns differ across learning of statistical regularities depending on the modality and domain of the presented stimuli. The current task design is sensitive to these differences and can identify task-specific regions of activation similar to past studies[20,25].

fMRI Behavioral Results

To demonstrate learning in the fMRI portion of this study, we have included in-scanner behavioral results from 28 adults (M age = 20.8, SD = 3.53, 20 females), as the data from 9 children was not enough to compute reliable statistics. Our findings in adults indicate that learning successfully occurred in all tasks for the structured sequence, supported by significantly quicker response time in the structured as compared to the random condition, except in the case of the tone task (see **Table 2** for statistics).

Taken together, our web-based measures of accuracy, and increased activation for structured versus random sequences in the scanner, indicate this protocol may be implemented with developmental populations to gauge statistical learning across domains and modalities within an individual. Our behavioral MRI results in an adult population further emphasis the utility of this protocol in measuring learning of structured sequences as it unfolds in real-time, as well as the ability to implement the web-based and fMRI protocols independently.

FIGURE AND TABLE LEGENDS:

Figure 1: Familiarization phase of all four statistical learning tasks. Example triplets across each task are depicted in this figure. Each visual stimulus appeared for 800 ms with a 200 ms ISI, and each auditory stimulus was heard for 460 ms with a 20 ms ISI.

Figure 2: Familiarization modification for fMRI statistical learning tasks. The fMRI task was similar to the web-based familiarization phase but introduced a random sequence that was counterbalanced across domains.

Figure 3: Average statistical learning (SL) accuracy in the web-based task compared against chance-level. Results indicate individuals performed significantly above chance on all four tasks, ***one-tailed $p < .001$, ** < 0.01, * < 0.05.

Figure 4: Mean reaction time slope in the web-based task against zero. A more negative slope indicates faster acceleration in the target detection during familiarization. Target detection significantly improved over the course of exposure during the syllable task. †one-tailed $p = .07$, * < .05.

Figure 5: Web-based between-task correlations across all four statistical learning tasks. (**a**) Non-significant values at an alpha of .05 are shown with a white background. All comparisons with a colored background denote significant effects. (**b**) Sample size for each pairwise comparison.

Figure 6: Neural activation at the group-level for structured blocks compared to random blocks within each modality and domain. Significant clusters were thresholded at voxel-level $p < 0.001$ and cluster-level $p < 0.05$ for each task. Horizontal slices were selected to depict the cluster with

the maximum z-value. The color bar in the bottom, right corner reflects the same scale for all plots.

Table 1: Web-based accuracy by condition. One-sample t-tests represent group differences compared to 0.5 chance-level.

Table 2: MRI behavioral performance differences on random versus structured sequences across all four tasks in adults. Paired-samples t-tests represent group differences in learning of structured versus random sequences.

DISCUSSION:

The methods presented in the current protocol provide a multimodal paradigm for understanding the behavioral and neural indices of statistical learning across the course of development. The current design allows for the identification of individual differences in statistical learning ability across modalities and domains, which can be used for future investigation of the relationship between statistical learning and language development. Since an individuals' statistical learning ability is found to vary across domains and modalities[6,8,9], it is optimal if participants complete all four tasks. Findings from typically developing children and adults indicate that an individuals' performance across statistical learning domains/modalities can differentially relate to vocabulary[4] and reading[5,6] outcomes. Therefore, we recommend additional measures of cognitive and language abilities be taken to relate to the measures of statistical learning taken in the current protocol.

Research has reported reasonable internal consistency and test-retest reliability of these statistical learning tasks for adults[8,42]. However, concerns about task reliability for children[42] and a recent discussion on general measurement issues[9] indicates an urgent need to develop measures of statistical learning, that take into account children's developmental characteristics. While our previous research, as well as the preliminary data from the current protocol, indicates high internal consistency for the non-linguistic statistical learning tasks in school-aged children between 8 and 16 years old[6], our research also confirmed a less satisfying task reliability, particularly in auditory linguistic statistical learning which has been reported before[42]. The differences in internal consistency between tasks are particularly intriguing in light of recent findings on the impact of a learner's prior linguistic experiences on statistical learning outcomes[18,43,44]. Language and reading development change rapidly during the school years. The learnability of each auditory linguistic triplet might differ substantially within each child, depending on their developmental stage and current language abilities. Combining our protocol with other individual difference measures will offer an exciting opportunity to study the cascading effect between existing skills and subsequent learning underlying the heterogeneity of statistical learning performance across the course of development.

An important benefit of the current design is in its' utility for measuring statistical learning via an online web-platform. Researchers should be aware of the following when considering the accuracy of reaction time measurements via a web browser. de Leeuw and Motz (2016)[45] found the response times measured via a web browser were approximately 25 ms longer than those measured via other standard data presentation software. Importantly, this delay was found to be constant across trials. Because our measure of real-time learning in the web-based tasks is the slope of change in reaction time, the effects of the delay in reaction time has been minimized using within-subject comparisons. de Leeuw (2015)[30] has also acknowledged that reaction time measured via jsPsych may be affected by factors such as the processing speed of the computer or the number of tasks loaded in the background. To minimize these effects, we recommend normalizing response time within each individual participant before computing the response time slope[30].

The current protocol, providing methods to demonstrate a large variability in learning behavior across domains and modalities, is designed to investigate individual differences of statistical learning. However, this protocol is not suitable for investigating questions such as whether visual statistical learning is inherently easier than auditory statistical learning. The interpretation for group-level performance difference between tasks is difficult due to all the confounding factors that we are not able to control, such as stimuli familiarity[14,43,46,47] , sensory salience, and processing speed[28]. Related to stimuli familiarity, it is well established that an individual's prior experiences with the stimuli may influence their statistical learning performance. Additionally, the visual and auditory tasks are difficult to directly compare due to differences in the salience of the stimuli and presentation rate across these modalities. Therefore, our methods are designed with the aim of investigating individual differences in statistical learning. However, with advanced fMRI analysis approaches, our protocol is suitable for studying theoretical questions about the nature of statistical learning, for example we can ask which brain networks are sensitive to regularities in each domain and how the patterns of neural engagement differ/overlap.

The current protocol was developed to be child-friendly and easily accessible to maximize research in neurotypical and atypical populations. During the implementation of this protocol with young children or those with developmental disorders, a critical step is to give breaks between each SL task to avoid fatigue. Each condition of the web-based tasks can be disseminated individually to ease cognitive demands. Prior to scanning, the mock scanner can be used to reduce child anxiety and head motion in preparation for the real fMRI task. An additional issue researcher should be aware of relates to a general concern when conducting any neuroimaging study: motion. A rotational head movement of just 0.3 mm can cause artifacts to manifest. In an effort to minimize the likelihood of motion artifacts, the current protocol has limited each run to last less than 5 minutes[48]. Participants should be encouraged to stay still during each 5-minute run but allowed to move or stretch between runs in order to reduce motion during actual scanning. We also recommend rigorous data analysis techniques to correct motion-related artifacts on the fMRI data[49].

Given the critical contribution of statistical learning ability on later language acquisition, it is necessary to develop more comprehensive and reliable measures that assess both real time and offline learning of statistical regularities. The current proposal is a first step towards delineating how individual differences in statistical learning ability based on domain/modality may account for variations in later language outcomes.

ACKNOWLEDGMENTS:

We thank Yoel Sanchez Araujo and Wendy Georgan for their contribution in the initial design of the web-based platform. We thank An Nguyen and Violet Kozloff for their work on improving the web-based statistical learning tasks, implementing the fMRI tasks, and piloting the tasks in adult participants. We thank Violet Kozloff and Parker Robbins for their contribution in assisting data collection in children. We thank Ibrahim Malik, John Christopher, Trevor Wigal, and Keith Schneider at the Center for Biological and Brain Imaging at the University of Delaware for their assistance in neuroimaging data collection. This work is funded in part by the National Institute on Deafness and other Communication Disorders (PI: Qi; NIH 1R21DC017576) and the National Science Foundation Directorate for Social, Behavioral & Economic Sciences (PI: Schneider, Co-PI: Qi & Golinkoff; NSF 1911462).

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES:

1. Saffran, J. R., Newport, L. L., Aslin, R. N. Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*. **35**, 606-621 (1996).

2. Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., Dow, K. A. Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology*. **42** (4), 643–655 (2006).

3. Graf Estes, K., Gluck, S. C. W., Grimm, K. J. Finding patterns and learning words: Infant phonotactic knowledge is associated with vocabulary size. *Journal of Experimental Child Psychology*. **146**, 34–49 (2016).

4. Evans, J. L., Saffran, J. R., Robe-Torres, K. Statistical Learning in Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*. **52** (2), 321–335 (2009).

5. Arciuli, J., Simpson, I. C. Statistical Learning Is Related to Reading Ability in Children and Adults. *Cognitive Science*. **36** (2), 286–304 (2012).

6. Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D. E., Arciuli, J. Hearing Matters More Than Seeing: A Cross-Modality Study of Statistical Learning and Reading Ability. *Scientific Studies of Reading*. **23** (1), 101–115 (2019).

7. Walenski, M., Tager-Flusberg, H. B., Ullman, M. T. Language in autism. *Understanding autism: From Basic Neuroscience to Treatment*. eds Moldin S. O., Rubenstein J. L. R., Boca Raton: Taylor and Francis Books, 175–203 (2006).

8. Siegelman, N., Frost, R. Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*. **81**, 105–120 (2015).

9. Erickson, L. C., Kaschak, M. P., Thiessen, E. D., Berry, C. A. S. Individual Differences in Statistical Learning: Conceptual and Measurement Issues. *Collabra*. **2** (1), 14 (2016).

10. Shufaniya, A., Arnon, I. Statistical Learning Is Not Age-Invariant During Childhood: Performance Improves With Age Across Modality. *Cognitive Science*. **42** (8), 3100–3115 (2018).

11. Qi, Z. et al. An online platform for visual and auditory statistical learning for school-aged

children (Version 1.0.0). Zenodo. http://doi.org/10.5281/zenodo.3820620 (2020).
12. Conway, C. M., Christiansen, M. H. Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science*. **17** (10), 905–912 (2006).
13. Saffran, J. R., Johnson, E. K., Aslin, R. N., Newport, E. L. Statistical learning of tone sequences by human infants and adults. *Cognition*. **70** (1), 27–52 (1999).
14. Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., Frost, R. Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*. **177**, 198–213 (2018).
15. Armstrong, B. C., Frost, R., Christiansen, M. H. The long road of statistical learning research: past, present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **372** (1711), 20160047 (2017).
16. Siegelman, N., Bogaerts, L., Christiansen, M. H., Frost, R. Towards a theory of individual differences in statistical learning. *Transactions of the Royal Society B*. **372** (1711), 20160059 (2017).
17. Krogh, L., Vlach, H. A., Johnson, S. P. Statistical Learning Across Development: Flexible Yet Constrained. *Frontiers in Psychology*. **3** (JAN), 598 (2013).
18. Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., Frost, R. Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*. **177** (August 2017), 198–213 (2018).
19. Plante, E., Patterson, D., Dailey, N. S., Kyle, R. A., Fridriksson, J. Dynamic changes in network activations characterize early learning of a natural language. *Neuropsychologia*. **62** (1), 77–86 (2014).
20. Milne, A. E., Wilson, B., Christiansen, M. H., Petkov, C., Marslen-Wilson, W. Structured sequence learning across sensory modalities in humans and nonhuman primates This review comes from a themed issue on The evolution of language. *Current Opinion in Behavioral Sciences*. **21**, 39–48 (2018).
21. Mcnealy, K., Mazziotta, J., Dapretto, M. Cracking the Language Code: Neural Mechanisms Underlying Speech Parsing. *Journal of Neuroscience*. **26** (29), 7629–7639 (2006).
22. McNealy, K., Mazziotta, J. C., Dapretto, M. Age and experience shape developmental changes in the neural basis of language-related learning. *Developmental Science*. **14** (6), 1261–1282 (2011).
23. Karuza, E. A., Emberson, L. L., Aslin, R. N. Combining fMRI and behavioral measures to examine the process of human learning. *Neurobiology of Learning and Memory*. **109**, 193–206 (2014).
24. Petersen, S. E., Dubis, J. W. The mixed block/event-related design. *NeuroImage*. **62** (2), 1177–1184 (2012).
25. Batterink, L. J., Paller, K. A., Reber, P. J. Understanding the Neural Bases of Implicit and Statistical Learning. *Topics in Cognitive Science*. **11**, 482–503 (2019).
26. Boersma, P., Weenink, D. Boersma, P., Weenink, D. Praat: doing phonetics by computer [Computer program]. Version 6.1.14 (2020).
27. Boersma, P. Praat, a system for doing phonetics by computer. *Glot International*. **5** (9), 341–345 (2001).
28. Conway, C. M., Christiansen, M. H. Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive*

Psychology. **21** (4), 561–580 (2009).

29. Emberson, L. L., Conway, C. M., Christiansen, M.H. Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology*. **64** (5), 1021–1040 (2011).

30. de Leeuw, J. R. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*. **47** (1), 1–12 (2015).

31. Kana, R. K. et al. Probing the brain in autism using fMRI and diffusion tensor imaging. *Journal of Visualized Experiments*. (55), e3178 (2011).

32. Casey, B. J. et al. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*. **32**, 43–54 (2018).

33. Gorgolewski, K. J. et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*. **3** (1), 1–9 (2016).

34. Halchenko, Y. nipy/heudiconv v0.6.0 (Version v0.6.0). Zenodo.

35. Esteban, O. et al. FMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods.* **16**, 111-116 (2019).

36. Esteban, O. et al. Analysis of task-based functional MRI data preprocessed with fMRIPrep. *bioRxiv*. 694364 (2020).

37. Cox, R. W. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*. **29** (3), 162–173 (1996).

38. Avants, B. B. et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*. **54** (3), 2033–2044, doi: 10.1016/j.neuroimage.2010.09.025 (2011).

39. Dale, A. M., Fischl, B., Sereno, M. I. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*. **9** (2), 179–194 (1999).

40. Woolrich, M. W. et al. Bayesian analysis of neuroimaging data in FSL. *NeuroImage*. **45** (1 Suppl), S173-86 (2009).

41. Klein, A. et al. Mindboggling morphometry of human brains. *PLoS Computational Biology*. **13** (2), e1005350 (2017).

42. Arnon, I. Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior Research Methods*. **52** (1), 68–81 (2020).

43. Finn, A. S., Hudson Kam, C. L. The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*. **108** (2), 477–499 (2008).

44. Krogh, L., Vlach, H. A., Johnson, S. P. Statistical learning across development: Flexible yet constrained. *Frontiers in Psychology*. **3** (JAN), 598 (2013).

45. De Leeuw, J. R., Motz, B. A. Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods.* **48**, 1-12 (2016)

46. Lew-Williams, C., Saffran, J. R. All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition*. **122** (2), 241–246 (2012).

47. Poulin-Charronnat, B., Perruchet, P., Tillmann, B., Peereman, R. Familiar units prevail over statistical cues in word segmentation. *Psychological Research*. **81** (5), 990–1003 (2017).

48. Leonard, J., Flournoy, J., Lewis-de los Angeles, C. P., Whitaker, K. How much motion is too much motion? Determining motion thresholds by sample size for reproducibility in

developmental resting-state MRI. *Research Ideas and Outcomes*. **3**, e12569 (2017).

49. Power, J. D. et al. Ridding fMRI data of motion-related influences: Removal of signals with distinct spatial and physical bases in multiecho data. *Proceedings of the National Academy of Sciences of the United States of America*. **115** (9), e2105–e2114 (2018).

Figure 1

Click here to access/download;Figure;Figure

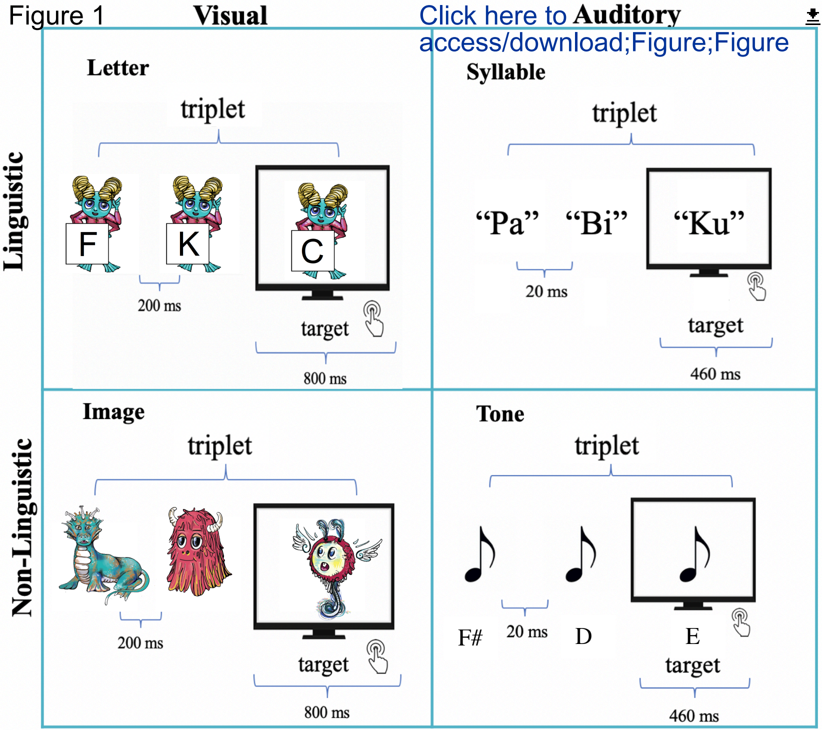


Figure 3

Click here to access/download;Figur

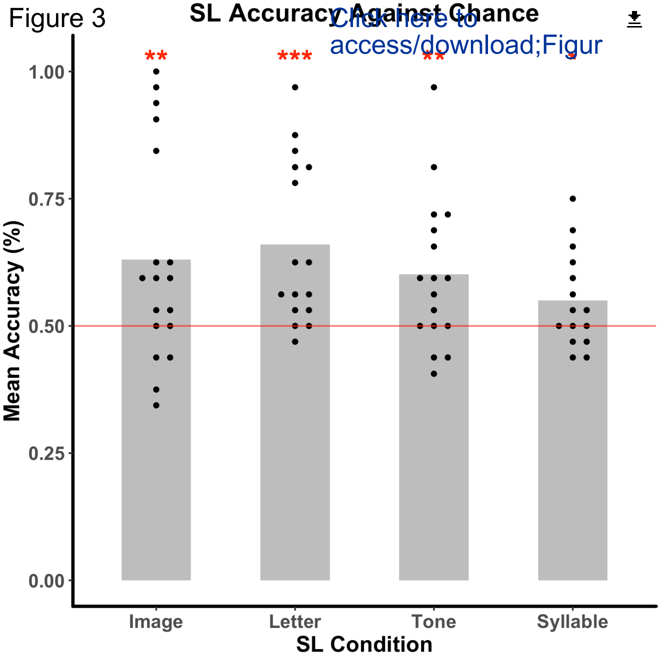


Figure 4

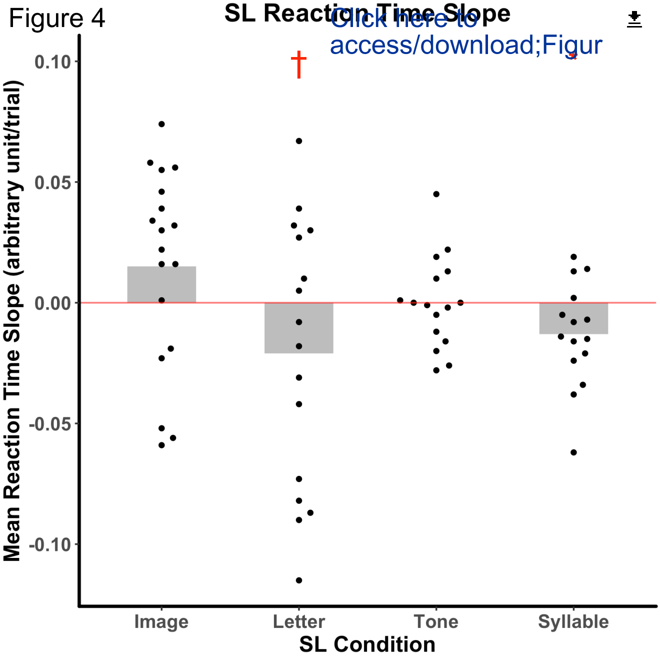

SL Reaction Time Slope
Mean Reaction Time Slope (arbitrary unit/trial)
0.10
0.05
0.00
-0.05
-0.10
†
Image
Letter
Tone
Syllable
SL Condition

Figure 5

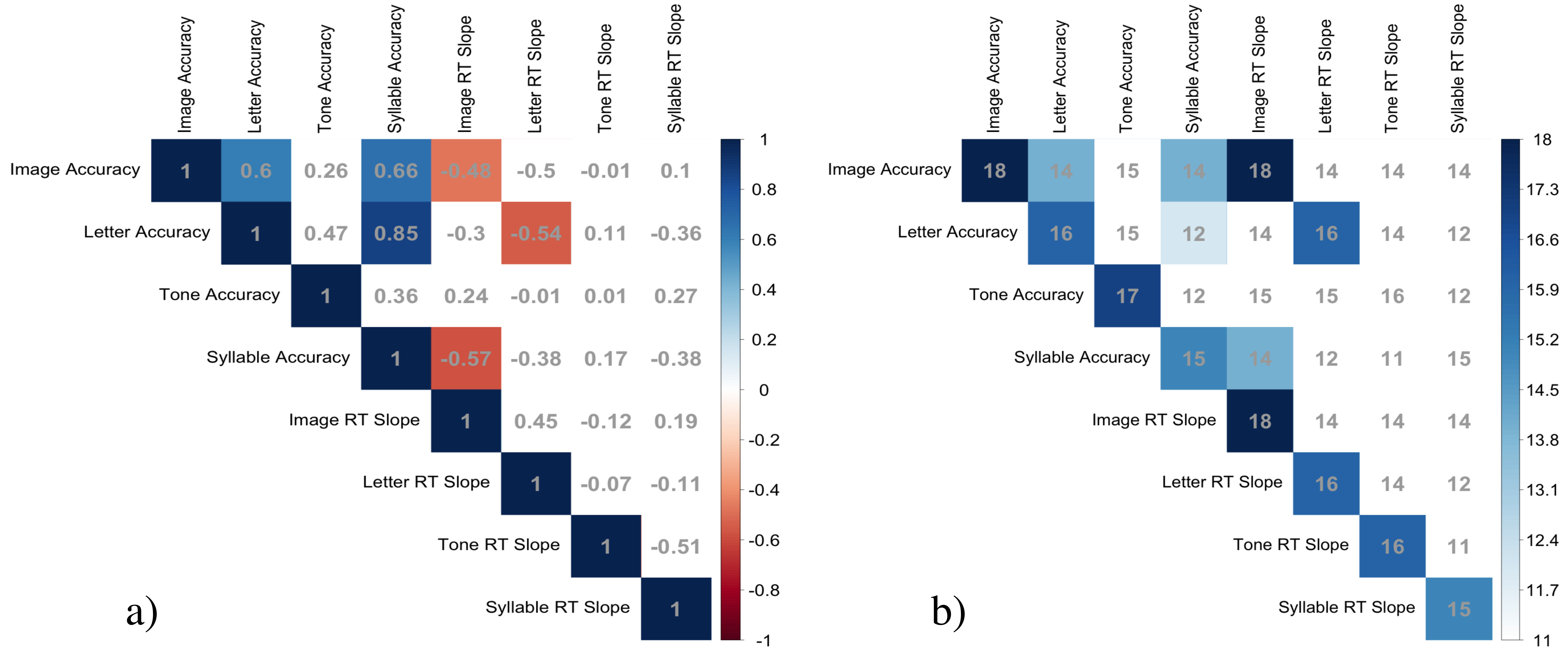
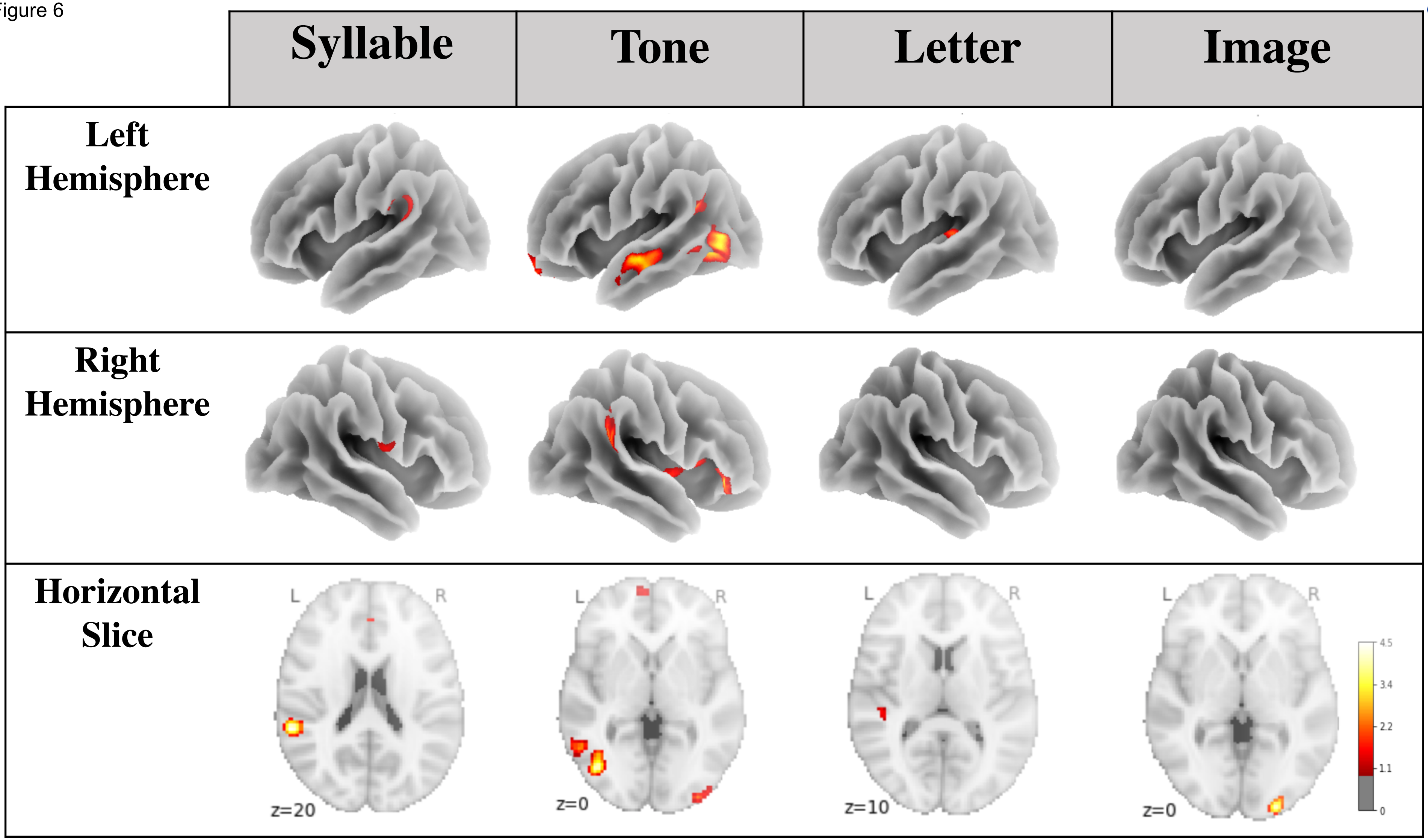


Image Accuracy
Letter Accuracy
Tone Accuracy
Syllable Accuracy
Image RT Slope
Letter RT Slope
Tone RT Slope
Syllable RT Slope
1 0.6 0.26 0.66 -0.48 -0.5 -0.01 0.1
1 0.47 0.85 -0.3 -0.54 0.11 -0.36
1 0.36 0.24 -0.01 0.01 0.27
1 -0.57 -0.38 0.17 -0.38
1 0.45 -0.12 0.19
1 -0.07 -0.11
1 -0.51
1
1 0.8 0.6 0.4 0.2 0 -0.2 -0.4 -0.6 -0.8 -1
a)
18 14 15 14 18 14 14 14
16 15 12 14 16 14 12
17 12 15 15 16 12
15 14 12 11 15
18 14 14 14
16 14 12
16 11
15
18 17.3 16.6 15.9 15.2 14.5 13.8 13.1 12.4 11.7 11
b)

Figure 6


Syllable
Tone
Letter
Image
Left Hemisphere
Right Hemisphere
Horizontal Slice
L
R
z=20
z=0
z=10
z=0
4.5
3.4
2.2
1.1
0

Condition	Mean	Standard Deviation	One-tailed T-test
Image	0.63	0.21	$t(17) = 2.64, p = .009$
Letter	0.66	0.16	$t(15) = 3.98, p < .001$
Tone	0.60	0.15	$t(16) = 2.83, p = .006$
Syllable	0.55	0.1	$t(14) = 2.06, p = .03$

	Structured		Random		
Condition	Mean	Standard Deviation	Mean	Standard Deviation	Paired Samples T-test
Image	468.1	76.04	493.4	60.33	$t(28) = -2.01, p = .05$
Letter	374.72	143.59	502.1	68.75	$t(28) = -4.97, p < .001$
Tone	422.11	165.75	416.97	169.69	$t(28) = 0.90, p = .13$
Speech	586.1	176.98	676.9	58.97	$t(28) = -2.70, p = .01$

Figure 2

Click here to access/download
Video or Animated Figure
Figure2.svg

 Click here to access/download;Table of Materials;JoVE_Table_of_Materials.xlsx

Name of Material/Equipment	Company	Catalog Number
4 Button Inline Response Device	Cambridge Research Systems	SKU: N1348
Short/Slim Canal Tips	Comply Foam	SKU: 40-15028-11
jsPsych	jsPsych	https://www.jspsych.org/
Speech Synthesizer	Praat	Version 6.1.14
Web-based statistical learning tasks	Zenodo	http://doi.org/10.5281/zenc

Comments/Description

An fMRI reponse pad used for measuring in-scanner response time
Short & slim in-ear canal tips are recommended for children to protect hearing and allow for them to hear the stimuli while in the scanner.
jsPsych is a JavaScript library for running behavioral experiments in a web browser.
This program is an artificial speech synthesizer which was used to create the syllable stimuli.

All web-based statistical learning tasks are available for free access on Zenodo.

We greatly appreciate all the reviewers' thoughtful critiques of our protocol. We have addressed each point and feel the manuscript has been strengthened by these changes.

Reviewers' comments:
Reviewer #1:
Overview
The authors have prepared a manuscript that aptly describes their novel protocol for empirical measurement of cognitive and brain outcomes related to multimodal statistical learning. I applaud their commitment to providing the neuroscience community with their materials and the software necessary to administer these tests. There is no question that this work represents a substantial contribution to the methodology of statistical learning. My (minor) comments are only intended to enhance clarity in descriptions of the approach.

Introduction
* Paradigms: The authors might note that passive exposure to stimulus streams is not the only paradigm available for investigations of statistical learning. Other approaches may also offer benefits that the authors identify as important (e.g., opportunities to measure on-line learning through behavioral responses during exposure, either explicit associative tests or implicit measures of, for example, eye movements).
Response: We have acknowledged other approaches that may offer benefits in moving beyond the standard 2-AFC measures in the introduction.

Protocol
* Familiarization phase, Image task: Are triplets arranged in a unique random sequence for each participant, a single pseudo-random sequence for all participants, or some further arrangement? Also, "no triplet can be repeated twice in a row" is redundant; consider "no triplet can be immediately repeated" or "no triplet can be presented twice in a row".

Response: We have updated the wording to reflect that "no triplet can be immediately repeated" and clarified that triplets are arranged in a single pseudo-random sequence. This was because we are studying the individual differences, it is important to keep the exposure consistent across people. We thank the reviewer for requesting this clarification.

* Test phase, Image task: Target triplets are presented four times at test. Are foil triplets also presented multiple times during the test phase? If not, the multiple presentations of target triplets (vs. foil triplets) may unintentionally present incidental learning opportunities. Also, the presence or absence of feedback during the test phase should be indicated.
Response: Each target triplet is presented 8 times in total at test. Each foil triplet is also presented 8 times in total at test. Each foil-target pair are repeated once. 4 target x 4 foil x 2 repetition of each foil-target pair in test phase = 32 trials in total for each condition. We have updated this information to be clearer in the protocol.

* Auditory task labels: Note that the parenthetical descriptions of the two auditory tasks, "(auditory linguistic)" and "(auditory non-linguistic)" are incorrect due to a swap.
Response: Thank you very much for identifying this error, which has now been corrected.

* Stimuli, auditory non-linguistic: Were the tones pure tones?
Response: Tones were in fact pure tones, and this information is now provided in the manuscript on page 4.

* Stimuli, auditory linguistic: Were syllables produced synthetically or recorded speech? If recorded speech, what was the sex of the speaker? Also, was a single recording used, or were multiple recordings of each syllable used (to simulate variability in natural speech)?
Response: The syllable stimuli were made using an artificial speech synthesizer and recorded as separate files in Praat. We have included this in section 1.1.2.

Reviewer #2:
I found this manuscript quite easy to follow and the protocol was mostly clear. Specific comments follow below:

Major(ish) Concerns:
I worry about systematic differences between the four tasks. I understand however at the same time that they might be hard to avoid. This includes e.g. differences in length, repeats in the familiarity phase, target selection rules (lines 166-174)
Response: We thank the reviewer for raising this concern. We have made the discussion of task differences much more thorough and have emphasized that the current study aims to clarify individual differences and is not designed with the ability to directly compare across tasks.

Line 183: One can become faster over time without learning anything about sequences, i.e. one would be expected to become faster for the stimulus in the third position of a triplet both due to general learning effects and statistical learning, but becoming faster for a stimulus in a first position would be more general learning but not statistical learning. It seems impossibly to unconfound these two in the protocol. It also varies by task, as the tone task allows for a target not just in the third position.
Response: The reviewer brings up an excellent point, that all targets occur in the triplet final position. To ascertain that learning is statistical in nature in the structured condition (and in the web-based tasks) the MRI paradigm includes a random task in which target detection occurs in a random pattern. We have added preliminary results from an adult sample who was tested outside of an MRI scanner to determine whether the "structured" stream did in fact elicit faster RT's than the "random" stream. Given mean RTs were faster for structured streams on all tasks (except in the auditory, non-linguistic, tone task) we feel confident that our tasks are measuring statistical learning. Additionally, the target for the tone task is in the third position and we have re-worded section 3.2.1. to clarify this point.

Unless I missed it, there are no stats on the internal consistency/reliability of the tasks. If people want to adopt this protocol to study individual differences in SL, then they need that information. Also, I suggest correlating the implicit measures (online RT changes) with the explicit measures (accuracy) - they might not necessarily tap into the same underlying mechanisms. One measure could also be more reliable than the other. It is also unclear to which degree the four tasks overlap in their mechanisms. I suggest providing a correlation matrix with RTs and accuracy from all four tasks, along with internal consistency estimates (alpha or such).

Response: We have provided data for internal consistency/reliability of the tasks and a correlation matrix between implicit and explicit measures of statistical learning in the results section (Figure 5).

Minor Concerns:
Line 120: "…twelve letter images… showing the same alien holding up twelve signs with capital letters written on them" What is the reasoning behind this choice, i.e. why not just have the twelve letter images without the aliens? This will become the most visually complex subtask, and seemingly unnecessarily so.
Response: We designed all four tasks in a theme of "meet your alien friends". An alien image on the screen also made the task more engaging for children. Based on our preliminary results in children, the letter task designed this way does not appear to challenging.

Line 142: The tones are not linguistic but are described as such.
Response: Thank you very much for identifying this writing error, which has now been corrected.

Line 143: Here you could indicate the auditory foil triplets as you did with the letter task.
Response: We have now included the foil triplets for the auditory tasks under section 1.2.

Line 151: Speech is described as nonlinguistic but should likely be linguistic.
Response: Thank you very much for identifying this writing error, which has now been corrected.

Line: 152: Explain CV acronym (consonant vowel).
Response: We have explained the CV acronym.

Line 158: What if people want to run the protocol from another platform, e.g. if their IRB only allows them to collect data from a local server (e.g. GDPR)? Can they download the jsPsych script, and then how?
Response: We are in the process of making our scripts and materials available for the community. All information can be found via the DOI: 10.5281/zenodo.3820620. Researchers can modify our scripts and run the experiments locally on any web browser as long as all the paths for the output files are set up appropriately. This information is provided in 3.1.

Lines 208-210: I don't understand this sentence, please clarify. What does "different type of stimuli" mean? How can random blocks be presented every two experimental blocks if the experiment contains the same number of experimental and random blocks? Or doesn't it?
Response: We appreciate the reviewer bringing our attention to this statement. We have clarified that we meant different domain: "To maximize learning of the structured sequences, the random blocks in each run contain a different domain from the structured sequence (e.g. speech structured sequences are presented together with tone random sequences in one run, and speech random sequences are presented together with tone structured sequences in another run)" (Section 4.1.3).

Line 238: "This practice should not expose them to the actual statistical regularities of any task" -

I suggest adding that this should not expose them to any of the stimuli actually used. If they are exposed to the same stimuli, then this needs some justification.
Response: The practice stimuli are the same as those utilized in the task; however, children are only exposed to the random sequence and not the structured sequences. The purpose of doing this was to ensure they were comfortable with the actual stimuli being used and found the stimuli to be recognizable. We have provided this information in a note following section 4.2.5.

Line 254: Is the child not given a squeeze ball or something like that to notify the experimenter if they feel uncomfortable in the scanner? Or is someone inside the scanner room with the child at all times?
Response: We thank the reviewer for acknowledging the need for this information, which has now been provided. Specifically, older, neurotypical children are given a squeeze ball to notify the experimenter if they are distressed or need to stop. Younger children and special populations are given a squeeze ball but are also provided with a scan buddy to assist them (described in 4.2.11).

Line 254: I don't know how idiot-proof this needs to be, but should it be said explicitly that e.g. headphones and button box needs to be scanner compatible?
Response: In addition to our recommendations for scanner compatible headphones and button box in the 'Table of Materials', we have also included a statement that both should be scanner compatible.

Line 269: "Now we're going to play a button-pressing game. You will hear the aliens say words and play music" - this is the first time that we hear about the aliens saying something or playing music. You have already mentioned that the aliens are holding letters, I think that this info belongs in the same section.
Response: This is the portion of the protocol where instructions are explicitly given to the participant describing what they are to do next. The instructions the reviewer is referring to are associated with the auditory tasks, and the next line reviews instructions for the visual tasks, one condition of which aliens do hold letters (in no other conditions do aliens hold letters). We have made it more explicit which tasks each set of instructions are for.

Line 294: My fMRI skills are rusty, but are you treating this as an event-related design instead of a block design? "Task regressors were determined by convolving a vector of event onset times with their durations with a canonical hemodynamic response function." Should you not convolve a boxcar instead of "event spikes"? And are you treating each image as an event?
Response: The current study implements a mixed block/event-related design. We are treating each mini-block as an event (e.g. random sequence is an event, structured sequence is an event, etc.). However, the task is also designed so that one can model each stimulus as an event if they are interested. We have provided more information about this as a note in section 4.3.

Line 385: "Current theories of statistical learning are mixed on whether this skill is domain-general versus domain-specific" - I think this is true that this is has not been fully settled. However, you then contradict yourselves, e.g. line 360: "…individuals' statistical learning ability is found to vary across domains and modalities".

Response: While some theories advocate for a domain-general perspective, other advocate that an individuals' statistical learning ability is found to vary across domains and modalities, indicating the capacity might be constrained by domain-specific cognitive processes. We removed this sentence to avoid confusion.

I don't recall seeing a description of instructions given to the participants in the web-based paradigm. Are they in their tasks e.g. explicitly required to emphasize accuracy, are they asked to respond quickly etc.? I would suggest adding these as it can affect participant behavior.
Response: We thank the reviewer for pointing out the lack of descriptions of web-based task procedure. We have now added the procedure in section 3.2.1-3.2.4. In the familiarization phase, participants were instructed to respond as soon as they see/ hear the target stimulus. In the test phase, though there was no emphasis on accuracy, participants must see/ hear both the target and foil triplets and must give a response.

There is no check to see whether participants recognize the stimuli. I put this in minor concerns as this is generally not done in statistical learning paradigms - but it should be done in my opinion. E.g. if you don't recognize the aliens on their own, then this would lead to lower 2AFC accuracy in the testing phase, but this would not reflect poor statistical learning, but poor recognition abilities. Adding a 2AFC task where participants distinguish between old and new aliens (or syllables or whatnot) would be my suggestion so that this can be ruled out.
Response: We appreciate the reviewers' concern regarding participant recognition of stimuli. We do include a practice session using the same stimuli utilized in the task to ensure participants recognize the aliens in this study; however, we do not explicitly test them on this knowledge afterwards. We have included this information generally in the discussion focusing on limitations in the study related to familiarity of stimuli (or lack thereof).

Reviewer #3:
Manuscript Summary:
The paper describes a paradigm for testing statistical learning (SL) abilities across modalities (auditory vs. visual) and across domains (linguistic vs. nonlinguistic), which assesses learning in three distinct ways: (1) an implicit behavioral measure (RTs); (2) an explicit behavioral measure (accuracy on a 2AFC task); and (3) a neural measure (regions of interest in fMRI). The paradigm is especially adapted for children, as well as for large-scale behavioral data collection via an online website.

While the paradigm is clearly closely based on existing tasks that have already been described in detail in previous studies, it does nicely bring together several methodological advantages. For example, many behavioral studies compared SL across the visual and auditory modalities, and many fMRI studies tested auditory or visual learning in isolation, but I am unaware of any fMRI study to date that compared SL across modalities. Similarly, there are only a handful of papers that looked at statistical learning as it unfolds over time during the learning phase. The protocol described in the paper has multiple strong points in its favor.
In general, I found this paper to be clear, concise, and well-written, with good methodological motivations and a detailed protocol. This is a timely paper given the recent criticism on classic SL tasks with offline measures and with a small sample size, and fits well with the current research. It is likely to be of interest to linguists, developmentalists, and experimental

psychologists who are interested in learning mechanisms.

I do have several important comments and requests for clarifications, which I would like the authors to address. While these points do not exclude publication, they should be addressed in the revised version:

Major Concerns:
1. While the abstract claims that the protocol allows "for the measurement of real-time learning" as change in RTs during the familiarization phase, no such analysis is actually provided (not in the protocol description, and not in the preliminary results). This is quite misleading. If this is indeed one of the strong sides of the paradigm, this missing data should be provided (i.e., how did learning unfold over time). Otherwise, it's best to remove this claim from the abstract.
Response: We agree that providing measures of real-time learning as change in RTs during the familiarization phase is a critical goal of the current study. We have included measures of RT change across the familiarization period for each task in the results (Figure 4) and protocol (Section 3.2.5./3.3.4.) and have emphasized how this measure relates to real-time learning more clearly.

2. Similarly, the abstract and introduction emphasize the comparison of SL abilities across domains and modalities. However, no such comparison is actually made. Albeit preliminary, the result should still be presented in a scientifically valid manner. Even though reporting the mean accuracy and RTs in tables and plots is a great visual demonstration, it is not sufficient for claiming anything about performance in auditory vs. visual modalities, or linguistic vs. nonlinguistic domains. Since this is arguably one of the main promises of the current protocol, please provide some statistical tests to demonstrate it.
Response: The current protocol aims to emphasize the examination of individual differences in statistical learning abilities across different domains and modalities. Specifically, in the current protocol we are interested in demonstrating whether a) these tasks are learnable and whether b) they are sensitive to variability across participants. There are a few caveats in the method that prevent us from doing direct comparisons on a group level that have now discussed in great detail in the discussion (related to stimuli saliency, stimuli familiarity, and processing difficulty). We have also included between-task correlations (Figure 5).

3. The abstract also mention examining individual differences in learning, yet no such measures/analyses are provided. All the results are presented at the group level. There is not even a correlation of performance across modalities (e.g., are participants consistent in their SL abilities in the four different tasks). While this is not the main aim of the protocol, the reader is led to believe that all these measures/analyses would be provided by the paradigm.
Response: While the current protocol does in fact allow for the measurement/analysis of this data, it was not included due to sample size/a lack of power (as mentioned by the reviewer in the Minor Concerns #5). However, our group results include individual values to demonstrate that, even with a limited sample size, individual variability exists within and across tasks. We have also provided some measures of individual differences utilizing between-task correlations of implicit and explicit learning (Figure 5).

4. Since we know that statistical learning is heavily influenced by prior knowledge (e.g.,

Siegelman et al., 2018), it actually seems quite problematic to compare the performance on linguistic and non-linguistic stimuli in the current paradigm. Given that the non-linguistic stimuli chosen for both modalities is quite novel (as the authors mention, tones are quite hard to distinguish for non-musicians, and the alien figures are completely novel to participants), it is not really possible (or fair) to compare them to well-known stimuli like letters and syllables, with which participants have plenty of experience. As such, in order to allow for a proper comparison across domains, it would make much more sense to use non-linguistic stimuli that is also familiar. For example, instead of novel aliens, the visual task should use triplets of known inanimate items (e.g., table, flower, brush) or animate animals (e.g., dog, tiger, bear). Similarly, the auditory task should use known environmental sounds (e.g., dog bark, glass breaking; See Shufaniya & Arnon, 2018), which are processed much more easily than tones.
Response: We completely agree with the reviewer that prior experiences have effects on SL learning outcomes. The suggestions of using familiar stimuli for non-linguistic tasks is indeed a reasonable step to reduce task differences. However, despite these steps, there are still other differences in the stimuli that are difficult to control. Therefore, it is not easy to make a direct comparison between tasks. We now emphasize in a number of places throughout our protocol that the method was designed to investigate the variation of performance across individuals. We also included a section in the discussion about the caveats of direct comparisons on task performance.

5. I was missing a more in-depth discussion of the protocol weaknesses. This is especially relevant in light of point 4 above (prior knowledge) and also points 3-4 made below (e.g., presentation rates, type of regularity).
Response: We have included a more in-depth overview of protocol weaknesses throughout the discussion.

Minor Concerns:
1. I found the literature review to be quite sloppy, and lacking key papers that would actually help support the authors' argument. For example, Karuza, Emberson & Aslin (2014) strongly advocated for combining fMRI with behavioral methods, and Milne, Wilson and Christiansen (2018) have an extensive review of behavioral and fMRI studies that examined SL across multiple modalities. I am not an author on any of these papers and have nothing to gain by recommending this - I just believe that the current paper would benefit from engaging more with the existing literature.
Response: We have revised the introduction and included a more thorough review of key papers and the relevance of combining MRI and behavioral methods.

2. Similarly, I was missing some discussion on how the results relate to previous studies with similar tasks. I understand that this is a methods paper and therefore there is less emphasis on the results, but in order to ensure the validity of the protocol it is important to establish that it produces results that are in line with those already obtained in the literature (e.g., a learning advantage in the visual modality).
Response: Our preliminary results demonstrated that all these tasks are learnable and there is substantial variation across individuals. We now explained more clearly in the revision that this protocol is not intended for a direct comparison between tasks, but rather is designed to provide robust measures for individual differences in SL. While we agree with the reviewer that more in-

depth discussion about our results and how they are related to past literature will strengthen our paper, we also have to adhere the page limit requirements and style guidelines set by the journal.

3. Given that processing of visual and auditory stimuli differs substantially, I completely understand why the presentation rates were adapted between the auditory and the visual task (i.e., much faster presentation rates in auditory modality). However, it is unclear to me why the number of repetitions is doubled in the auditory modality? I understand that given the different presentation rates across modalities, doubling the repetitions of the auditory stimuli results in a familiarization phase of similar length in both auditory and visual tasks. But if participants received twice as much input in one modality, this also makes it hard to compare learning across modalities. Is this addressed in the comparison between the auditory and visual task? And if not, why not?
Response: The decisions on the repetition times were based on previous literature (Saffran et al., 1997 for auditory and Arciuli & Simpson, 2012 for visual). We tried to keep the overall exposure time similar between the two modalities (around 5 minutes) and we also found comparable performance at the group level between the two non-linguistic tasks in school-aged children, suggesting fewer exposure times in the visual domain did not lead to worse performance (Qi et al., 2019). We have included a much more thorough discussion of differences between the auditory and visual tasks in the discussion section.

4. I wonder why the visual task is temporal, rather than spatial? One of the things that have been established in previous work (e.g., Emberson, Conway & Christiansen, 2011) is that the auditory and visual modalities differ not only in their preferred presentation rates, but also in their preferred regularities (i.e., temporal patterns are harder to learn in the visual modality, as opposed to spatial ones). In this case, why not fully optimize processing and have the visual task with spatial triplets instead of temporal? I am aware that most visual SL studies use a temporal design as employed here, but given the way the current protocol aims to improve on existing studies, I just wonder why not go all the way.
Response: This method was developed in particular to study the relationship between statistical learning and language, which contains patterns distributed in the temporal domain. Spatial pattern learning, on the other hand, engages different neural systems (Karuza et al., 2017) and have different behavioral profiles in special populations (Roser et al., 2014). Therefore, all these tasks were designed to study learning in the temporal domain.

5. The Age ~ Accuracy correlation which is reported in the results section doesn't actually tell us much given that there are only 22 kids in total. Given the small dataset and the lack of power to conduct such an analysis, it seems inappropriate.
Response: The correlation has been removed from the current analysis.

6. The objectives of the fMRI tasks are less clear - is it just about locating relevant brain areas associated with each modality/domain? Given that there is already quite some data on this (for review see Milne, Wilson and Christiansen (2018), isn't it more relevant to use this paradigm in order to examine other things, for example, the strength of activation? Moreover, what is the proposed relationship between the neural measures and the behavioral measures? Since both measures are collected in the current paradigm, surely the authors have some predications as to

how these measure are correlated. For example, would stronger activation lead to better learning?
Response: We have provided more background in the introduction on the merits and contributions of our fMRI paradigm. A combination of neural and behavioral measures allows future brain-behavior correlation analyses for both online and off-line statistical learning. An individual's performance might be constrained by not only activation strength at certain brain regions, but also connectivity between the domain-specific sensory and domain-general learning regions. Our paradigms provide a large range of possibility for analyses including both univariate and multivariate approaches. We appreciate the reviewers' recommendation to include Milne, Wilson and Christiansen (2018), as the current manuscript can be strengthened by support from their research.

7. The instructions regarding the use of a mock scanner are written as if this is something to be done in future studies. Was this actually done here, with the children reported in the paper?
Response: A mock scanner was used with all children in the current study and we have edited the style of wording throughout the manuscript to better reflect this.

8. There is a mix-up in the description of sections 1.2.4 and 1.2.5. Since it mentions tones, section 1.2.4 seems to be describing the auditory nonlinguistic task (and not the auditory linguistic task, as it written in parenthesis), and the opposite goes for section 1.2.5.
Response: Thank you very much for identifying this writing error, which has now been corrected.

References
Karuza, E. A., Emberson, L. L., & Aslin, R. N. (2014). Combining fMRI and behavioral measures to examine the process of human learning. Neurobiology of learning and memory, 109, 193-206.
Milne, A. E., Wilson, B., & Christiansen, M. H. (2018). Structured sequence learning across sensory modalities in humans and nonhuman primates. Current opinion in behavioral sciences, 21, 39-48.
Emberson, L.L., Conway, C.M., & Christiansen, M.H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. Quarterly Journal of Experimental Psychology, 64, 1021-1040.
Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. Cognition, 177, 198-213.
Shufaniya, A., & Arnon, I. (2018). Statistical Learning Is Not Age-Invariant During Childhood: Performance Improves With Age Across Modality. Cognitive Science, 42, 3100-3115.
Response: We had previously included Emberson, Conway & Christiansen (2011), as well as Sigeleman et al. (2018); however, we appreciate the recommendation to include these other references, and have done so throughout the protocol.

Reviewer #4:
Manuscript Summary:
Dr. Schneider and colleagues present a protocol for the study of statistical learning (SL) applied to school children. The protocol is sound and offers one great advantage over other similar types

of tasks: it allows for real-time learning. This is an exciting innovation, as most of the studies assessing SL focus on the outcome (they use post learning tests). The authors provide fMRI and web versions of the task, as well as different paradigms to address linguistic and non-linguistic SL in both visual and auditory domains, which is of further significance. I recommend publication of this work pending some minor comments detailed below.

Major Concerns:
None

Minor Concerns:
1 Could the authors provide a better justification of the 24 repetitions of each of the 4 triplets for the visual tasks? Why 24? Is this usually enough for learning to occur?
Response: It seems that increased exposure does not lead to increased performance in SL tasks after a certain point (see Arnon, 2019, and Siegelman, et al. 2018 for a demonstration and discussion) and that 24 repetitions is appropriate to measure learning in this task based on our preliminary results.

2 Why double the trials for auditory stimuli than for visual? Is this just to make sure that all the tasks have a similar length in time? Is auditory SL harder than visual?
Response: The purpose to doubling the trials for auditory stimuli was to ensure that all tasks have a similar length in time, and we do not believe the auditory task was any harder based on our results.

3 How accurate is the measurement of Reaction Times via a web browser?
Response: de Leeuw, the creator of jsPsych, and Motz found the response times measured via a web browser "were approximately 25 ms longer" than those measured via MATLAB (2016). Importantly, this delay was found to be constant across trials. Because our measure of real-time learning in the web-based tasks is the slope of change in reaction time, the effects of the delay in reaction time has been minimized using within-subject comparison. de Leeuw (2015) has also acknowledged that reaction time measured via jsPsych may be affected by factors such as the processing speed of the computer or the number of tasks loaded in the background. To minimize these effects, the author has recommended using within-subject comparison for response times across trials (de Leeuw, 2015). We have added this information to the discussion.

4. For the fMRI protocol, the authors say "In contrast, random sequences contain the same 12 stimuli as presented in the structured sequences" and also "in order to maximize learning of the structured sequences, the random blocks in each run contain a different type of stimuli from the structured sequence". Do the authors mean that for the latter the difference is the domain? Visual or auditory? Please clarify"
Response: The reviewer is correct that the latter difference is related to domain. We have clarified this by stating "In order to maximize learning of the structured sequences, the random blocks in each run contain a different domain from the structured sequence (e.g. speech structured, tone random/image structured, letter random). This was counterbalanced across the four runs."