



Published in final edited form as:

*J Cogn Neurosci*. 2016 October ; 28(10): 1484–1500. doi:10.1162/jocn\_a\_00990.

## Sampling over Nonuniform Distributions: A Neural Efficiency Account of the Primacy Effect in Statistical Learning

Elisabeth A. Karuza<sup>1</sup>, Ping Li<sup>2</sup>, Daniel J. Weiss<sup>2</sup>, Federica Bulgarelli<sup>2</sup>, Benjamin D. Zinszer<sup>3</sup>, and Richard N. Aslin<sup>3</sup>

<sup>1</sup>University of Pennsylvania

<sup>2</sup>Pennsylvania State University

<sup>3</sup>University of Rochester

### Abstract

Successful knowledge acquisition requires a cognitive system that is both sensitive to statistical information and able to distinguish among multiple structures (i.e., to detect pattern shifts and form distinct representations). Extensive behavioral evidence has highlighted the importance of cues to structural change, demonstrating how, without them, learners fail to detect pattern shifts and are biased in favor of early experience. Here, we seek a neural account of the mechanism underpinning this primacy effect in learning. During fMRI scanning, adult participants were presented with two artificial languages: a familiar language (L1) on which they had been pretrained followed by a novel language (L2). The languages were composed of the same syllable inventory organized according to unique statistical structures. In the absence of cues to the transition between languages, posttest familiarity judgments revealed that learners on average more accurately segmented words from the familiar language compared with the novel one. Univariate activation and functional connectivity analyses showed that participants with the strongest learning of L1 had decreased recruitment of fronto-subcortical and posterior parietal regions, in addition to a dissociation between downstream regions and early auditory cortex. Participants with a strong new language learning capacity (i.e., higher L2 scores) showed the opposite trend. Thus, we suggest that a bias toward neural efficiency, particularly as manifested by decreased sampling from the environment, accounts for the primacy effect in learning. Potential implications of this hypothesis are discussed, including the possibility that “inefficient” learning systems may be more sensitive to structural changes in a dynamic environment.

### INTRODUCTION

Our environment is composed of a vast and uncertain number of perceptual patterns. A successful learning mechanism must detect the statistical regularities inherent in these patterns and crucially determine whether samples from the environment reflect a single common pattern or multiple separable patterns. Thus, the processes of learning and structural change detection are fundamentally intertwined. The importance of differentiating

between patterns is highlighted when considering multilingual acquisition in young children (see Weiss, Poepsel, & Gerfen, 2015; Genesee, 1989; Meisel, 1989; Lindholm & Padilla, 1978). An infant raised in a multilingual home is exposed to linguistic structures that may or may not be clearly marked as mapping onto separable languages. Particularly in the case of conflicting information, successful acquisition hinges on distinguishing between rather than combining across the statistics of multiple languages (Weiss, Gerfen, & Mitchel, 2009; Bosch & Sebastián-Gallés, 2003). Fortunately for the naive learner, the disambiguation of linguistic structures is often facilitated by external cues. These cues may be social or contextual in nature; for example, an infant may associate a specific language with a specific parent, gender (male vs. female speaker), or social context (e.g., within vs. outside the home). Alternatively, these cues might be tied to the input itself; for example, languages to which a young child is exposed might differ markedly in terms of phonemic inventory or inflectional patterns (Gervain & Werker, 2013; see Bosch & Sebastian-Galles, 2001).

In addition to making use of perceptual and contextual cues, learners can proficiently detect changes in underlying statistical representations when explicit feedback is available (e.g., in the form of reward; Behrens, Woolrich, Walton, & Rushworth, 2007; Gallistel, Mark, King, & Latham, 2001). However, when neither source of evidence for structural shifts is present, individuals tend to fall short of an ideal learner. For example, Weiss et al. (2009) demonstrated that adults could segment words from interleaved syllable streams containing competing statistics only when the speech streams were paired with an indexical voice cue corresponding to the pattern change. When the streams were produced by the same speaker, neither stream was accurately segmented. However, when the conflicting streams were produced by different speakers (male or female), adults were able to segment the two statistical patterns with similar levels of accuracy. This finding offers evidence that adults are sensitive to contextual cues to structural change, and that in the absence of such cues, learning may be significantly diminished.

Overlaid onto the challenge of detecting structural changes in the environment is a bias toward efficiency. Cognitive resources are finite—learners have limitations in processing speed, attention, and memory (e.g., Osaka, Logie, & D’Esposito, 2007; Miller, 1956). Evidence that our neural systems are geared toward efficiency is abundant. A classic case is light/dark adaptation—when overall illumination is suddenly altered, sensitivity to small changes in illumination shifts to the mean level of illumination. This adaptive gain-control mechanism is highly efficient—it enables neural firing of retinal ganglion cells to concentrate its high sensitivity around the mean illumination rather than spreading that sensitivity across several orders of magnitude of illumination (which would flatten its sensitivity profile; see Ohzawa, Sclar, & Freeman, 1985). This phenomenon, facilitated processing of or adaptation to the most common or frequent stimulus in the immediate environment, is not unique to low-level perception and has also been demonstrated in complex language comprehension tasks that measure processing behaviorally (Fine, Jaeger, Farmer, & Qian, 2013). Thus, adaptive mechanisms enable the efficient use of processing resources for predictable stimuli at both the neural and the behavioral levels. Broadly construed, neural correlates of “efficiency” could take a variety of forms ranging from the topological organization of functional networks (i.e., high levels of local clustering and shorter connections between these clusters; Sporns, 2011) to decreased recruitment of

cognitive control regions during attentionally demanding tasks (Kozasa et al., 2012). In addition to monitoring the overall activation of brain areas during language exposure, we focus here on a metric of efficiency that is likely to have specific consequences for change detection tasks: decreased sampling of the environment as a function of learning. In other words, we predict that functional connections with sensory cortex should weaken when learners encounter familiar patterns.

The brain therefore contends with competing forces: the potential benefit of monitoring for new structures in the environment and a bias toward minimizing use of its resources. Evidence of a bias toward efficiency can be found in certain types of behavioral learning tasks, in which failure to detect changes in continuous input manifests as insensitivity to downstream patterns. For example, Jungé, Scholl, and Chun (2007) demonstrated such an early bias effect on learning in a contextual cueing task. One group of participants was first exposed to a visual display where distractor shapes were not correlated with target location. When the displays were subsequently altered, such that the distractor shapes were then correlated with target location, those participants showed no facilitation effect on motor RT. In other words, initial experience with the noisy (uncorrelated) displays prevented learners from capitalizing on predictive information later in learning. In contrast, participants who were first exposed to displays where the distractors were predictive of target location later showed the expected facilitatory effects on motor RT, even after an intervening distractor phase. The authors proposed that learning does not necessarily unfold uniformly over time and that initial hypotheses about the statistics of the environment may be resistant to updating. Although it is certainly possible that learners in this case were unwilling to update an already entrenched representation of the environment, another related possibility is that entrenched representations actually induce learners to sample less from downstream input.

In addition to situations in which early input is noisy, there is evidence that early information is favored in learning situations that contain multiple patterns. Gebhart, Aslin, and Newport (2009) conducted a segmentation experiment similar to that of Weiss et al. (2009), but in which participants were presented with two languages (which we will hereafter refer to as L1 and L2) for 5.5 min each in direct succession, as opposed to interleaved in shorter blocks. When the transition to L2 was marked with a 30-sec pause, participants successfully discriminated high probability syllable triplets (“words”) from low probability triplets (“partwords”) in both languages. When the transition was unmarked and the L2 stream followed continuously from the L1 stream, participants learned only the first language. This primacy effect held even when L2 consisted of the same inventory of consonants and vowels that comprised L1 (i.e., the language streams were perceptually overlapping). In the unmarked condition, L2 learning was achieved only by tripling the amount of L2 exposure relative to L1.

Thus, the existing behavioral data and theoretical perspectives suggest that (1) learning of multiple structures in a continuous context is heavily dependent upon indexical cues (Mitchel & Weiss, 2010; Gebhart et al., 2009; Weiss et al., 2009) and (2) in the absence of such cues learning favors early experience (Gebhart et al., 2009; Jungé et al., 2007). We hypothesize that this primacy effect results from a bias toward efficiency: If learners expect an environment to be stationary, it is inefficient to sample constantly from it once they have

robust knowledge of its underlying structure. This hypothesis, if supported, could have significant implications for understanding the constraints of language acquisition and plasticity of learning.

Here, we seek to characterize through fMRI the specific mechanism by which neural efficiency impedes learning of a novel language after participants have formed a representation of a perceptually similar language. In the present experiment, we first pretrained participants on a statistically structured syllable stream. Next, we recorded brain responses as participants relistened to that syllable stream (L1), followed immediately by a novel stream consisting of the same syllables organized in a different pattern (L2). As in the unmarked condition of Gebhart et al. (2009), we removed any accompanying indexical cue. Following exposure, we evaluated learning of each language by asking participants to discriminate between high and low probability triplets from L1 and L2. We then used a combination of univariate and functional connectivity measures to examine how decreases in activation levels and the interplay between high-level association and low-level sensory-specific cortical regions might give rise to the primacy effect in learning.

Our work diverges from prior studies of learning in nonstationary environments in that we examine neural responses to prolonged (>5 min), rather than short-term (<10 sec), exposures with nonuniform patterns (Tobia, Iacovella, Davis, & Hasson, 2012; Tobia, Iacovella, & Hasson, 2012). Indeed, converging neural and behavioral evidence shows that learners are adept at change detection on such rapid time scales (Tobia, Iacovella, Davis, et al., 2012; Tobia, Iacovella, & Hasson, 2012; Huettel, Mack, & McCarthy, 2002), suggesting that early pattern changes may override the primacy effect and the expectation for stationarity observed after lengthy L1 exposure. Other work has examined neural response to statistical learning over longer time frames (e.g., approximately 2.5 min; Andric & Hasson, 2015; McNealy, Mazziotta, & Dapretto, 2006), but in those cases researchers examined a highly structured sequence relative to a less structured or random sequence. Here, we investigate the acquisition of two different but equally structured language systems.

In examining how the neural processes engaged during L1 exposure might interfere with L2 acquisition, we hypothesize that, because the brain prioritizes efficiency over change detection, it will sample less from its auditory environment as a function of the efficacy of L1 learning. We hypothesize that this diminished sampling rate should manifest itself as an uncoupling (i.e., decreased connectivity) between auditory cortex, where sensory input is initially processed, and fronto-basal ganglia regions, particularly the inferior frontal gyrus and dorsal striatum, where some evidence suggests statistical learning is mediated (Karuza et al., 2013; Turk-Browne, Scholl, Chun, & Johnson, 2009). Given prior evidence suggesting temporal lobe involvement in statistical learning, it is also possible that early perceptual cortex might dissociate from more medial areas (i.e., the hippocampus; Turk-Browne, Scholl, Johnson, & Chun, 2010; Turk-Browne et al., 2009) and superior temporal association areas (Plante et al., 2015; McNealy et al., 2006).

Overall, we predict a negative correlation between L1 learning and connectivity with auditory cortex; learners who have acquired L1 most robustly should have neural learning systems that are not only less active but also less integrated with sensory cortex. However,

this decrease in sampling should come at a cost: When interregional connections are weaker, learners should be less likely to pick up on the shift to L2 because the input to higher-level systems has been attenuated. Thus, we also predict a positive correlation between connectivity during L1 exposure and L2 learning as measured at posttest. Learners whose systems are still functionally integrated and actively monitoring the sensory input during L1 processing should be more sensitive to the new statistical information present in L2. This account is compatible with prior neuroimaging work that has demonstrated deactivation of control-related frontal and parietal areas as a function of successful language learning (Grant, Fang, & Li, 2015; Stein et al., 2009; Chee, Hon, Lee, & Soon, 2001). It also accords with recent observations that sensorimotor and fronto-cingulate networks become functionally independent during a sequence-learning task (Bassett, Yang, Wymbs, & Grafton, 2015).

## METHODS

### Participants

Thirty-five participants recruited from either the University of Rochester or the Pennsylvania State University completed a behavioral prequalification portion of this study (see Procedure section). Seventeen of them completed the subsequent scanning session. One of the 17 participants was excluded because it was discovered that she had already participated in an earlier pilot version of the study. The 16 participants included in these neuroimaging analyses were divided evenly between both institutions (fMRI scanning was conducted on identical scanners using identical imaging sequences; see below for scanning details<sup>1</sup>). The study was approved by the institutional review boards at both institutions. All participants were right-handed, native speakers of English between the ages of 18 and 30 (mean age = 20.67 years,  $SD = 1.01$  years; 6 men, 10 women) with normal or corrected-to-normal vision. They provided informed consent and were compensated financially according to university guidelines.

### Stimuli

Auditory stimuli were adapted from Gebhart et al. (2009) and Newport and Aslin (2004). Two distinct “languages” were created by concatenating 16 unique trisyllabic words three times each, resulting in a continuous 48-word stream of approximately 34 sec that was then looped to create a much longer exposure stream, interrupted by silence after each 34-sec block. Streams were generated with the constraints that the same word never appeared twice in a row and each word-final syllable was followed by only one of two possible word-initial syllables. Words in both languages had an underlying structure of *CV-CV-CV*, but in one stream the informative statistic was the transitional probability between consonants within a word ( $TP = 1.0$ ) and in the other the relevant statistic was between vowels within a word ( $TP = 1.0$ ). All other TPs, that is, between adjacent and nonadjacent syllables as well as between segments spanning word boundaries, ranged from 0.42 to 0.58. Note that the syllabic inventory in each language was identical (*pa, ku, be, po, gi, tae, ki, gu, do, da, te, bae*),

<sup>1</sup>To maximize the comparability of data, we used the same stimulus materials and presentation software/protocols. For fMRI data, we minimized cross-site scanning variability through quality assurance protocols for both sites to check field homogeneity, ghosting and geometry, and the stability of spatial and temporal signal-to-noise ratios.

although the words themselves differed (e.g., *pa-ku-te* vs. *te-da-ku*) because of the underlying consonant or vowel frame structure (Figure 1). The syllable streams were synthesized using a monotone, female voice setting in MacinTalk, ensuring that there were no informative pauses or cues to word boundaries other than the statistical regularities in the input. Syllable length ranged from 180 to 220 msec, and the mean intersyllable interval was approximately 18 msec within words/13 msec between words for the vowel frame language and 17 msec within words/19 msec between words for the consonant frame language.

A two-alternative forced-choice test following exposure to each language assessed whether participants had successfully segmented the streams. During the posttest, participants were asked to discriminate the predictable triplet sequences (words) from one of two types of partwords: a 3-1-2 partword composed of one word-final syllable and the first two syllables from the next word and a 2-3-1 partword composed of two word-final syllables and one word-initial syllable. We selected four words and four partwords from each stream and paired them exhaustively to create a list of 16 test pairs per language. Crucially, Gebhart et al. (2009) found no significant difference in posttest performance for separate groups of participants when comparing the consonant and vowel frame languages presented in isolation, suggesting that the languages are roughly equally learnable.

## Procedure

**Phase 1 Testing**—Participants were initially recruited for a behavioral testing session that took place in a mock scanner. Because of variability in learning that has been observed under fMRI scanning conditions (Karuza et al., 2013; Turk-Browne et al., 2009; McNealy et al., 2006), we included this preliminary session to ensure that participants robustly acquired L1. It is important to note that our question in this study was not which brain areas underlie learning of statistical regularities (a topic that has been addressed by the aforementioned studies); rather, we sought to uncover the mechanism by which knowledge of L1 impacts the learning of patterns embedded in L2. It was therefore crucial to establish an L1 learning effect.

Participants were positioned comfortably in the mock scanner, equipped with headphones, and instructed to view an experimental display through a rearview mirror mounted to a mock head coil. They were told that they would be exposed to an alien language, parts of which might become familiar to them. They were also informed that, following exposure, there would be a testing phase. As in the subsequent scanning phase, stimuli were presented in a blocked design (33.7 sec on, 20 sec off). While the language was playing, a black-and-white cartoon image of an “alien” (fictional creature) was presented in the center of the screen. During the off periods, a fixation cross was displayed. Auditory fade-in and fade-out effects (Audacity Team, 2012; Audacity v. 2.0) were applied to the first and last two syllables of each sound stream to eliminate potential segmentation cues. Finally, ambient scanner noise was played during the session so that the setup would match an actual scanning session as closely as possible.

Following 10 blocks of exposure (5.6 min of L1 and 3.7 min of interstimulus rest; 9.3 min total), participants were told that they would hear pairs of items with a silence between them. Within each pair, one item was a word and the other a partword. Using a handheld



button box, they selected “1” if the first triplet sounded like it belonged to the language they just heard and “2” if the second triplet belonged to the language. Participants who met or exceeded a predetermined cutoff of 11 of 16 correct (68.75%) were invited to participate in a follow-up scanning session. Participants were not informed pretesting that their performance in the mock scanner determined their eligibility for scanning nor were ineligible participants informed that they failed to meet criterion.

**Phase 2 Language Exposure**—Qualifying participants completed an fMRI scan within 48 hr of the original mock-scanner session. At each institution, half of these participants had been exposed to the consonant frame language on Day 1 and half to the vowel frame language. They were told that the session in the mock scanner was a practice session to make sure they would feel comfortable in the real scanner while they performed a longer version of that experiment. Block duration, ISI, visual presentation, and instructions were identical to the mock-scanner protocol. However, in this phase of the experiment, we extended the exposure to include 4 “transition” blocks and 5.6 additional minutes of L2 exposure (plus interleaved baseline periods). That is, in the first run, we reexposed participants to the L1 stream they heard in the mock scanner over the course of 10 blocks (either the consonant or vowel frame language), followed by two additional blocks that were 33% and 50% L2, respectively (total run duration was 11.1 min). We then started a second run (also 11.1 min) with two more transition blocks (50% and 67% L2, respectively) followed by 10 full blocks of L2 (Figure 2).<sup>2</sup> Regardless of the language playing, the same alien was present on the screen, and participants were given no indication that they would be exposed to a different language. In fact, the intended purpose of the transition blocks, along with the consistent presentation of the alien image, was to obscure cues to the L2 shift that might be induced by the pause between scanning runs. During scanning, participants were therefore led to believe that they were simply listening to more L1, although half of the Phase 2 presentation actually consisted of L2. Following exposure to both languages, participants were asked to complete a “longer version of the first posttest.” In the General Discussion section, we consider how potential consequences of explicitly leading participants to expect a single language might bear upon interpretation of our results.

No imaging data were collected during the posttest, though for continuity participants remained in the scanner during this final portion of the experiment. Unbeknownst to participants, the first 16 items of the posttest examined L2 learning whereas the second 16 items reexamined L1 learning. Thus, we could ask whether L1 learning was maintained, even after the L2 pattern shift. Moreover, if test performance were transient, we would capture any learning of L2 because it was always tested before L1. This design feature enabled us to examine learning of L2 without potential interference effects from L1 test items; however, when interpreting L1 test scores following scanning, we cannot completely rule out potential interference effects from the preceding L2 test. To be clear, any such interference effect is likely to be quite negligible because neither words nor partwords in the L2 posttest overlapped in any way with the words and partwords in the subsequent L1

---

<sup>2</sup>Transition blocks were formed by splicing together each sound stream with an intersyllable interval of 15–35 msec to eliminate coarticulatory cues to the transition. The three types of transition blocks (33%, 50%, and 67%) were structured as follows: initial 32 words L1/final 16 words L2, initial 24 words L1/final 24 words L2, initial 16 words L1/final 32 words L2.

posttest. It was not, for example, possible for learners to endorse a word during the L2 posttest and then reject that same triplet as a partword in the L1 test. Although an online, more implicit behavioral measure of learning undoubtedly has its benefits, we opted to test participants' knowledge with an offline posttest for the following reasons: (1) the nature of our exposure streams, which contained the rapid and continuous presentation of syllables, did not easily lend itself to the collection of online measures; (2) depending on the task used, the collection of online measures has unknown effects on the learning process; (3) the broader field of statistical learning offers considerable evidence that learning can be indexed by an offline posttest, as it is the most commonly used measure; and (4) online implicit measures, such as RT, have been shown to correlate with offline posttest performance (Karuza, Farmer, Smith, Fine, & Jaeger, 2014).

**Phase 2 Setup and Acquisition Parameters**—Before beginning the scanning sequences, the participant's head was secured using foam padding. As in the mock scanner, rear-projected visual stimuli appeared through a small mirror mounted above the eyes at an angle of 45°. Participants were provided with earplugs to reduce scanner noise, and they wore noise-canceling headphones to enable them to hear the auditory stimuli. Posttest responses were recorded on a custom-built MR-safe button box held in their right hand.

Imaging data acquired at both sites (Penn State and the University of Rochester) were collected on a Siemens (Erlangen, Germany) 3T MRI scanner equipped with a 12-channel head coil. At the start of the session, a high-resolution, T1-weighted anatomical image was acquired using an MPRAGE sequence (repetition time [TR] = 2530 msec, echo time = 3.44 msec, flip angle = 7°, voxel size = 1 mm<sup>3</sup>). For the subsequent two functional runs (334 time points each), 30 axial slices covering the whole brain were collected in an interleaved order using a T2\*-weighted gradient-echo EPI sequence (TR = 2000 msec, echo time = 30 msec, flip angle = 90°, voxel size = 4 mm<sup>3</sup>).

## RESULTS

### Behavioral Performance

Average L1 learning in the mock scanner was significantly greater than chance level (50%), even when including those participants who did not meet the qualification criterion (mean accuracy = 70.89% correct;  $t(34) = 5.87, p < .0001$ ). Of those 16 participants who qualified and were available for the scanning session, mean L1 performance in the mock scanner was 87.50% correct ( $t(15) = 15.94, p < .0001$ ). Bar plots in Figure 3A indicate that participants generally maintained their L1 performance (86.72% correct,  $SD = 14.22\%$ ;  $t(15) = 10.32, p < .0001$ ) following the fMRI phase of L1 and L2 exposure. However, they did not, on average, successfully acquire L2 (50.39% correct,  $SD = 16.99\%$ ;  $t(15) = 0.09, p = .93, ns$  different from chance). On the scanning day, L1 scores ranged from 50% to 100% and L2 scores ranged from 25% to 81.25% (Figure 3B). A Levene's test revealed no significant difference in variance ( $F = 1.93, p = .18$ ). Interestingly, L1 and L2 performance were significantly inversely correlated ( $r = -0.56, p = .02$ ), suggesting that robust L1 acquisition may have impeded subsequent L2 acquisition.



## Neuroimaging Results

**General Preprocessing**—All fMRI analyses were carried out using FEAT (fMRI Expert Analysis Tool, v. 6.00), a general linear modeling package that is part of FSL (FMRIB's Software Library; Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012). Two dummy volumes from the start of each functional run were discarded to eliminate start-up equilibration artifacts. The following standard corrective procedures were then performed: skull-stripping with BET to remove nonbrain material, motion correction with MCFLIRT (FMRIB's Linear Image Registration Tool; Jenkinson, Bannister, Brady, & Smith, 2002), slice timing correction (interleaved), spatial smoothing with a 5-mm 3-D Gaussian kernel, and high-pass temporal filtering with a cutoff of 55 sec. Six motion parameters estimated from MCFLIRT were added as nuisance regressors in all subsequent modeling.

**Group Level Univariate Analysis**—A series of univariate analyses were then performed to elucidate the relationship between activation levels during language exposure and ultimate learning measures. As the basis for these group level analyses, we modeled individual subject activation separately for each of the two functional runs (so-called “first level analysis” carried out using FILM: FMRIB's Improved Linear Model). The waveform corresponding to each language (L1 in the first run or L2 in the second run) was modeled by specifying the onset and duration of each of the 10 language blocks. The two transition blocks (mixture of L1 and L2) in each run were coded as separate regressors of no interest. This waveform then underwent double-gamma convolution to best match it to the measured hemodynamic response function. To reduce unexplained noise, we also added in a fraction of the temporal derivative from the original waveform and applied a temporal filtering process. Finally, native image transformation to the standard MNI-152 structural template (2 mm) was completed using FLIRT (Jenkinson et al., 2002): Participants' functional scans were first coregistered to their corresponding structural scan, and the combined maps were warped to the MNI template via an affine transformation with 12 degrees of freedom.

We began by examining the effect of posttest score on each run separately.<sup>3</sup> First-level parameter estimates from either L1 or L2 were entered into FMRIB's Local Analysis of Mixed Effects (FLAME1). In addition to modeling the main effect of task in each run, we included a demeaned predictor corresponding to L1 or L2 posttest score. As FLAME1 is one of the more conservative cluster-based parametric methods (Eklund, Nichols, & Knutsson, 2015), all group level maps were thresholded at the single voxel level using a cutoff of  $Z > 1.96$ . However, because of concerns about inflated family-wise error rates for certain parametric statistical methods (Eklund et al., 2015), we also report which results were maintained when a more stringent cluster defining threshold of  $Z > 2.6$  was applied (corresponding to a one-tailed  $p < .005$ ). Finally, according to the cluster correction method

<sup>3</sup>In an analysis suggested by a reviewer, we also examined, through an additional first-level regressor, whether parametric increases in activation might predict L1 and L2 posttest scores. We find that for L1, linear decreases in activity across exposure blocks were correlated with L1 accuracy in two fronto-temporal clusters (Cluster 1: extent = 1886 voxels,  $p = .0003$ ,  $Z$ -max of 3.33 in right paracingulate gyrus  $x = 2$ ,  $y = 10$ ,  $z = 52$ ; Cluster 2: extent = 2731 voxels,  $p < .0001$ ,  $Z$ -max of 3.31 in right central operculum  $x = 64$ ,  $y = -12$ ,  $z = 12$ ;  $ns$  at  $Z > 2.6$ ) and a subcortical cluster (extent = 1320 voxels,  $p = .004$ ,  $Z$ -max of 2.97 in right caudate  $x = 14$ ,  $y = 116$ ,  $z = -4$ ;  $ns$  at  $Z > 2.6$ ). Conversely, linear increases in activity during L2 were correlated with L2 accuracy in a single bilateral fronto-temporal cluster (cluster extent = 4301 voxels,  $p < .0001$ ,  $Z$ -max of 3.5 in left superior frontal gyrus  $x = -18$ ,  $y = 18$ ,  $z = 42$ ;  $ns$  at  $Z > 2.6$ ).

described in Worsley (2001), we compared significance levels associated with contiguous groupings of these remaining voxels to a cluster probability threshold of  $p < .05$ .

Although we find no significant effect of L2 posttest performance on L2 activation levels (at either  $Z$ -statistic threshold), we observe a widespread set of areas exhibiting a significant negative correlation between L1 posttest scores and L1 activation. This diffuse negative association encompasses bilateral temporal cortex, posterior parietal cortex, as well as fronto-subcortical regions (Table 1). These results suggest that learners with higher L1 posttest scores appeared to allocate fewer processing resources during L1 exposure, presumably because they had already robustly learned L1 in the prior mock-scanner session. Noting that this inverse relationship was quite robust, we also display in Table 1 cluster information when the analysis was run using the more conservative threshold of  $Z > 2.6$ . In Figure 4, we display all regions significantly active above and below baseline for L1 and L2 regardless of behavioral performance (i.e., corresponding to the main effect of task at  $Z > 1.96$  and  $Z > 2.6$  in the L1 and L2 runs; see also Table 2).

Next, we contrasted activation between the runs directly, asking whether, across participants, activation differences in response to L1 and L2 exposure were modulated by the strength of the primacy effect measured behaviorally. To accomplish this, all parameter estimates from first-level analyses of L1 and L2 were entered into a FLAME model; our group level model contained a fixed categorical factor with two levels corresponding to each of the two language exposure runs (L1 or L2), a numeric primacy predictor for each participant (L1–L2 scores centered with respect to the group mean), and random subject intercepts. The primacy effect predictor was computed by subtracting accuracy scores from the L2 posttest from accuracy scores from the L1 posttest (i.e., L1–L2 performance, where a positive value indicates a particular participant scored higher on L1 than L2). First, we observe a significant main effect of language (cluster-corrected at a threshold of  $p < .05$  and  $Z > 1.96$ ; *ns* at  $Z > 2.6$ ). When comparing L1 and L2, we find a medial fronto-subcortical cluster with a peak in right fronto-orbital cortex (cluster extent = 1492 voxels,  $p = .0004$ , peak of  $Z = 3.55$  in  $x = 16$ ,  $y = 28$ ,  $z = -14$ ) and a posterior parietal cluster with a peak in right precuneus (cluster extent = 917 voxels,  $p = .01$ , peak of  $Z = 3.51$  in  $x = 8$ ,  $y = -56$ ,  $z = 64$ ) to be more strongly activated in the L2 exposure run. Thus, participants on average showed greater activation in these two clusters during the novel language (L2) when compared with the familiar language (L1), and no brain areas displayed the reverse effect. Second, our results reveal a significant negative interaction between these predictors within posterior parietal cortex (Figure 5A; *ns* at  $Z > 2.6$ ). We find an approximately medial cluster (extent = 1960 voxels,  $p < .0001$ ) with maxima in right and left precuneus (right:  $Z = 3.25$ ,  $x = 14$ ,  $y = -64$ ,  $z = 28$ ; left:  $Z = 3.14$ ,  $x = -10$ ,  $y = -74$ ,  $z = 38$ ) and extending into right posterior cingulate (PCC) gyrus (peak in  $x = 2$ ,  $y = -46$ ,  $z = 26$ ,  $Z = 2.90$ ). To illustrate this effect, we have extracted for the precuneal and PCC activation peaks an estimate of the L1–L2 activation difference for each participant and plotted it against their behavioral difference scores (Figure 5B). Note that when L1 learning was higher than L2 learning, participants showed lower levels of posterior parietal activation during L1 exposure relative to L2 exposure. Conversely, those participants whose L2 learning exceeded L1 learning showed greater posterior parietal activation during L1 relative to L2 during exposure.

**Group Level Connectivity Analysis**—Subsequently, we investigated the neural connectivity co-occurring with the observed activation decreases during L1 processing. That is, we sought to characterize through a psychophysiological interaction (PPI) analysis the interregional functional integration that accompanies changes in mean activation levels (O’Reilly, Woolrich, Behrens, Smith, & Johansen-Berg, 2012). This temporally fine-grained approach may be of particular importance in light of findings that, during learning, early perceptual areas respond on a much shorter time scale than higher-level areas such as frontal cortex (Harrison, Bestmann, Rosa, Penny, & Green, 2011). If the brain indeed samples less from its environment as a function of learning, then we would expect a corresponding attenuation of the interaction between auditory cortex and downstream cortical and subcortical regions that mediate statistical learning. PPI analyses thus enable us to reveal regions that become more or less tightly coupled during exposure to L1. This method contrasts with a pure correlational analytic approach that explores which regions are integrated in general, regardless of the timing of the input. To differentiate between the PPI approach used here and correlational approaches commonly associated with functional connectivity, we use the term “interactivity” below when referring to the association between time series in auditory cortex and the rest of the brain.

Because we were specifically interested in interactivity with perceptual regions, we defined our seed ROI from the activation peak in left temporal cortex determined at the group level via a simple language versus baseline contrast (i.e., activation corresponding to a main effect of task in L1). This task-general peak for L1 ( $x = -42, y = -30, z = 8$ ) was localized to primary auditory cortex, falling near the junction of left planum temporale and Heschl’s gyrus. To perform the PPI analysis, we first transformed this peak voxel to each individual’s native functional space through their unique structural images, then dilated a spherical region around that voxel using a 5-mm kernel. From the resulting 7-voxel seed, we extracted each participant’s mean time series from their preprocessed (i.e., filtered and motion-corrected) functional images for L1 (see General Preprocessing section). Next, we reran the previously described first-level analysis, but this time including the following regressors: (1) a psychological regressor indicating the timing of the L1 task blocks (this regressor was centered such that the zero point fell halfway between the task blocks [+1] and the baseline periods [-1]); (2) a *physiological* regressor consisting of the activation time course of our single auditory seed sphere (this regressor was centered by subtracting the mean intensity across the time series from the intensity value at each TR); (3) an interaction regressor modeling the relationship between our block timing waveform and the physiological time series; and (4) a nuisance regressor specifying the timing of the mixed blocks. Of the various approaches to PPI analyses (cf. Gitelman, Penny, Ashburner, & Friston, 2003), we adhere to the approach detailed by O’Reilly et al. (2012). That is, our task regressor was first convolved with a double-gamma hemodynamic response function and subsequently combined with the filtered time series from our left temporal seed region. Neither the physiological regressor nor the interaction term required convolution or temporal filtering, as they already represented the real-time state of the brain during scanning.

To investigate how interactivity during L1 experience relates to ultimate learning outcomes (and therefore gives rise to the observed behavioral primacy effect), we then performed two

group-level analyses. In the first analysis, we entered first-level parameter estimates of the PPI effect from L1 into FLAME and asked where L1 posttest performance (demeaned L1 posttest scores) significantly predicted L1 interactivity with our auditory seed. In a second, related analysis, we examined the relationship between L1 interactivity and L2 learning by inputting L2 posttest scores. Thus, we tested whether infrequent L1 sampling, which we operationalize as low interactivity with auditory cortex during L1 processing, might result in lower L2 learning outcomes. All significant interactivity results reported below were obtained using a threshold of  $Z > 1.96$  (i.e.,  $ns$  at  $Z > 2.6$ ).

Results from the first analysis show a complementary relationship between L1 activation and L1 interactivity, at least in the context of robust learning. As Figure 6A demonstrates, we found a significant negative correlation between L1 learning and L1 interactivity with auditory cortex in diffuse bilateral frontal and subcortical regions (cluster extent = 8001 voxels,  $p < .0001$ ), including left inferior frontal gyrus (peak of  $Z = 2.52$  in  $x = -50$ ,  $y = 12$ ,  $z = 10$ ), left caudate (peak of 2.81 in  $x = -10$ ,  $y = 12$ ,  $z = 8$ ), left putamen (peak of 2.66 in  $x = -22$ ,  $y = 0$ ,  $z = 8$ ), right caudate (peak of 2.53 in  $x = 8$ ,  $y = 6$ ,  $z = 4$ ), and right putamen (peak of 2.64 in  $x = 26$ ,  $y = 2$ ,  $z = 8$ ). In Figure 6B, we show how these five peaks, selected for visualization because of their implication in other fMRI studies of statistical learning (e.g., Karuza et al., 2013; Turk-Browne et al., 2009), are less functionally integrated during L1 processing in strong L1 learners. This pattern does not hold when comparing L1 interactivity estimates in these five regions with L2 learning, although we do see a slight positive trend (Figure 6B). Given the inverse correlation between L1 and L2 posttest scores, note that it would be impossible for those regions that are negatively correlated with L1 learning to also negatively correlate with L2 learning.

For the second group level analysis, in which L2 posttest scores were used as a predictor of L1 interactivity, we do observe a positive relationship between these measures in a different set of areas. Specifically, we find a bilateral frontoparietal cluster (extent = 6026 voxels,  $p < .0001$ ), with peak parietal activation in left precuneus ( $Z = 3.33$ ,  $x = -2$ ,  $y = -40$ ,  $z = 54$ ) and right precuneus ( $Z = 3.33$ ,  $x = 2$ ,  $y = -40$ ,  $z = 52$ ), and peak frontal activation in left anterior cingulate gyrus ( $Z = 3.11$ ,  $x = -4$ ,  $y = 6$ ,  $z = 36$ ). This finding represents the only significantly positive relationship we observed thus far; the greater the interactivity between early auditory cortex and this frontoparietal system during L1 processing, the higher the L2 learning scores. Participants who were still integrating auditory input with high-level cognitive systems during L1 (i.e., sampling more from their environment) were more likely to be affected by novel statistical information contained in L2 streams.

**Investigating Complementary Systems**—In a final analysis, we asked whether the positive relationship between L1 interactivity and L2 learning would be mirrored in the relationship between L1 activation levels and L2 behavioral performance. We ran an identical correlational analysis, but input L1 first-level parameter estimates for the main effect of task instead of the PPI estimates. We find that L1 activation and interactivity are complementary in posterior parietal cortex. When analyzing the effect of L2 posttest scores on mean activation during L1 at  $Z > 1.96$ , we observe two significant clusters with peaks in the right superior parietal lobule (cluster extent = 3136 voxels,  $p < .0001$ , peak of  $Z = 3.90$  in  $x = 10$ ,  $y = -48$ ,  $z = 74$ ), and right PCC (cluster extent = 845 voxels,  $p = .04$ , peak of  $Z =$

3.40 in  $x = 4$ ,  $y = -48$ ,  $z = 10$ ). At  $Z > 2.6$ , this L2 learning-related activation was reduced to a single cluster with a peak in right postcentral gyrus (cluster extent = 310 voxels,  $p = .03$ , peak of  $Z = 3.56$  in  $x = 42$ ,  $y = -32$ ,  $z = 64$ ). For comparison purposes, Figure 7 illustrates the overlap between L1 activation predicted by L2 posttest and L1 interactivity predicted by L2 posttest. In summary, both functional activation and integration during processing of an initial language are significantly related to learning of an upcoming language. Learners who expended more neural resources during L1 processing were more sensitive to a new set of distributional statistics; learners with the most efficient neural systems (decreased activation and interactivity levels—an “entrenchment” effect) had the lowest scores on the L2 posttest. As an additional insight, these analyses, both of which include L2 posttest performance as a behavioral regressor, demonstrate that our hypotheses about sampling and expenditure of neural resources are supported not only by L1 response correlated with L1 learning (test items which may be prone to interference effects) but also by L1 response correlated with L2 learning (which, because of the nature of the L2–L1 test order following scanning, cannot be explained by test interference effects).

## GENERAL DISCUSSION

Results of this study offer new insight into the complex interrelationships among learning, structural change detection, and neural efficiency. Both fronto-subcortical interactivity with early auditory cortex and diffuse activation within multiple association systems (i.e., downstream processing areas outside early sensory areas) were negatively correlated with L1 learning. We find that participants who expended the fewest neural resources in processing the first of two successive speech patterns ultimately displayed the most robust offline knowledge of the initially presented structure (i.e., were most accurate in discriminating between L1 words and partwords on posttest). Thus, we have used functional neuroimaging to expand on a behavioral effect that has been clear since the earliest serial RT studies (Nissen & Bullemer, 1987): Humans are more efficient (i.e., faster) in processing sequential stimuli when they have knowledge of its underlying structure (see also Karuza et al., 2014). However, at least at the neural level, facilitatory processing effects appear to involve a trade-off. We find a markedly different relationship between L1 processing and L2 learning; participants with stronger activation and interactivity between auditory and parietal regions during L1 exposure, particularly in the precuneus and PCC, attained the highest scores on the L2 posttest. Thus, we propose that sampling less from sensory input and minimizing the engagement of high-level neural systems, while potentially beneficial from an efficiency standpoint, may impede change detection mechanisms that lead to the generation of additional structural representations. On the other hand, resource-intensive processing and continuous monitoring of the environment (as evidenced by richer functional connections with primary sensory cortex) may increase sensitivity to and thereby spur stronger hypotheses about new distributions of statistical information.

### Efficiency in a Learning Context

On the surface, the efficiency/change detection trade-off appears to contrast with certain accounts of adaptation: It is in part because the sensory system depresses its response to unvaried input that it is better able to detect changes in the environment (Puccini, Sanchez-

Vives, & Compte, 2006; Ulanovsky, Las, & Nelken, 2003; Muller, Metha, Krauskopf, & Lennie, 1999; for a review, see Malmierca, Sanchez-Vives, Escera, & Bendixen, 2014). By this account, however, it is relevant to note that repetition–suppression effects are typically, though not always, measured within the confines of early perceptual cortex (cf. priming effects during sentence processing observed in higher-level temporal and frontal areas; Devauchelle, Oppenheim, Rizzi, Dehaene, & Pallier, 2009; Hasson, Nusbaum, & Small, 2006). Here, our activation and functional connectivity measures revealed differences between L1 and L2 extending beyond primary auditory areas (i.e., Heschl’s gyrus). Thus, although sensory cortex may have been sensitive to changes in syllable patterns, our data suggest that this information generally failed to be threaded upward to association regions that would induce the formation of additional structural representations. Moreover, we stress that perceptual adaptation studies typically examine sensory neurons already tuned to key properties of the visual and auditory world (i.e., meaningful variations in basic features of the environment such as the luminance, shape, and orientation of objects or the duration, frequency, and amplitude of sounds). Likewise, in the domain of linguistic processing, higher-level adaptation effects measured behaviorally (Fine et al., 2013) and neurally (Devauchelle et al., 2009; Hasson et al., 2006) are observed in participants with years of experience with a mature language system.

While we pretrained participants on L1 in this study, we still examined change detection within a dual-structure learning context, one in which structural representations of familiar input were less mature and the languages shared an identical syllable inventory. In this particular context, our results suggest that learners favor early statistical information because their neural systems are biased toward efficient processing. This bias is linked to the assumption, supported in this case by task instructions, of a single underlying statistical structure. The strongest L1 learners were those who had the most robust expectations about L1 structure and saw those expectations supported as the L1 exposure run progressed. When a second, perceptually identical, statistical structure was then presented without indexical cues, this efficiency bias then hampered the accumulation or consideration of evidence about the presence of alternative structures that differed from those initially encountered. Of course, subsequent learning of new structures is not blocked forever, and as revealed by Gebhart et al. (2009), L2 may “break through” this primacy effect with extensive additional exposure. In this study, we focused on the early phase of L2 exposure when the primacy effect was generally still robust.

In the current set of results, the question of why some participants appeared to overwrite a previously learned language in favor of downstream input (as evidenced by the inverse L1/L2 correlation) remains open. The PARSER model proposed by Perruchet and Vinter (1998) and empirically supported by Perruchet, Poulin-Charronnat, Tillmann, and Peereman (2014) offers one possible explanation. This account posits that, in the case of syllable overlap, a well-formed representation of a previously learned “chunk” will cause learners to mis-segment a second language (particularly when an L1 word transitions to an L2 partword). By this account, learners who most strongly acquired the word structure of L1 would most egregiously mis-segment L2, at least initially. On the other hand, the observed negative relationship between L1 and L2 learning is contrary to evidence collected by Franco, Cleeremans, and Destrebecqz (2011) that the weakest L1 learners tended to also be



the weakest L2 learners. Despite their dissimilar conclusions, both Perruchet et al. (2014) and Franco et al. (2011) investigated learning only when participants were explicitly cued to the presence of two languages (i.e., in instruction, voicing, or a combination of the two). In the present case, learners were informed only of a single language, suggesting that the negative correlation between L1 and L2 knowledge may be influenced by learners' expectations about the number of structures in their environment. Moreover, we note that a direct comparison between studies is not possible because our screening procedures ensured that all scanned participants were originally robust L1 learners.

### Brain Basis of the Primacy Effect

In addition to probing the general activation and interactivity patterns corresponding to our observed behavioral outcomes, it is useful to consider the brain regions in which these effects functionally arise. Focusing on the inverse relationship between neural activation during L1 exposure and L1 learning outcomes, we have implicated a diffuse set of prefrontal, superior temporal, and subcortical regions that, in other studies of statistical learning, have been positively associated with behavioral performance (Karuza et al., 2013; Cunillera et al., 2009; Turk-Browne et al., 2009; McNealy et al., 2006). In the current experimental context, we suggest that the inverse relationship between activation in these areas and behavioral performance can be attributed to key differences in the particular stage of learning captured during scanning. In typical fMRI studies of statistical learning, participants are first exposed to novel input while in the scanner, resulting in heavy initial recruitment of learning systems (indicated by a boost in neural activation related to posttest performance). Here, we screen participants based on L1 learning scores from a mock scanner session and confirm that our participants were already familiar with L1. As a result, our imaging data are likely more representative of processes related to recognition, maintenance, or consolidation, and it follows that the strongest learners would engage learning systems less than they would have in the first stages of acquisition (including functionally uncoupling parts of these systems from sensory areas). Importantly, lowered activation of frontal structures (e.g., left inferior frontal gyrus) and other control-related regions has been shown to result from extensive training on various tasks, including foreign language learning (e.g., Grant et al., 2015; Yang, Gates, Molenaar, & Li, 2014; Ventura-Campos et al., 2013; Stein et al., 2009; Chee et al., 2001) and motor sequence learning (for a review, see Patel, Spreng, & Turner, 2013). Previous studies of learning-related changes in functional connectivity offer complementary insight; specifically, integration within cognitive control networks has been shown to progressively decrease during novel word or phonological learning, presumably because of less demanding, more automatic processing (Ghazi Saidi et al., 2013). Although we instead focus on interactivity with sensory cortex, results of our functional connectivity analyses show reduced coupling with parallel areas, such as the pFC, but also encompassing the basal ganglia.

In addition to the fronto-subcortical substrates commonly observed in statistical learning tasks, our analyses also indicate a unique role of posterior parietal cortex. Interestingly, we show activation within this region to be negatively correlated with the strength of individual learners' primacy effect, or L1–L2 scores (Figure 5). Examining each posttest separately, we observe that stronger L1 learners exhibited reduced L1 activation in PCC and precuneus,

whereas stronger L2 learners exhibited the opposite effect. Tobia, Iacovella, and Hasson (2012) also found increased PCC activation in response to unstable input patterns on a far more rapid time scale (see also Speer, Zacks, & Reynolds, 2007; Zacks et al., 2001). Our data complement this finding, showing that participants with strong activation and interactivity with posterior parietal cortex during L1 are more sensitive to new structures in the learning environment (i.e., better learning of L2; Figure 7).

Although the role of PCC is not yet fully understood, accumulating evidence suggests that it is a densely connected, hypermetabolic associative region that acts as a central hub in the default mode network (Greicius, Supekar, Menon, & Dougherty, 2009). As such, it has been called (alongside its sister area, the precuneus) a task-negative region, exhibiting greater levels of activation when the brain is in self-referential, endogenously generated mental states relative to when it is engaged in exogenously driven, cognitively demanding tasks (Fox et al., 2005). Of particular relevance to the current findings, Pearson, Hayden, Raghavachari, and Platt (2009; see also Pearson, Heilbronner, Barack, Hayden, & Platt, 2011) have framed PCC as flexibly guiding behavior in dynamic environments. Using reward-based learning tasks, they draw a distinction between exploratory and exploitative behavior and show that neurons in PCC distinguish between these two strategies. Pearson et al. (2011) propose that a primary role of the PCC, as a crucial node in the DMN, is in monitoring and detecting change in a shifting world. They note, as we do here, that this region is suppressed in contexts where stimuli are well learned and no longer demand high levels of cognitive control.

Extending Pearson et al.'s proposal to implicit probabilistic learning tasks lacking explicit reward, we offer preliminary evidence that "inefficient" brains with more active association regions (specifically PCC and precuneus) are less anchored to early input and more sensitive to downstream statistical information (i.e., the shift to L2). These results find additional support from a recent behavioral study demonstrating that when learners are advanced to L2 immediately after acquiring L1, they are less likely to exhibit a primacy effect relative to learners who received additional L1 input after successful L1 acquisition before encountering L2 (Bulgarelli & Weiss, 2016). In the context of this study, neural patterns from these "inefficient" L1 learners contrast with those "efficient" learners in whom decreased activation in posterior parietal cortex during early exposure is associated with lower L2, but higher L1, learning outcomes, a result not altogether surprising given the inverse correlation between L1 and L2 posttest scores.

### Methodological Considerations

In considering other possible interpretations of the observed data patterns, we acknowledge that the task instructions in this study were specifically constructed so as to mask the transition to a new language (i.e., participants were told before scanning that they would hear "more of the language from the mock scanner session"). Although these instructions may have biased learners against L2, we suggest that they cannot fully account for findings central to our efficiency account; particularly, the inverse relationship between L1 posttest scores and L1 neural engagement (activation and interactivity with primary auditory cortex). In other words, explicit expectation of a single, familiar language, which was equated in

instructions across participants, does not explain why learners who most strongly disengaged neural systems during L1 achieved higher L1 posttest scores, and vice versa. The sum of these results points toward broader trade-offs in deploying processing resources, latching on to early structures, and detecting shifts in the environment. Moreover, Gebhart et al. (2009), who less explicitly biased learners in favor of a single language, found that L2 acquisition could be induced when exposure was increased threefold.<sup>4</sup> If mere mention of a single language were always sufficient to block L2 learning, then we would not expect learners to acquire an uncued language under any circumstances.

## Conclusions and Future Directions

Taken together, our combined univariate activation and functional interactivity data offer a neural account of the mechanisms underlying statistical learning in the face of multiple structures: We propose that the behavioral primacy effect stems, at the neural level, from the tension between processing efficiency and the potential benefit of change detection. However, the present data cannot address whether differences in task instruction, such as stressing the number of languages, and individual biases toward processing efficiency may have interactive effects on ultimate learning outcomes (i.e., perhaps “efficient” L1 processors are more likely to assume a single structure and therefore more likely to weight experimenter instructions to this effect). An investigation aimed at teasing apart these influences would be one valuable area of future research, as it remains an open question whether learners explicitly instructed to monitor for change or, given no instructions at all, would exhibit the same patterns of neural activity.

## Acknowledgments

This research was supported by an NIH grant to R. N. A. (HD-037082), an NSF grant to P. L. (BCS-1349110), an NIH grant to D. J. W. (HD-067250), and NSF GRFs to E. A. K. and F. B. We are grateful to Alex Teghipco for assistance in data collection at the University of Rochester and to Uri Hasson for helpful comments on this work.

## References

- Andric M, Hasson U. Global features of functional brain networks change with contextual disorder. *Neuroimage*. 2015; 117:103–113. [PubMed: 25988223]
- Audacity Team. Audacity (R): Free audio editor and recorder. 2012. Version 2.0.0 retrieved from <http://audacity.sourceforge.net>
- Bassett DS, Yang M, Wymbs NF, Grafton ST. Learning-induced autonomy of sensorimotor systems. *Nature Neuroscience*. 2015; 18:744–751. [PubMed: 25849989]
- Behrens TE, Woolrich MW, Walton ME, Rushworth MF. Learning the value of information in an uncertain world. *Nature Neuroscience*. 2007; 10:1214–1221. [PubMed: 17676057]
- Bosch L, Sebastián-Gallés N. Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*. 2001; 2:29–49.
- Bosch L, Sebastián-Gallés N. Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*. 2003; 46:217–243. [PubMed: 14748445]

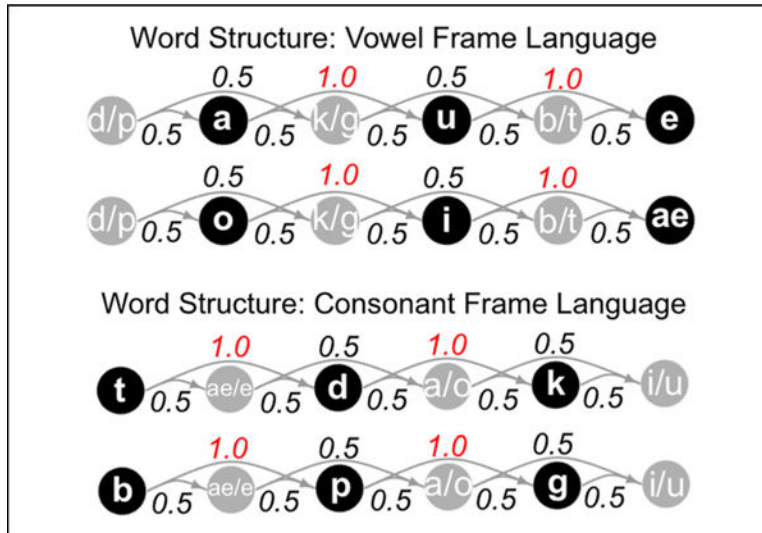
<sup>4</sup>In the case of dual-language exposure, Gebhart et al. instructed participants: “to listen attentively to a recording of a continuous speech stream that would sound a little like *a foreign language*” (p. 1095, emphasis ours).

- Bulgarelli F, Weiss DJ. Anchors aweigh: The impact of overlearning on entrenchment effects in statistical learning. *Journal of Experimental Psychology Learning, Memory, and Cognition*. 2016; doi: 10.1037/xlm0000263
- Chee MWL, Hon N, Lee HW, Soon CS. Relative language proficiency modulates BOLD signal change when bilinguals perform semantic judgments. *Neuroimage*. 2001; 13:1155–1163. [PubMed: 11352621]
- Cunillera T, Camara E, Toro JM, Marco-Pallares J, Sebastian-Galles N, Ortiz H, et al. Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage*. 2009; 48:541–553. [PubMed: 19580874]
- Devauchelle AD, Oppenheim C, Rizzi L, Dehaene S, Pallier C. Sentence syntax and content in the human temporal lobe: An fMRI adaptation study in auditory and visual modalities. *Journal of Cognitive Neuroscience*. 2009; 21:1000–1012. [PubMed: 18702594]
- Eklund, A., Nichols, T., Knutsson, H. Can parametric statistical methods be trusted for fMRI based group studies?. 2015. Retrieved from [arxiv.org/abs/1511.01863](https://arxiv.org/abs/1511.01863)
- Fine AB, Jaeger TF, Farmer TA, Qian T. Rapid expectation adaptation during syntactic comprehension. *PLoS One*. 2013; 8:e77661. [PubMed: 24204909]
- Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences, USA*. 2005; 102:9673–9678.
- Franco A, Cleeremans A, Destrebecqz A. Statistical learning of two artificial languages presented successively: How conscious? *Frontiers in Psychology*. 2011; 229:1–12.
- Gallistel CR, Mark TA, King AP, Latham PE. The rat approximates an ideal detector of changes in rates of reward: Implications for the law of effect. *Journal of Experimental Psychology: Animal Behavior Processes*. 2001; 27:354–372. [PubMed: 11676086]
- Gebhart AL, Aslin RN, Newport EL. Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science*. 2009; 33:1087–1116. [PubMed: 20574548]
- Genesee F. Early bilingual development: One language or two? *Journal of Child Language*. 1989; 16:161–179. [PubMed: 2647777]
- Gervain J, Werker JF. Prosody cues word order in 7-month-old bilingual infants. *Nature Communications*. 2013; 4:1490.
- Ghazi Saidi L, Perlberg V, Marrelec G, Pelegrini-Issac M, Benali H, Ansaldo AI. Functional connectivity changes in second language vocabulary learning. *Brain and Language*. 2013; 124:56–65. [PubMed: 23274799]
- Gitelman DR, Penny WD, Ashburner J, Friston KJ. Modeling regional and psychophysiological interactions in fMRI: The importance of hemodynamic deconvolution. *Neuroimage*. 2003; 19:200–207. [PubMed: 12781739]
- Grant A, Fang S, Li P. Second language lexical development and cognitive control: A longitudinal fMRI study. *Brain and Language*. 2015; 144:35–47. [PubMed: 25899988]
- Greicius MD, Supekar K, Menon V, Dougherty RF. Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral Cortex*. 2009; 19:72–78. [PubMed: 18403396]
- Harrison LM, Bestmann S, Rosa MJ, Penny W, Green GGR. Time scales of representation in the human brain: Weighing past information to predict future events. *Frontiers in Human Neuroscience*. 2011; 5:1–8. [PubMed: 21283556]
- Hasson U, Nusbaum HC, Small SL. Repetition suppression for spoken sentences and the effect of task demands. *Journal of Cognitive Neuroscience*. 2006; 18:2013–2029. [PubMed: 17129188]
- Huettel SA, Mack PB, McCarthy G. Perceiving patterns in random series: Dynamic processing of sequence in prefrontal cortex. *Nature Neuroscience*. 2002; 5:485–490. [PubMed: 11941373]
- Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*. 2002; 17:825–841. [PubMed: 12377157]
- Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *Neuroimage*. 2012; 62:782–790. [PubMed: 21979382]

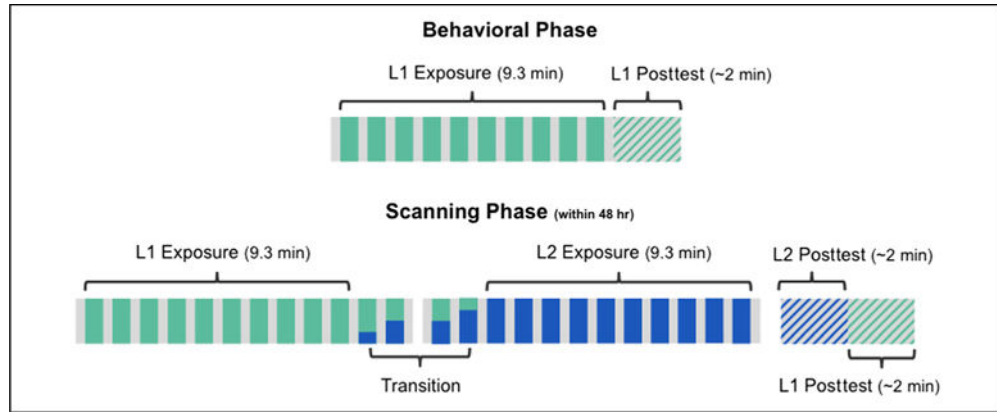
- Jungé JA, Scholl BJ, Chun MM. How is spatial context learning integrated over signal versus noise? A primacy effect in contextual cueing. *Visual Cognition*. 2007; 15:1–11. [PubMed: 18725966]
- Karuza E, Farmer TA, Smith FX, Fine AB, Jaeger TF. On-line measures of prediction in a self-paced statistical learning task. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*. 2014:725–730.
- Karuza, EA., Newport, EL., Aslin, RN., Starling, SJ., Tivarus, ME., Bavelier, D. *Brain and Language*. Vol. 127. Austin, TX: Cognitive Science Society; 2013. Neural correlates of statistical learning in a word segmentation task: An fMRI study; p. 46-54.
- Kozasa EH, Sato JR, Lacerda SS, Barrios MA, Radvány J, Russell TA, et al. Meditation training increases brain efficiency in an attention task. *Neuroimage*. 2012; 59:745–749. [PubMed: 21763432]
- Lindholm KJ, Padilla AM. Language mixing in bilingual children. *Journal of Child Language*. 1978; 5:327–335.
- Malmierca MS, Sanchez-Vives M, Escera C, Bendixen A. Neuronal adaptation, novelty detection and predictive coding in audition. *Frontiers in Systems Neuroscience*. 2014; 8:111. [PubMed: 25009474]
- McNealy K, Mazziotta JC, Dapretto M. Cracking the language code: Neural mechanisms underlying speech parsing. *Journal of Neuroscience*. 2006; 26:7629–7639. [PubMed: 16855090]
- Meisel, JM. Early differentiation of languages in bilingual children. In: Hyltenstam, K., Obler, L., editors. *Bilingualism across the lifespan: Aspects of acquisition, maturity, and loss*. Cambridge: Cambridge University Press; 1989. p. 13-40.
- Miller GA. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*. 1956; 63:81–97. [PubMed: 13310704]
- Mitchel A, Weiss DJ. What's in a face? Visual contributions to speech segmentation. *Language and Cognitive Processes*. 2010; 25:456–482.
- Muller JR, Metha AB, Krauskopf J, Lennie P. Rapid adaptation in visual cortex to the structure of images. *Science*. 1999; 285:1405–1408. [PubMed: 10464100]
- Newport EL, Aslin RN. Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*. 2004; 48:127–162. [PubMed: 14732409]
- Nissen MJ, Bullemer P. Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*. 1987; 19:1–32.
- Ohzawa I, Sclar G, Freeman RD. Contrast gain control in the cat's visual system. *Journal of Neurophysiology*. 1985; 54:651–667. [PubMed: 4045542]
- O'Reilly JX, Woolrich MW, Behrens TEJ, Smith SM, Johansen-Berg H. Tools of the trade: Psychophysiological interactions and functional connectivity. *Social Cognitive and Affective Neuroscience*. 2012; 7:604–609. [PubMed: 22569188]
- Osaka, N., Logie, RH., D'Esposito, M. *The cognitive neuroscience of working memory*. New York: Oxford University Press; 2007.
- Patel R, Spreng RN, Turner GR. Functional brain changes following cognitive and motor skills training: A quantitative meta-analysis. *Neurorehabilitation and Neural Repair*. 2013; 27:187–199. [PubMed: 23093519]
- Pearson JM, Hayden BY, Raghavachari S, Platt ML. Neurons in posterior cingulate cortex signal exploratory decisions in a dynamic multioption choice task. *Current Biology*. 2009; 19:1532–1537. [PubMed: 19733074]
- Pearson JM, Heilbronner SR, Barack DL, Hayden BY, Platt ML. Posterior cingulate cortex: Adapting behavior to a changing world. *Trends in Cognitive Science*. 2011; 15:143–151.
- Perruchet P, Poulin-Charronnat B, Tillmann B, Peereman R. New evidence for chunk-based models in word segmentation. *Acta Psychologica*. 2014; 149:1–8. [PubMed: 24632521]
- Perruchet P, Vinter A. PARSER: A model for word segmentation. *Journal of Memory and Language*. 1998; 39:246–263.
- Plante E, Patterson D, Gómez R, Almryde KR, White MG, Asbjørnsen AE. The nature of the language input affects brain activation during learning from a natural language. *Journal of Neurolinguistics*. 2015; 36:17–34. [PubMed: 26257471]

- Puccini GD, Sanchez-Vives MV, Compte A. Selective detection of abrupt input changes by integration of spike-frequency adaptation and synaptic depression in a computational network model. *Journal of Physiology*. 2006; 100:1–15. [PubMed: 17095200]
- Speer NK, Zacks JM, Reynolds JR. Human brain activation time-locked to narrative event boundaries. *Psychological Science*. 2007; 18:449–455. [PubMed: 17576286]
- Sporns, O. *Networks of the brain*. Cambridge, MA: MIT Press; 2011.
- Stein M, Federspiel A, Koenig T, Wirth M, Lehmann C, Wiest R, et al. Reduced frontal activation with increasing 2nd language proficiency. *Neuropsychologia*. 2009; 47:2712–2720. [PubMed: 19501603]
- Tobia MJ, Iacovella V, Davis B, Hasson U. Neural systems mediating recognition of changes in statistical regularities. *Neuroimage*. 2012; 63:1730–1742. [PubMed: 22906790]
- Tobia MJ, Iacovella V, Hasson U. Multiple sensitivity profiles to diversity and transition structure in non-stationary input. *Neuroimage*. 2012; 60:991–1005. [PubMed: 22285219]
- Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK. Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*. 2009; 21:1934–1945. [PubMed: 18823241]
- Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM. Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience*. 2010; 30:11177–11187. [PubMed: 20720125]
- Ulanovsky N, Las L, Nelken I. Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*. 2003; 6:391–398. [PubMed: 12652303]
- Ventura-Campos N, Sanjuan A, Gonzalez J, Palomar-Garcia MA, Rodriguez-Pujadas A, Sebastian-Galles N. Spontaneous brain activation predicts learning ability of foreign sounds. *Journal of Neuroscience*. 2013; 33:9295–9305. [PubMed: 23719798]
- Weiss DJ, Gerfen C, Mitchel A. Speech segmentation in a simulated bilingual environment: A challenge for statistical learning? *Language Learning and Development*. 2009; 5:30–49. [PubMed: 24729760]
- Weiss, DJ., Poepsel, T., Gerfen, C. Tracking multiple inputs: The challenge of bilingual statistical learning. In: Rebuschat, P., editor. *Implicit and explicit learning of languages*. Amsterdam: John Benjamins Press; 2015. p. 167-190.
- Worsley, KJ. Statistical analysis of activation images. In: Jezzard, P., Matthews, PM., Smith, SM., editors. *Functional MRI: An introduction to methods*. New York: Oxford University Press; 2001.
- Yang J, Gates KM, Molenaar PCM, Li P. Neural changes underlying successful second language word learning: An fMRI study. *Journal of Neurolinguistics*. 2014; 33:29–49.
- Zacks JM, Braver TS, Sheridan MA, Donaldson DI, Snyder AZ, Ollinger JM, et al. Human brain activation time-locked to perceptual event boundaries. *Nature Neuroscience*. 2001; 4:651–655. [PubMed: 11369948]



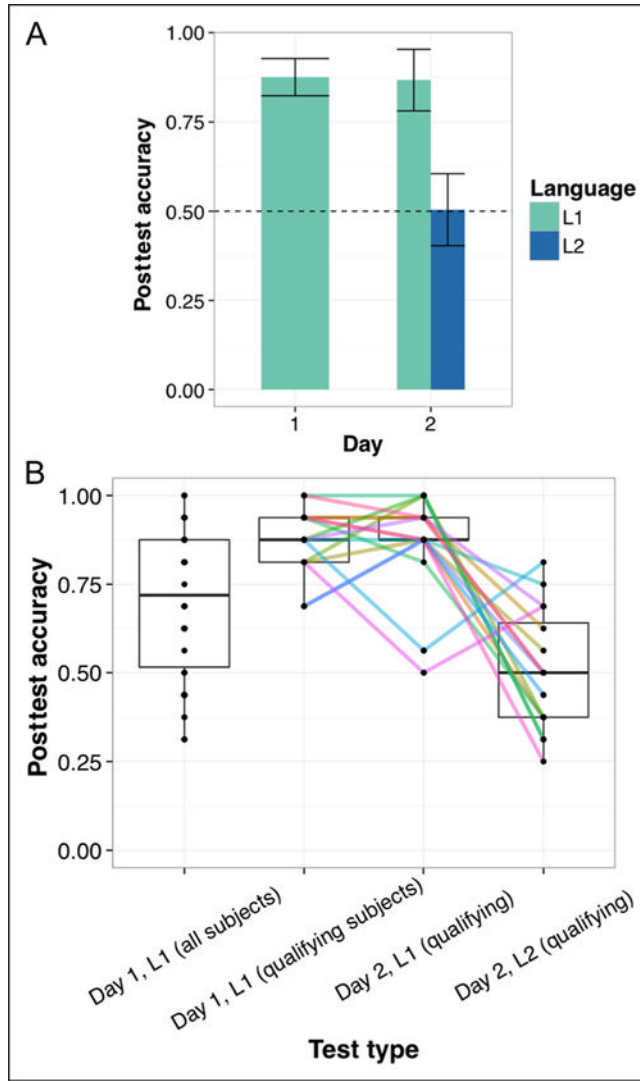


**Figure 1.** Directed graph displaying underlying word structure in the vowel and consonant frame languages. Nodes represent segments and edges represent transitional probabilities between them. Note that, within each word, the maximally informative statistic is the perfectly predictive relationship between either vowels (top plot) or consonants (bottom plot).

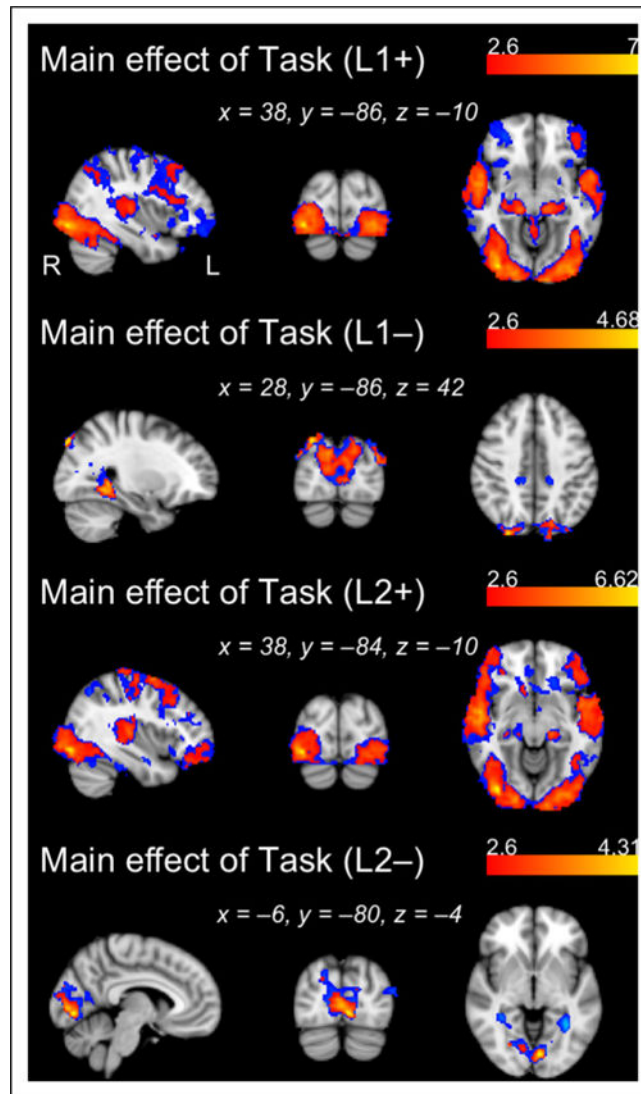


**Figure 2.**

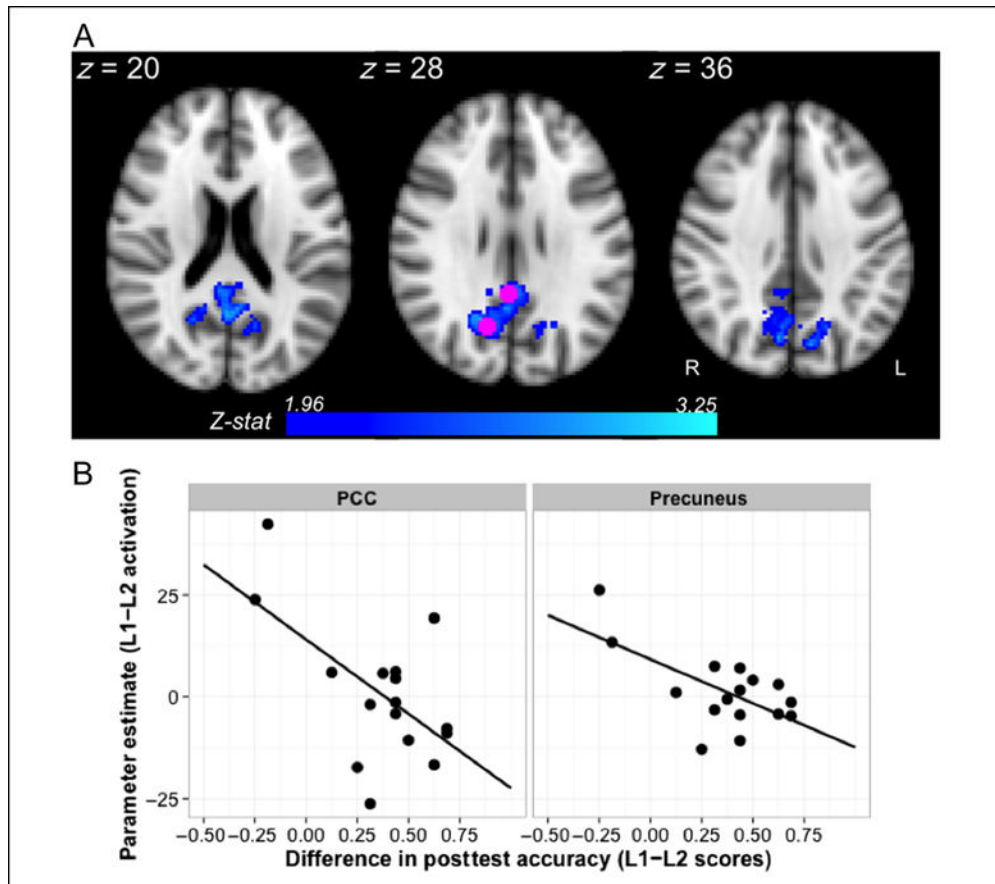
Experimental design. Participants were first exposed to L1 in a mock scanner. Those with qualifying scores were invited to participate in a scanning session within 48 hr of the initial session. During the scanner phase, participants were exposed to an additional 9.3 min of L1 (teal) followed by 9.3 min of L2 (blue). The transition between the two streams was not explicitly marked, and presentation of each language was accompanied by an identical visual stimulus (an “alien”).



**Figure 3.** Mean posttest scores for qualifying participants (A) and the spread of individual participant performance across testing days (B). Note that even after exposure to L2, qualifying participants tended to maintain their L1 representations in the Day 2 scanning session. Error bars in A represent 1 *SEM* corrected for within-subject comparisons, and colored lines in B indicate individual trends in performance across tests.

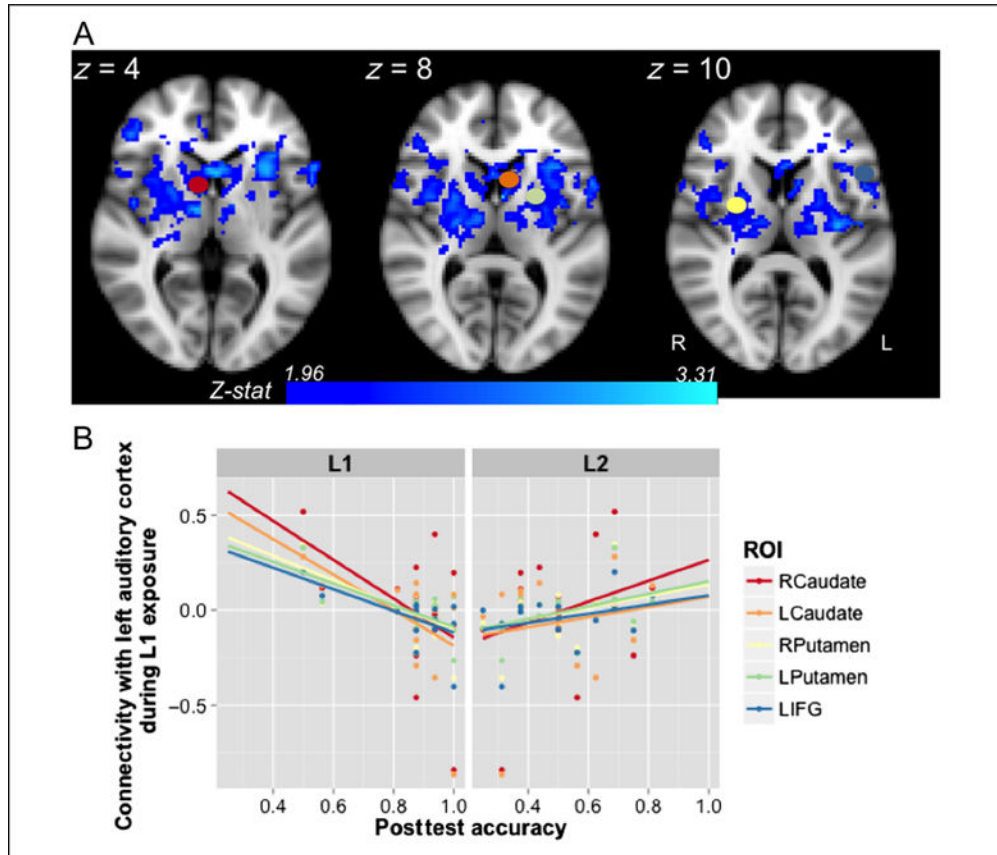


**Figure 4.** Regions showing positive and negative main effects of Task for each language, L1 and L2 (see also Table 2). Activation maps generated with a cluster-determining threshold of  $Z > 2.6$  are overlaid (in red) onto activation maps generated using  $Z > 1.96$  (in blue).



**Figure 5.**

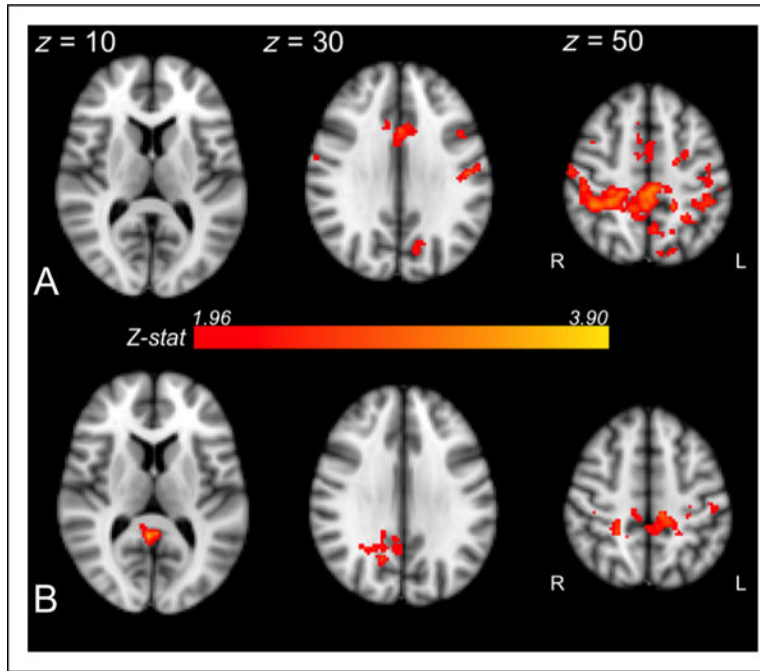
Inverse relationship between L1 versus L2 activation and the strength of the primacy effect measured behaviorally. Activation in PCC and precuneus was modulated by differences between L1 and L2 posttest accuracy (A;  $Z > 1.96$ ,  $p < .05$ ). As illustrated in B, lower L1 relative to L2 activation in these regions was related to a stronger primacy effect. To create the bottom plot, parameter estimates for each participant were extracted using the group level peaks in right precuneus and PCC, shown in pink, and plotted against individual difference scores (L1-L2). Negative values on either axis indicate lower L1 activation relative to L2 ( $y$  axis) or lower L1 accuracy relative to L2 ( $x$  axis).



**Figure 6.**

Inverse relationship between L1 connectivity with early perceptual cortex and L1 learning outcomes. The top plot (A) displays the results of a whole-brain PPI analysis in which we observed a negative correlation between L1 posttest scores and connectivity with left auditory cortex during L1 exposure ( $Z > 1.96$ ,  $p < .05$ ). Decoupling from this sensory area, particularly in pFC and the basal ganglia, was associated with high accuracy on the L1 posttest. For illustrative purposes, the bottom plot (B) shows this relationship for each of the fronto-subcortical peaks color-coded in A. Note that we find no significant relationship between L1 connectivity with fronto-subcortical regions and L2 learning (but refer to Figure 7).





**Figure 7.** L1 connectivity (top plot, A) and L1 activation (bottom plot, B) positively correlated with L2 posttest performance in overlapping posterior parietal regions ( $Z > 1.96$ ,  $p < .05$ ). When precuneus and PCC were actively engaged during L1 processing (indicated both by connectivity and activation measures), participants were more sensitive to the L2 shift, as evidenced by robust L2 behavioral performance.

**Table 1**  
Regions Exhibiting a Lowered Level of L1 Activation in the Highest-achieving L1 Learners (L1 Posttest -)

Predictor	Cluster	Extent (Vox)	Region	x	y	z	Z Statistic
L1 posttest (-)	1	3134	R planum polare	42	0	-18	4.51
			R frontal operculum	46	16	6	3.51
			R temporal pole	30	18	-50	3.4
			R superior temporal gyrus*	68	4	-6	3.4
			R middle temporal gyrus	72	-12	-6	3.39
	2	1137	L supramarginal gyrus	-58	-52	28	4.13
			L angular gyrus	-38	-50	28	3.56
			L lateral occipital complex	-34	-58	34	3.52
			R inferior temporal gyrus	62	-50	-12	3.81
			R temporal occipital fusiform	40	-46	-20	3.75
	3	11910	R angular gyrus	48	-52	26	3.74
			L posterior cingulate gyrus	-10	-52	26	3.69
			L precuneus	-12	-58	8	3.52
			L central operculum	-42	-8	12	3.79
			L putamen	-34	-6	-4	3.45
4	1869	L insula	-44	-4	4	3.38	
		L frontal pole	-36	40	34	3.3	
		L inferior frontal gyrus	-40	24	12	3.12	
		R planum polare	42	0	-18	4.51	
		R putamen	34	-12	4	3.21	
L1 posttest (-)	474	R temporal pole	30	10	-32	2.9	
		L supramarginal gyrus	-58	-52	28	4.13	
		L angular gyrus	-38	-50	28	3.56	
		L lateral occipital complex	-34	-58	34	3.52	
		R inferior temporal gyrus	62	-50	-12	3.81	
L1 posttest (-)	288	R temporal occipital fusiform	40	-46	-20	3.75	
		R planum polare	42	0	-18	4.51	
		R putamen	34	-12	4	3.21	
		R temporal pole	30	10	-32	2.9	
		L supramarginal gyrus	-58	-52	28	4.13	

<i>Predictor</i>	<i>Cluster</i>	<i>Extent (Vox)</i>	<i>Region</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>Z Statistic</i>
			R middle temporal gyrus	54	-50	-4	3.24
4	673		L posterior cingulate gyrus	-10	-52	26	3.69
			L precuneus	-12	-58	8	3.52

For each significant cluster, we present up to the top 5 activation peaks in anatomically distinct areas (in MNI coordinate space) as well as their corresponding peak *Z* statistics. The first four clusters represent significant learning-related activation observed with a cluster-determining threshold of  $Z > 1.96$ ; the subsequent four clusters represent the same activation map, this time thresholded at  $Z > 2.6$ .

Regions Showing Positive and Negative Main Effects of Task Are Also Presented (Main Effect of L1 +/- and L2 +/-)

Table 2

Predictor	Cluster	Extent (Vox)	Region	x	y	z	Z Statistic			
Main effect of task (L1+)	1	61559	R lateral occipital complex	38	-86	-10	7			
			R occipital pole	30	-94	0	6.09			
			L planum temporale	-42	-30	8	6.03			
			R superior temporal gyrus	58	-6	-6	5.85			
			L superior temporal gyrus	-58	-26	0	5.7			
Main effect of task (L1-)	1	8966	R lateral occipital complex	28	-86	42	4.68			
			L lateral occipital complex	-40	-82	22	4.63			
			L lingual gyrus	-30	-46	-6	4.53			
			R lingual gyrus	34	-42	-6	4.31			
			R occipital pole	14	-92	16	4.17			
			R precuneus*	10	-54	56	3.55			
			L precuneus*	-12	-46	50	3.48			
			L posterior cingulate gyrus*	-14	-30	38	2.75			
			R posterior cingulate gyrus*	16	-32	40	2.63			
			L precentral gyrus*	-16	-34	40	2.63			
Main effect of task (L2+)	1	55707	R lateral occipital complex	38	-84	-10	6.62			
			R superior temporal gyrus	56	-22	-2	6.18			
			R planum temporale	54	-18	4	5.85			
			L Heschl's gyrus	-42	-24	10	5.76			
			L planum temporale	-42	-30	10	5.71			
			R lateral occipital complex	46	-62	48	3.63			
			R angular gyrus	52	-56	36	3.08			
			R superior parietal lobule	32	-48	42	2.93			
			Main effect of task (L2-)	1	4233	L lingual gyrus	-6	-80	-4	4.31

<i>Predictor</i>	<i>Cluster</i>	<i>Extent (Vox)</i>	<i>Region</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>Z Statistic</i>
			L intracalcarine	-8	-86	2	4.14
			R intracalcarine	2	-82	4	3.93
			R occipital pole	10	-90	6	3.8
			R supracalcarine	2	-80	8	3.61
	2	1080	L precuneus*	-30	-56	8	3.86
			L temporal occipital fusiform*	-32	-50	-4	3.71
			L lateral occipital complex*	-38	-72	10	3.32
			L intracalcarine*	-28	-60	8	3.19
			L lingual gyrus*	-16	-42	-12	2.72

For each significant cluster (determined by  $Z > 1.96$ ,  $p < .05$ ), we present up to the top 5 activation peaks in anatomically distinct areas (in MNI coordinate space) as well as their corresponding peak  $Z$  statistics. Asterisks (\*) represent regions that did not survive a more conservative cluster-determining threshold of  $Z > 2.6$ .