ANNALS of THE NEW YORK ACADEMY OF SCIENCES

Neural bases of learning and recognition of statistical regularities

Journal:	*Ann NY Acad Sci*
Manuscript ID	annals-2000-196.R4
Manuscript Type:	Original Article
Date Submitted by the Author:	18-Dec-2019

Complete List of Authors:	Ordin, Mikhail; Basque Center on Cognition Brain and Language, Consciousness; Ikerbasque, Consciousness research group Polyanskaya, Leona; Basque Center on Cognition Brain and Language, Spoken Language Research Group Soto, David; BCBL, Consciousness research group; Ikerbasque, Basque Foundation for Science

Keywords:	statistical learning, segmentation, statistical generalization

Neural bases of learning and recognition of statistical regularities

Mikhail Ordin[1,2] and Leona Polyanskaya[1] and David Soto[1,2]

[1]BCBL – Basque Centre on Cognition, Brain and Language and [2]Ikerbasque – Basque Foundation for Science, San Sebastián, Spain

Address for correspondence: Basque Centre on Cognition, Brain and Language, Paseo Mikeletegi 69, San Sebastián, 20009 Spain. m.ordin@bcbl.eu

Short title: *Neural bases of statistical learning*

Graphical abstract

Previous studies have only assessed brain responses to trained words and novel non-words, and hence do not provide sufficient information on how the brain mediates the recognition of word-like units versus mere statistical regularities within sequences. The present study addresses this issue, as well as determines whether learning of statistical regularities embedded into a continuous sensory input and discrete constituents comprising this input, and subsequent recognition of extracted constituents, relies on the same mechanisms.

Abstract

Statistical learning is a set of cognitive mechanisms allowing for extracting regularities from the environment and segmenting continuous sensory input into discrete units. The current study used functional MRI (N = 25) in conjunction with an artificial language learning paradigm to provide new insight into the neural mechanisms of statistical learning, considering both the online process of extracting statistical regularities and the subsequent offline recognition of learned patterns. Notably, prior fMRI studies on statistical learning have not contrasted neural activation during the learning and recognition experimental phases. Here we found that learning is supported by the superior temporal gyrus and the anterior cingulate gyrus, while subsequent recognition relied on the left inferior frontal gyrus. Besides, prior studies only assessed the brain response during the recognition of trained words relative to novel non-words. Hence a further key goal of this study was to understand how the brain supports recognition of discrete constituents from the continuous input vs. recognition of mere statistical structure that is used to build new constituents that are statistically congruent with the ones from the input. Behaviorally, recognition performance indicated that statistically congruent novel tokens were less likely to be endorsed as parts of the familiar environment than discrete constituents. fMRI data showed that the left intraparietal sulcus and angular gyrus support the recognition of old discrete constituents relative to novel statistically congruent items, likely reflecting an additional contribution from memory representations for trained items.

Keywords: statistical learning; segmentation; statistical generalization; fMRI; sensory input; information

Introduction

Statistical learning allows agents to detect regularities in the world around them. These statistical cues can be used to split the continuous flow of sensory information (visual, auditory, tactile) into discrete constituents, a process called segmentation. The neural mechanisms underlying segmentation are evolutionary ancient[1–4] and shared by a diverse range of species[5–8]. Segmentation based on statistical cues contained in the sensory input operates across different domains. In humans, this includes splitting speech into words and phrases[9], separating distinct rhythms and other musical properties [in musical compositions][10–12], parsing sequences of events as well as discerning discrete sequences of actions in a continuous series of human activities[13,14]. For example, while viewing a series of still images representing a continuous dynamic activity, viewers´ gaze tends to dwell longer on those slides that illustrate the boundaries between unfolding events: dwell times are longer for slides that show the grasp of a glass has been completed than those that show the grasping action still unfolding. This segmentation takes place at various levels: slides that depict boundaries between distinct higher-level actions, for example, the boundary between emptying a dishwasher (which includes the lower-level action of grasping a glass to take it out of the machine) and starting a new sequence of sweeping the floor, attracts even longer gazes[15]. It is more difficult to predict actions that follow boundaries so they attract more attention. By contrast, when the next action can be easily predicted based on previously observed events, less attention is required – and dwell times diminish – because the further unfolding of events is highly predictable. These results were interpreted within the framework of statistical learning[16]. Besides, it has been suggested that the segmentation of actions, of continuous sensory input across modalities, and segmentation of speech into linguistic constituents like words and phrases – all rely on the same cognitive processes related to statistical learning[14,16].

Statistical learning operates on a variety of cues, including (but not limited to) conditional regularities known as transitional probabilities (TPs). TPs refer to the probability of an event B happening given that event A has occurred. Higher TPs characterize the events that commonly happen sequentially one after another, while lower TPs are aligned with the boundaries between the sequences of commonly co-occurring events. Thus, the differences between high and low TPs between events allow breaking continuous flow of events into discrete sequences, with events within sequences being more predictable than those spanning the edges of these sequences. Although this tokenization, or segmentation mechanism has been shown to operate across domains (see references above), it has been most extensively studied in the context of speech processing. For instance, during speech processing, continuous stream of syllables (i.e., events) can be segmented into separate words by calculating transitional probabilities (TPs) between adjacent syllables[1]. For example, in the phrase "*pretty baby",* the probability that the syllable *"ty*" will follow syllable *"pret"* is higher than the probability *"ba"* will follow *"ty"*. Minima in TPs between adjacent syllables, compared to surrounding TPs (i.e., local minima), are aligned with word boundaries. They can be used by infants learning their first language or adults exposed to a foreign language to segment a sequence of syllables into discrete words[17].

What is still debated, however, is how new constituents are identified. Do listeners detect word boundaries between consecutive constituents based on lower TPs, or do they merge smaller frequently co-occurring units into a single constituents[9]? Some researchers advocate for clustering mechanisms[18,19], while others argue in favor of boundary-finding mechanisms[20–22]. Some studies demonstrate that both human and non-human animals employ various strategies, which might rely on different neural mechanisms, and show that the choice of the strategy is determined by individual preferences, peculiarities of sensory input, and environmental circumstances[9,23].

The use of statistical cues in speech segmentation is usually studied within the artificial language learning paradigm[24]. A set of artificial words is concatenated into a continuous acoustic stream, with each word in the stream recurring multiple times. The syllable pairs with lower inter-syllabic TPs are more likely to straddle word boundaries than syllable pairs with higher TPs, which, in turn, are more likely to be confined within word boundaries. This enables the segmentation of the continuous syllable stream into words. Performance in speech segmentation tasks is tested by habituating the listener to the constructed acoustic stream, then administering a recognition test, in which participants need to endorse or reject test tokens as word candidates. Test tokens can either be statistical words from the learning stream, or sequences that violate the statistical regularities embedded in the habituation stream. Empirical studies convincingly demonstrate that words are endorsed as legal word candidates, while syllabic sequences that violate statistical regularities are rejected[25]. We aim to explore the neural bases of statistical learning in the context of this speech segmentation paradigm using the interesting case of so-called phantom words. Phantoms are test tokens that conform to the acquired statistical regularities but never occurred during habituation. For instance, listeners may be exposed to recurrent syllabic triplets, including XYA and BYZ triplets, with 0.5 transitional probabilities between syllables within triplets. Although syllabic pairs XY and YZ frequently occur in the familiarization stream, the sequence XYZ is never presented during habituation. It is still not clear if or how the brain differentiates between familiar structural units (i.e. artificial words) and novel units that are structurally congruent with the old ones. Addressing this question promises to provide important insights into the neurocognitive bases of statistical learning, namely, whether the brain relies on structural regularities, memory representations or a mixture of both while endorsing different types of tokens as legitimate candidates as discrete constituents in a continuous environmental input. Observing differences in brain responses to words and phantoms would suggest that different cognitive mechanisms are employed in the recognition of these tokens, notably, even when tokens of both types are not differentiated behaviorally. Phantoms can only be endorsed based on the recognition of their congruence with the statistical regularities embedded in the sensory input, while words may additionally rely on memory representations of whole discrete elements.

To date, there is conflicting evidence as to whether, after exposure to artificial language, phantoms emerge as perceptual units during recognition. Not all researchers have observed that phantoms are confused with holistic triplets (i.e. word-like structures)[21,26]. Furthermore, individual differences and the native language of the listener can influence whether or not phantoms are confused with words[22,27]. A key goal of the present study is to provide novel insights into the mechanisms supporting statistical learning using a novel behavioral and neuroimaging protocol to partial out the processing of words, phantoms and pseudorandom sequences during learning and subsequent recognition.

Earlier studies have shown that learning based on transitional probabilities (TPs) in the auditory modality is supported by the superior temporal gyrus (STG)[28–31] and inferior frontal gyrus (IFG)[28,32,33], mainly – but not exclusively – in a left-lateralized network. These studies have suggested that the STG is involved in learning TPs in the acoustic input, while the IFG is hypothesized to support the learning of word-like units[28]. As learning progresses, brain responses to occasional violations of statistical regularities can also be observed in frontoparietal cortex[34,35], particularly in the control network, which includes the right angular and bilateral anterior cingulate gyri[28,29,35,36]. Detection of violations of statistical structure during training is also supported by the temporoparietal junction[34,37], in line with its established role in attentional re-orienting to unexpected stimuli[38].

However, previous studies have only assessed brain responses to trained words and novel non-words. Hence, they do not provide sufficient information on how the brain mediates the recognition of word-like units versus the recognition of mere statistical regularities (i.e., TPs) within sequences. The present study addresses this issue for the first time. A second key objective of this study is to determine whether learning of statistical regularities embedded into a continuous sensory input and discrete constituents comprising this input on the one hand, and subsequent recognition of extracted constituents on the other hand relies on the same mechanisms[9,28,35]. This is an issue that is hard to tackle using behavioral measures alone.

Cognitive processes related to memory encoding and retrieval may rely on shared neural mechanisms[39,40]. However, most prior fMRI studies that used a statistical learning paradigm only recorded brain activity online (during learning) or offline (later, during recognition). Critically, brain responses were not recorded during both learning and recognition within the same experiment. Additionally, the only fMRI study that examined both the learning and recognition stages[29] did not test for neural differences between these two phases. Here, we recorded BOLD responses during both the learning and recognition phases in order to determine whether online statistical learning and subsequent offline recognition were supported by similar or different brain substrates.

Methods

Participants

We analyzed the data from 25 native Spanish participants (11 males between 20 and 33, average age 25.5 years, SD = 3.29). MRI data from one participant was discarded because he did not follow instructions (i.e. during the recognition test he pressed the same button for all responses). All participants had acquired Basque in childhood after the age of two as their second language and were using it daily. We note that prior studies have not revealed differences between monolingual Spanish and bilingual Spanish-Basque participants in the segmentation of statistical units in off-line recognition tests[41,42]. Also, to mitigate the possibility that individual differences related to bilingualism might influence statistical learning processes, we homogenized our sample by matching participants by proficiency and age of acquisition (2–3 years) of the second language, as well as the self-reported extent of their daily exposure to and use of Basque. None of the participants had any prior history of neurological disorders. The experiment was approved by the BCBL ethical committee. All participants provided informed consent.

Experimental materials and procedures

Learning phase. The habituation stream was composed of alternating structured and random blocks. Structured blocks consisted of concatenated syllable triplets, so that higher TPs between syllables within these triplets and lower TPs between syllables straddling the triplet boundaries allow for predicting the following syllable after both triplet-initial and triplet-medial syllables, but not after triplet-final syllables. Random blocks consisted of the same syllables unsystematically concatenated so that the TPs between them were uniform throughout and therefore did not allow for segmentation.

The habituation stream was synthesized using MBROLA, with ES1 voice and fundamental frequency invariably set to 110Hz. We used a set of 18 consonant-vowel syllables. The duration of each syllable was 240 ms (100 ms for consonants and 140 ms for vowels). These syllables were used to construct 12 trisyllabic statistical words, with TPs between syllables within triplets set to 0.5. These triplets were randomly concatenated 21 times with the restriction that the same word was never repeated consecutively. TPs between syllables straddling triplet boundaries were approximately 0.16. The difference in TPs between syllables within triplets and TPs straddling triplet boundaries provided statistical cues for the segmentation of the continuous stream of syllables into recurrent trisyllabic sequences (words). This procedure was repeated three times to create three structured blocks of 181.44 sec. We also pseudorandomly concatenated all 18 syllables from the artificial language syllabic inventory six times, such that no syllable was ever repeated consecutively. We prepared three pseudorandom blocks of 25.92 sec each. The same syllable inventories were used during structured and pseudorandom blocks because our goal was to test for differences in brain responses to the presence or absence of statistical structure without introducing any confounds due to differences between the acoustic properties of the stimuli. Also, we elected not to use rest periods as the baseline but rather contrasted activity in structured and pseudorandom blocks during learning. It was likely that overall brain state in pseudorandom blocks would be more similar to that in structured blocks than during rest. Therefore, we could be confident that any differences in neural responses to random and structured blocks were elicited because structured blocks included recognizable structure with extractable and learnable constituents. Contrasting structured blocks to rest would also have introduced the problem of individual variability because mental activity during rest can vary, engaging different mechanisms and different networks[43].

The habituation syllabic stream was prepared by alternating structured and pseudorandom blocks three times. At the end of the stream, we added an additional 36 randomized syllables (each of the 18 syllables from the inventory repeated twice for a total of 8.64 sec) and applied a fade-out filter. A fade-in filter was also applied at the beginning of each stream, thus preventing any potential anchoring effects of stream-initial and stream-final syllables on segmentation. As statistical learning mechanisms are constantly operating on incoming sensory input, it is possible – though unlikely, given the difference in exposure time to structured and random blocks – that participants collapsed conditional statistics across conditions. However, this possibility does not undermine the validity of the cues for statistical segmentation since the syllable pairs with higher TPs were more likely to fall within the recurrent triplets.

We prepared two similar habituation streams which had unique orders for recurrent triplets within structured blocks and for syllables within pseudorandom blocks. Streams 1 and 2 were used for the first and second runs of the learning session. During the learning phase, participants were asked to listen to an "extraterrestrial language" and to try to detect and memorize the words from that language.

Recognition phase. The recognition test was comprised of four test runs each comprising 63 trials. In each trial, we randomly concatenated either 4 different triplets from the habituation stream, 4 different phantoms – triplets that fit the statistical regularities of the habituation input but had never occurred in the learning stream as whole constituents, or 4 non-words (i.e., triplets composed of syllables that never occurred consecutively in the habituation stream). Each run included 21 trials of each type. The duration of the stimuli in the recognition test was 2880 ms. Each triplet was used an equal number of times across all trials. The stimuli were preceded and followed by 200 ms silence. After each stimulus presentation, participants were asked to decide whether that acoustic sequence had been presented during the learning session, and then to rate their confidence in the given response, on a 4-point scale. The period for each response was fixed to 2000 ms. The trials were separated by a jittered time interval according to a pseudo-exponential distribution from 3000 ms to 5000 ms in steps of 500 ms.

Both the learning and recognition phases were performed inside the scanner. The sound was played via in-ear Sensimetrics S14 headphones. A pair of headphones was placed above the in-ear headphones in order to dampen the noise of the scanner and to enable communication with the experimenter. The stimuli were back projected onto a screen by a mirror on the head coil. The area between the participant's head and the coil was padded with foam to make the participant more comfortable and to minimize head movements. We asked the participants not to move during scanning.

To familiarize the participants with the procedure and the experimental protocol and interface, a brief training session was organized outside the scanner, with a 40-second familiarization stream and 4 recognition trials. The syllables for this training session were different from those used in the actual experiment. The list of statistical words, phantoms and non-words are given in Table 1. The structure of the learning runs and recognition trials is illustrated in Figure 1.

Functional and structural MRI data acquisition

Whole-brain MRI data acquisition was conducted in a 3T MAGNETOM PRISMAfit MR scanner using a 64-channel coil. T1-weighted images were acquired using MP-RAGE sequences with the following parameters: TR = 2530 ms, TE = 2,36 ms, FoV = 256 mm, flip angle = 7 degrees, acquiring 176 contiguous 1 cubic mm slices per run. Functional images were acquired using a multi-band acceleration factor of 6 (multi-slice interleaved mode), with 66 contiguous 2.4 cubic mm slices, TR = 850 ms, TE = 35 ms, flip angle = 56 degrees. We achieved whole-brain coverage.

FMRI data pre-processing

Image pre-processing was performed in FSL 5.0.9 using the FEAT module. First, we used the brain extraction tool (BET[44]) to separate the brain matter from non-brain tissues. The first 11 volumes of each run, both in learning and in recognition tests, were discarded to control for magnetic saturation effects and allow for MR signal stabilization. We used a high-pass filter cutoff of 100 sec for the learning runs and of 60sec for the testing runs following FSL manual instructions for blocked and event-related designs. Scans were realigned by using MCFLIRT motion correction (spatial smoothing with a 6-mm FWHM Gaussian kernel applied). Translation parameters did not exceed half a voxel in any direction for any participant in any run. Functional images were registered to T1 structural images (7 degrees of freedom for testing runs and using the boundary based registration BBR algorithm[45] for the learning runs). Then, the images were registered to the standard MNI152 template using affine registration with 12 degrees of freedom, using full search setting.

FMRI data analysis

Learning phase. Statistical analysis for the learning stream was performed within the framework of the general linear model in the individual native space first, with statistical maps normalized to the standard space prior to higher-level analyses. Each pseudorandom block was entered as a separate explanatory variable (EV). Critically, chunks with duration of 25.92 seconds of structured blocks immediately preceding the pseudorandom blocks (also 25.92 seconds) were entered as separate EVs in order to match the two conditions in the number of scans so that the amount of data for the comparison of the BOLD responses during structured and pseudorandom blocks is equated. The numbers of scans were equated in order to have a similar signal for the contrast of blood oxygen level dependent (BOLD) differences between structured and pseudorandom blocks. See Rosenthal *et al*.[46] for detailed methodological justifications for the necessity to equate the amount of data while comparing the BOLD response between two conditions.

Each EV specified the onset of the pseudorandom block or the onset of the structured chunk. The EVs were introduced in the model along with their temporal derivatives. We applied FILM pre-whitening[47]. Standard and extended motion parameters were introduced in the GLM model as additional regressors of no interest[48]. Furthermore, we used the FSL motion outliers function (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLMotionOutliers) and included the regressors corresponding to the motion outliers in the design matrix in order to deal with the effect of intermediate-to-large motions that could potentially corrupt images beyond the level that the extended motion parameter regression methods could deal with. To detect the volumes containing motion outliers, we calculated the root mean square (RMS) head position difference to the reference volume and compared the 75 percentile +1.5 inter-quartile range of the distribution of RSM values for each run. A confound matrix was generated and used in the GLM to completely remove the effects of these timepoints on the analysis.

Parameter estimates were calculated for the following contrasts in which brain activity was higher for pseudorandom relative to structured blocks (R1>S1, R2>S2, R3>S3, R4>S4, R5>S5, R6>S6, where “R” stands for pseudorandom, “S” stands for structured, and the number indicates the sequential number of the pseudorandom block or immediately preceding structured chunk, i.e., R1, R2, and R3 are parts of run 1 and R4, R5, and R6 are parts of run 2. The resulting 6 contrasts of parameter estimates reflect acquisition of relevant sequence knowledge.

We then performed a second level, fixed-effects analysis within each participant using the 6 parameter estimates noted above. Here we tested for different temporal profiles of learning related activity (S<R and S>R) across the training phase. We assessed a logarithmic trend (specified using the following contrast: -3.125, -1.0, 0.5, 1.0, 1.25, 1.375) and an exponential increase (specified using the following contrast: -1.375, -1.25, -1.0, 0.5, 1.0, 3.125). The logarithmic increase suggests that the difference in BOLD change on pseudorandom and structured blocks is larger at the beginning of the learning phase and attenuates as learning progresses. This trend can be caused by faster learning at the beginning of exposure, with the learning rate then decelerating. Exponential increase is a reciprocal function, which, on the contrary, suggests that learning is slower at the beginning, and accelerates with time, causing the differences in the BOLD response on pseudorandom and structured blocks to increase more rapidly as habituation progresses.

The output of the contrasts was fed into a mixed-effect model using the FLAME 1 algorithm in FSL[49] in order to test for the consistent effect across participants (group Z >2.3, cluster significance threshold *P* = 0.05, corrected using Gaussian field theory).

Recognition test. Statistical analysis for the recognition test was first conducted within the framework of the general linear model in native space, with statistical maps normalized to the standard space prior to higher-level, group analyses. We created the EVs for words endorsed as words (*words_acc*), rejected non-words (*nonw_rej*), rejected and endorsed phantoms (correspondingly *phan_rej* and *phan_acc*). We modelled the onset of each EV with durations that corresponded to the length of the stimulus (2.88 secs). Regressors of no interest were introduced into the design matrix as separate EVs to control for variation in decision response time both for the first recognition response (i.e., whether the stimulus was presented during the learning session) and for the confidence rating. The EVs were introduced along with their temporal derivatives. We applied FILM pre-whitening with standard and extended motion parameters and introduced an additional EV for the motion outliers.

We then estimated contrasts of parameter estimates that were relevant to our study goals. In particular, we assessed the brain substrates that support (1) the recognition of words vs. non-words and critically (2) the recognition of words vs phantoms. In these analyses we used trials with correct responses (i.e. *words_acc* vs. *nonw_rej* for (1) and *words_acc* vs. *phantoms_rej* as well as *words_acc* vs. *phantoms_acc* for (2). The contrasts were estimated in both directions, i.e., for the contrast *words_acc* vs. *nonw_rej* we estimated contrasts of parameter estimates both for *words_acc* > *acc-nonw_rej* and *words_acc* < *nonw_rej*. We reasoned that accepting words can rely both on the recognition of word-like structural information present in the memory traces of the triplets and the statistical structure contained in the lower-level TPs. The comparison of the BOLD signal change between words and phantoms ought to reveal the brain substrates that support the recognition of word-like structural information. The trials with phantoms involve analysis of statistical structure and making decisions based only on statistical congruency. Alternatively, rejecting phantoms may rely on the fact that they are not supported by memory representations because the phantoms were not encountered and extracted as whole units during the learning phase. Hence, we expected the memory network to be more activated for accepted words than both rejected and accepted phantoms.

We performed within-subject, cross-run (fixed effects) analysis to estimate the individual mean of each contrast across all runs of the recognition test. In order to find effects that were consistent across subjects, all contrasts were fed into a mixed effect model for the whole-brain group analysis using FLAME 1, thresholding the statistic images (Z >2.3, *P* = 0.05 corrected using Gaussian Random Field theory[49,50].

Results

Behavioral results

Behavioral data was acquired only during the recognition test. We estimated the percentage of endorsed words, phantoms and non-words. Also, we calculated the mean confidence rating assigned to endorsed and rejected words, non-words and phantoms. Although the percentage of correct responses might seem somewhat lower than what is usually reported in artificial language learning experiments, it is not extraordinarily low. The environment, in which participants had to do the task was more challenging (i.e. performed inside the scanner against strong background noise), the number of discrete constituents was larger (12 triplets) than what is usually used in similar experiments (4 triplets), and the differences in TPs between syllables within words and between syllables spanning word boundaries was less pronounced (50% vs. 16%) than what is usually used (100% vs. 33%). All these factors increased the difficulty of the task. In addition, the interleaved random blocks could also have had a detrimental effect on learning.

We performed one-sample t-tests comparing the percentage of endorsed tokens in each condition (words, phantoms and non-words) with chance level performance (50%). The results

(Fig. 2) show that the percentage of endorsed words is significantly above what would be expected by chance, t(24) = 3.224, *P* = 0.004, M (mean difference) = 12.08%, 95%CI [4.35: 19.81], d = 0.645. The percentage of endorsed non-words is significantly below what would be expected by chance, (t(24) = -6.379, *P* < 0.0005, M = -23.63%, 95%CI [-34.27: -15.98], d = 1.276. The percentage of endorsed phantoms is not significantly different from chance, t(24) = 1.285, *P* = 0.211, M = 10.71%, 95%CI [-3.25: 13.95], d = 0.257.

Repeated-measures ANOVA revealed significant differences in the percentage of endorsed words, phantoms and non-words, F(2, 48) = 60.009, *P* < 0.0005, η_p^2 = 0.714 (*p* value corrected with the Greenhouse-Geisser method, *df* are reported uncorrected). Pairwise comparison (with the Bonferroni correction applied) showed that the proportion of endorsed words was significantly higher than that of endorsed phantoms, t(24) = 3.371, *P* = 0.008, M = 6.72%, 95%CI[2.6: 10.84], d = 0.674. The percentage of endorsed phantoms was also significantly higher than the percentage of endorsed non-words, t(24) = 7.245, *P* < 0.0005, M = 28.98%, 95%CI[20.73: 37.24], d = 1.449. Unsurprisingly, the proportion of endorsed words was also significantly higher than the proportion of endorsed non-words, t(24) = 8.922, *P* < 0.0005, M = 35.71%, 95%CI[27.45: 43.97], d = 1.785.

We then analyzed the confidence ratings. The results showed that endorsed words were assigned significantly higher confidence ratings compared to incorrect, rejected words, t(23) = -4.253, *P* < 0.0005, M = 0.334, 95%CI [-0.497: -0.172], Cohen's d = 0.96. The same pattern was found for non-words, t(24) = -2.428, *P* = 0.023, M = -0.23, 95%CI [-0.425: -0.035], Cohen's d = 0.65. Interestingly, the level of confidence assigned to endorsed phantoms was significantly higher than rejected phantoms, t(24) = 4.451, *P* < 0.0005, M = 0.353, 95%CI [0.189: 0.517], Cohen's d = 0.812. As Figure 3 shows, the level of confidence for accepted phantoms matches that of accepted words, t(24) = 1.118, *P* = 0.275. Similar results were found for confidence ratings on trials with rejected words and rejected phantoms, t(24) = 0.237, *P* = 0.814.

Taken together, these results show that participants have learnt discrete constituents from the auditory input and reliably endorse them later during the test, while rejecting those sequences which violated the statistical regularities embedded in the familiarization stream. However, when participants encountered phantoms, which were consistent with the statistical probabilities defining the structural constituents but had never been presented as holistic units during the learning phase, participants were at a loss, and could not unambiguously accept or reject them. They responded randomly, at a chance level, which resulted in a significantly higher proportion of accepted words than accepted phantoms. However, once a decision was made, participants assigned higher ratings to accepted than to rejected phantoms, showing that the metacognitive system treated acceptance of phantoms as a correct response. This lack of difference in confidence ratings assigned to correctly endorsed words and accepted phantoms suggests that metacognitively phantoms were treated as words, even though the cognitive system treated words and phantoms differently.

fMRI results

Learning phase. Our goal here was to delineate the neural correlates of learning related changes during the study phase. Accordingly, we tested the effect of training on the differences in BOLD response between the successive structured (S) and pseudorandom (R) chunks (i.e. our index of learning). We compared two types of models in which (1) learning related activity mainly occurred during the first training run and then remained constant during the second training run, and (2) learning related changes emerged in the second training run (see Methods). Based on prior research on perceptual sequence learning, we elected to focus our analyses on the S<R contrast[46,51–53]. For the sake of completeness, we also conducted a similar analysis based on S>R parameter estimates (as was done in Ref. 28), however, here we did not find any significant results at the group level.

The significant fit of learning related activity with a logarithmic trend indicates that learning-related brain activity builds faster at the beginning of the exposure and is then attenuated as training progresses. This was found in three clusters: (1) superior-frontal gyrus (SFG) extending to the paracingulate gyrus; (2) right superior temporal gyrus (STG); and (3) left STG. Figure 4 illustrates these results. Table 2 provides information regarding the peak voxels in MNI coordinates of the different contrasts. We did not find any evidence for linear or exponential increases in learning related activity across the two training runs suggesting that learning increases did not continue as training progressed further in the second run of the learning phase.

Recognition phase. Following the learning phase, participants were presented with a recognition test. On each trial, previously studied words, phantoms or non-words appeared for an old/new recognition decision followed by confidence ratings. Our key goal was to isolate the neural substrates implementing recognition of word units relative to phantoms (i.e. statistically congruent tokens that were not embedded in the learning input). We therefore ran three contrasts. First, we compared BOLD activity changes when words were accepted relative to when non-words were rejected. This contrast is similar to comparing pseudorandom and structured blocks during the learning phase. Most crucially, we then compared BOLD activity changes when words were endorsed relative to when phantoms were rejected as well as when words were endorsed relative to when phantoms were endorsed. These contrasts aimed to isolate the brain substrates activated by the recognition of word-like units as whole constituents relative to the recognition of merely statistical structure. We reasoned that phantoms may be accepted as legitimate elements of an artificial language due to their statistical congruency with word-like constituents[22,27]. Rejecting phantoms may therefore rely on the fact that they are not supported by memory representations; phantoms were not encountered and extracted as whole units during the learning phase. Hence, we reasoned that by splitting the tokens into accepted and rejected the chances of dissociating the brain basis of words vs phantom processing would increase.

Endorsed words, compared to rejected non-words (*words*_acc > *non-words_rej*), elicit BOLD response changes in the left inferior frontal gyrus (LIFG) around BA44, pars opercularis extending to par triangularis in BA45 (see Table 2). Critically, endorsed words relative to rejected phantoms (*words_acc* > *phan_rej*) elicited BOLD increases in the anterior part of the cingulate cortex, posterior division of the STG (strongly right lateralized), and in the left hemisphere in a cluster that involved the posterior division of the angular gyrus and anterior part of the intra-parietal sulcus (see Table 3). Figure 5 illustrates these results.

Finally, we report that no differences were found when comparing accepted words vs. accepted phantoms, which suggests that the recognition of equally familiar tokens may have a similar neural underpinning, whether the recognition is based merely on statistical congruency, or strengthened by memory representations of word-like units. This result is in line with the behavioral data on confidence, showing that accepted words are treated as accepted phantoms, and that participants processed the trials, in which they endorsed phantoms, as correct responses.

Comparison of learning and recognition phases. In order to compare neural substrates supporting online statistical learning with offline recognition of holistic constituents, we compared the brain activity maps associated with learning-related activity for structured vs. pseudorandom blocks, and the brain activity maps associated with the recognition of words relative to non-words. Note that structured/pseudorandom blocks during learning equate to the presentation of words/non-words during recognition, both acoustically and statistically. Hence, contrasting the recognition effect with the learning effect in terms of the associated BOLD activity maps will reveal whether these processes are supported by similar neural substrates.

Learning and recognition contrasts of parameter estimates were fed into a whole-brain mixed effects model paired t-test with subjects as random factors. We found that learning-related activity was stronger in the left STG and the right superior frontal gyrus, extending to the anterior cingulate cortex ($Z > 2.3$, $P = 0.05$ with a corrected significance level using Gaussian Random Field theory[49,50]). Table 4 and Figure 6 illustrates these results.

Discussion

The key objectives of the current study were (1) addressing the common and distinct neural substrates that support online statistical learning and the subsequent offline recognition of the learned constituents, and (2) defining the neural substrates that support the recognition of holistic constituents (learned words) as opposed to recognition of merely statistical structure (phantoms). Behaviorally, we found that participants could successfully recognize words vs non-words in terms of discrimination accuracy and response confidence. Higher confidence ratings were assigned to correctly endorsed words and correctly rejected non-words compared to the corresponding incorrect responses. The proportion of accepted words was also significantly higher than that of phantoms. Importantly, rejected phantoms were assigned lower confidence rating compared to endorsed phantoms, showing that on rejected phantoms participants estimate the likelihood of making an error to be higher than on accepted phantoms. At the same time, endorsed phantoms and words were assigned similar confidence. Overall, the results show that accepted phantoms were treated as accepted words, *metacognitively*, although in terms of accuracy (i.e., cognitive decisions), they were not confused with words. One of the driving forces that underlie endorsing sensory input as part of the environment is the concordance with the statistical regularities the individuals are familiar with. This motivates their response to both words and phantoms. Memory representations further strengthen the recognition of word-like units. Sometimes, phantoms are endorsed despite the lack of memory support for these tokens as whole constituents, suggesting that memory does not play a crucial role in the recognition of legitimate constituents[21,54]. The fact that statistically congruent tokens are sometimes endorsed in the absence of memory representations would allow for the transfer of processing skills from recently encountered to novel situations, as long as these novel situations exhibit recently encountered statistical features.

We turn now to the neuroimaging results. We found that the bilateral STG supports the online extraction of conditional statistical cues during learning. This is in keeping with a number of earlier studies both in the auditory and visual modalities[29,31,34] and with prior work demonstrating a role for STG in associative learning and relational memory[55,56], here related to statistical structure and the acquisition of the relational positions of syllables in a stream. Learning was also mediated by the cognitive control network. The level of processing load during auditory perception is known to regulate activity in the control network, especially in the paracingulate and anterior cingulate gyri[57]. Accordingly, we found an increased activity in these areas when pseudorandom sequences were presented during learning following structured sequences.

Notably, we did not observe that activity in the IFG changed differently for pseudorandom and structured blocks during the learning phase. This result seems at odds with the study of Abla and Okanoya[32], who showed that online segmentation of recurrent tone sequences in a continuous tone stream elicits higher levels of activity in the LIFG. Karuza et al.[28] however, found involvement of the LIFG for learning forward speech but not backward speech, and suggested that the results by Abla and Okanoya[32] and their own results[28] reveal the role of the LIFG in TP calculation and the formation of structural representations. Importantly, in the Abla and Okanoya's[32] study, participants were first trained on three-tone sequences presented in isolation, and later these tone triplets were concatenated in a continuous stream and presented in alternation with the same tones randomly concatenated (i.e., not built into triplets). As participants were already familiarized with the recurrent triplets prior to the exposure, the activation in the LIFG could actually indicate a neural response to the recognition of the already learnt constituents rather than the process of formation of new representations. The role of the LIFG in the recognition of learned constituents rather than online segmentation of TPs is shown by Turk-Browne *et al*.[56], who correlated familiarity ratings, assigned to discrete constituents during the recognition test with activity in the LIFG during learning exposure. Higher activity in the LIFG was observed for those constituents which were later rated as more familiar, indicating that neural responses in the LIFG differed for recognized vs. unrecognized tokens embedded into a continuous sensory input. Our findings are also in line with the hypothesis that activation in the LIFG is related to the recognition of discrete constituents rather than to the online segmentation of continuous input. By assessing learning and recognition processes independently within the same paradigm, our study allowed us to isolate the contributions of the left STG and the anterior cingulate cortex to learning, while the left IFG was critically involved in subsequent, offline recognition processes. We believe that the

LIFG is implicated in monitoring that statistically congruent sequences are indeed discrete constituents learned from the environment, which is in line with its proposed role as a general domain and modality independent sequence processor[28]. The right STG seems to be equally involved in learning and recognition. The conclusion that statistical learning and recognition are supported by different neural substrates also agrees with previous studies in other domains and modalities, which have shown differences in the neural networks supporting successful encoding (i.e., learning) and successful retrieval (i.e., recognition) of semantic and perceptual associations[55]. The lack of differential involvement of the IFG in the present study during learning and recognition is in keeping with the proposed account.

For the first time, in a functional MRI of statistical learning, phantom sequences were included in the recognition test phase alongside words. This allowed us to determine whether the processing of phantoms and words is mediated by a similar neurocognitive mechanism. Our behavioral evidence suggests that participants are confused by phantom cases and are as likely to accept as to reject them. Hence, we tested the extent to which processing of words and phantoms could be dissociated in brain responses. Overall, BOLD responses to endorsed words and endorsed phantoms did not differ, confirming the conclusion that accepted phantoms are metacognitively considered to be correct responses, and that once a phantom is accepted, it is processed as a legitimate structural constituent. However, when we compared BOLD responses for accepted words and rejected phantoms, we found stronger responses in the left angular gyrus and intraparietal sulcus associated with endorsed words. This reveals active activation of the memory network[39] elicited by retrieving memory representations of words as whole constituents presented during learning. Also, we found that the level of neural activity associated with accepted words and rejected phantoms differed in the anterior division of the cingulate cortex (ACC), which is frequently related to error detection and conflict resolution[59]. A significant difference in neural activation in the ACC for accepted words vs rejected phantoms suggests that on the trials in which phantoms were rejected a competing response was present, and thus the participants estimated the likelihood of making an error on such trials as high. This is manifested in the low confidence ratings assigned to rejected phantoms. Absence of difference in neural activation between accepted words and phantoms confirms our earlier conclusion, based on the behavioral results: accepted phantoms are treated as correctly endorsed words. This pattern of results also invites a tentative explanation, which nevertheless needs to be further empirically tested. We propose that the tokens are accepted mainly based on recognition of statistical structure, while rejection is based on the lack of memory representation. Thus, it is possible that memory representations do not yield additional support for recognizing constituents from the sensory input. It is rather the absence of memory representations that leads to the rejection of some statistically congruent items. We believe these findings are important considering that prior behavioral studies have not consistently observed that phantoms are confused with holistic triplets (i.e. word-like structures). Hence this work lays the foundation for future studies to further explore the conditions in which the brain can distinguish words vs phantoms, for instance, by manipulating the amount of exposure during the training phase. New insights may be provided by exploring the different cognitive mechanisms which underlie rejection and acceptance of novel statistically congruent tokens. This understanding might be important, for example, in the field of language learning, to explain the phenomenon of generalization, when a rule learnt on a small set of examples is generalized over previously unencountered cases, and the phenomenon of fossilization, when the transfer from known to novel situations does not happen and progress in learning is halted.

As we argued in the introduction, statistical learning mechanisms are evolutionary ancient, operate across a wide variety of taxonomically divergent species, and predate the emergence of language. Hence, although these mechanisms are engaged in speech processing and language acquisition, it is very unlikely that they evolved specifically for these purposes. We suggest that statistical learning mechanisms evolved to detect abrupt changes in the environment. The structure of ecologically relevant natural states is usually relatively stable, with rapid transitions between longer lasting stable states[60,61]. For survival and reproduction (i.e., fitness), organisms need to monitor the environment and detect and react to rapid ecologically-relevant changes as they suddenly occur. These fitness needs likely gave rise to the early emergence of statistical learning mechanisms during evolution and explain their spread across different taxa, domains and modalities. Detecting structural regularities is probably more important than detecting recurrent constituents, because any breach in ecological stability signals rapid and fitness-relevant environmental changes. The ability to detect statistical structure presumably predates and underlies the segmentation of the dynamical flow of experience and supports building abstract representations of segmented constituents that reflect the structure of the environment. As phantoms in our study were statistically congruent with words, the cognitive system might have confused some of them as being equally familiar as words at the behavioral level, because both correspond to stable states in the statistical structure of the acoustic environment. Nevertheless, additional support from memory representations (evidence for the memory support of words relative to phantoms is discussed below) affects confidence judgements. However, the presence of this support at the neural level does not override the importance of detecting breaches in statistical structure which signal environmental changes. In the environment of evolutionary adaptiveness that shaped the functions of the statistical learning mechanisms breaches in the statistical structure rather than recurrent constituents had to be monitored and required a behavioral response because they cued sudden ecological changes.

It may be argued that this activity contrast is related to endorsement vs. rejection, irrespective of underlying statistical structure. Several considerations argue against this. First, the BOLD response to the auditory sequence was modelled separately from that associated with the behavioral response in the recognition test that took place 4 seconds later. Critically, the BOLD activity patterns that we report are time locked to the onset of the auditory sequence. Second, the patterns of BOLD activity were found in putative substrates of statistical learning. Finally, the fact that there is a brain signal that distinguishes words (accepted) vs phantoms (rejected) likely reflects the contribution of memory representations derived throughout the training. Statistically congruent novel items (i.e., phantoms) that did not receive additional support from activation in the memory-related brain areas were rejected. If the level of the memory activation was the same for words and phantoms, then phantoms were accepted as constituents of the previously experienced sensory input. The role of the left intraparietal sulcus in memory is well established[39,58]. Previous studies have also found post-learning sensitivity in the angular gyrus to the presentation of statistically congruent sequences[28,35]. The angular gyrus also underlies discrimination of pseudo-words and real words from natural languages[62]. The relations between linguistic experience and the functionality of the left angular gyrus is supported by work by Mechelli *et al*.[63], which showed that bilinguals and highly proficient L2 leaners have stronger grey matter density in the anterior division of the angular gyrus than monolinguals. The important role of the angular gyrus in the recognition of artificial words may be facilitated by its strong connectivity with temporal cortices via the arcuate fasciculus[64] and also with the inferior frontal gyrus, both BA 44[65] and BA 45[66,] via the longitudinal fasciculus. These connections are ipsilateral, which explains the simultaneous left-lateralized activation in multiple cortical areas, communicating through a major connection hub of the left angular gyrus[67]. Functional connectivity studies have also revealed a broad cortical network involved in statistical learning, with strong functional connections between both left and right STG, which we also observed here, alongside the LIFG[68]. Overall, our recognition results are in line with the conclusion of Skosnik *et al*.[35] that grammaticality and recognition judgements rely on different networks. Since both words and phantoms are statistically congruent, the recognition of words must rely on additional support from memory representations derived from the learning phase, which recruits additional neural networks, different from those engaged in phantom recognition.

To conclude, we found that the neural substrates underlying online statistical learning processes and offline recognition of the learned patterns rely on different neural substrates, indicating that the neurocognitive mechanisms that support the initial formation and subsequent maintenance of structural representations are distinct. Also, we dissociated, at the neural level, recognition of discrete constituents from the sensory input vs. recognition of mere statistical structure that is used to build constituents, enabling recognition of novel constituents that were not experienced before. Mechanisms for statistical learning have been shaped by fitness needs in the environment of evolutionary adaptiveness. We suggest that statistical mechanisms for detecting breaches in statistical structure are more essential to fitness than those that detect structural units; they are more evolutionary ancient and prevail over those that allow us to recognize structural units.

Acknowledgements

The research was supported by the Spanish Ministry of Economy and Competitiveness (MINECO) through the "Severo Ochoa" Programme for Centres/Units of Excellence in R&D (SEV-2015-490), and project grants RTI2018-098317-B-I00 awarded to MO, by the Basque Government through project grant PI-2017-25 awarded to DS, and by European Commission as Marie Skłodowska-Curie fellowship DLV-792331 to LP.

Competing interests

The authors declare no competing interests.

References

1. Kikuchi, Y., Sedley, W., Griffiths, T. D. & Petkov, C. I. Evolutionarily conserved neural signatures involved in sequencing predictions and their relevance for language. *Curr. Opin. Behav. Sci.* **21**, 145–153 (2018).

2. Milne, A. E., Petkov, C. I. & Wilson, B. Auditory and Visual Sequence Learning in Humans and Monkeys using an Artificial Grammar Learning Paradigm. *Neuroscience* **389**, 104–117 (2018).

3. Wilson, B. *et al.* Auditory Artificial Grammar Learning in Macaque and Marmoset Monkeys. *J. Neurosci.* **33**, 18825–18835 (2013).

4. Wilson, B. *et al.* Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nat. Commun.* **6**, 8901 (2015).

5. Saffran, J. *et al.* Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition* **107**, 479–500 (2008).

6. Spierings, M. J. & Ten Cate, C. Budgerigars and zebra finches differ in how they generalize in an artificial grammar learning experiment. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E3977-84 (2016).

7. van Heijningen, C. A. A., Chen, J., van Laatum, I., van der Hulst, B. & ten Cate, C. Rule learning by zebra finches in an artificial grammar learning task: which rule? *Anim. Cogn.* **16**, 165–175 (2013).

8. Wilson, B., Smith, K. & Petkov, C. I. Mixed-complexity artificial grammar learning in humans and macaque monkeys: evaluating learning strategies. *Eur. J. Neurosci.* **41**, 568–578 (2015).

9. Thiessen, E. D., Kronstein, A. T. & Hufnagle, D. G. The extraction and integration framework: A two-process account of statistical learning. *Psychol. Bull.* **139**, 792–814 (2013).

10. Deutsch, D. Grouping Mechanisms in Music. in *The Psychology of Music* 183–248 (Elsevier, 2013). doi:10.1016/B978-0-12-381460-9.00006-7

11. Povel, D.-J. & Essens, P. Perception of Temporal Patterns. *Music Percept. An Interdiscip. J.* **2**, 411–440 (1985).

12. Ravignani, A., Delgado, T. & Kirby, S. Musical evolution in the lab exhibits rhythmic universals. *Nat. Hum. Behav.* **1**, 0007 (2017).

13. Baldwin, D. A. & Baird, J. A. Discerning intentions in dynamic human action. *Trends Cogn. Sci.* **5**, 171–178 (2001).

14. Hard, B. M., Meyer, M. & Baldwin, D. Attention reorganizes as structure is detected in dynamic action. *Mem. Cognit.* **47**, 17–32 (2019).

15. Hard, B. M., Recchia, G. & Tversky, B. The shape of action. *J. Exp. Psychol. Gen.* **140**, 586–604 (2011).

16. Baldwin, D., Andersson, A., Saffran, J. & Meyer, M. Segmenting dynamic human action via statistical structure. *Cognition* **106**, 1382–1407 (2008).

17. Gambell, Timothy, Yang, C. *Word segmentation: Quick but not dirty.* (2006).

18. Frank, M. C., Goldwater, S., Griffiths, T. L. & Tenenbaum, J. B. Modeling human performance in statistical word segmentation. *Cognition* **117**, 107–125 (2010).

19. Perruchet, P. & Vinter, A. PARSER: A Model for Word Segmentation. *J. Mem. Lang.* **39**, 246–263 (1998).

20. Elman, J. L. Finding structure in time. *Cogn. Sci.* **14**, 179–211 (1990).

21. Endress, A. D. & Mehler, J. The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *J. Mem. Lang.* **60**, 351–367 (2009).

22. Endress, A. D. & Langus, A. Transitional probabilities count more than frequency, but

might not be used for memorization. *Cogn. Psychol.* **92**, 37–64 (2017).

23. Knowlton, B. J. & Squire, L. R. Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *J. Exp. Psychol. Learn. Mem. Cogn.* **22**, 169–181 (1996).

24. Erickson, L. C. & Thiessen, E. D. Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Dev. Rev.* **37**, 66–108 (2015).

25. Aslin, R. N. & Newport, E. L. Statistical learning: From acquiring specific items to forming general rules. *Curr. Dir. Psychol. Sci.* **21**, 170–176 (2012).

26. Perruchet, P. & Poulin-Charronnat, B. Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *J. Mem. Lang.* **66**, 807–818 (2012).

27. Saksida, A., Langus, A. & Nespor, M. Co-occurrence statistics as a language-dependent cue for speech segmentation. *Dev. Sci.* **20**, e12390 (2017).

28. Karuza, E. A. *et al.* The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain Lang.* **127**, 46–54 (2013).

29. McNealy, K., Mazziotta, J. C. & Dapretto, M. Cracking the Language Code: Neural Mechanisms Underlying Speech Parsing. *J. Neurosci.* **26**, 7629–7639 (2006).

30. Cunillera, T. *et al.* Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage* **48**, 541–553 (2009).

31. Blakemore, S. J., Rees, G. & Frith, C. D. How do we predict the consequences of our actions? A functional imaging study. *Neuropsychologia* **36**, 521–9 (1998).

32. Abla, D. & Okanoya, K. Statistical segmentation of tone sequences activates the left inferior frontal cortex: A near-infrared spectroscopy study. *Neuropsychologia* **46**, 2787–2795 (2008).

33. Petersson, K.-M., Folia, V. & Hagoort, P. What artificial grammar learning reveals about the neurobiology of syntax. *Brain Lang.* **120**, 83–95 (2012).

34. Bischoff-Grethe, A., Proper, S. M., Mao, H., Daniels, K. A. & Berns, G. S. Conscious and unconscious processing of nonverbal predictability in Wernicke's area. *J. Neurosci.* **20**, 1975–81 (2000).

35. Skosnik, P. D. *et al.* Neural correlates of artificial grammar learning. *Neuroimage* **17**, 1306–14 (2002).

36. Furl, N. *et al.* Neural prediction of higher-order auditory sequence statistics. *Neuroimage* **54**, 2267–2277 (2011).

37. Nastase, S., Iacovella, V. & Hasson, U. Uncertainty in visual and auditory series is coded by modality-general and modality-specific neural systems. *Hum. Brain Mapp.* **35**, 1111–

28 (2014).

38. Corbetta, M., Patel, G. & Shulman, G. L. The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron* **58**, 306–324 (2008).

39. Vilberg, K. L. & Rugg, M. D. Memory retrieval and the parietal cortex: A review of evidence from a dual-process perspective. *Neuropsychologia* **46**, 1787–1799 (2008).

40. Rugg, M. D., Johnson, J. D., Park, H. & Uncapher, M. R. Chapter 21 Encoding-retrieval overlap in human episodic memory: A functional neuroimaging perspective. in *Progress in brain research* **169**, 339–352 (2008).

41. Ordin, M., Polyanskaya, L., Laka, I. & Nespor, M. Cross-linguistic differences in the use of durational cues for the segmentation of a novel language. *Mem. Cognit.* **45**, 863–876 (2017).

42. Ordin, M., Polyanskaya, L., & Soto, D. Metacognitive processing in language learning tasks is affected by bilingualism. *J. Exp. Psychol. Learn. Mem. Cogn.*

43. Gonzalez-Castillo, J. *et al.* Imaging the spontaneous flow of thought: Distinct periods of cognition contribute to dynamic functional connectivity during rest. *Neuroimage* **202**, 116129 (2019).

44. Smith, S. M. Fast robust automated brain extraction. *Hum. Brain Mapp.* **17**, 143–155 (2002).

45. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63–72 (2009).

46. Rosenthal, C. R., Andrews, S. K., Antoniades, C. A., Kennard, C. & Soto, D. Learning and Recognition of a Non-conscious Sequence of Events in Human Primary Visual Cortex. *Curr. Biol.* **26**, 834–41 (2016).

47. Woolrich, M. W., Ripley, B. D., Brady, M. & Smith, S. M. Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. *Neuroimage* **14**, 1370–1386 (2001).

48. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–41 (2002).

49. Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M. & Smith, S. M. Multilevel linear modelling for FMRI group analysis using Bayesian inference. *Neuroimage* **21**, 1732–1747 (2004).

50. Beckmann, C. F., Jenkinson, M. & Smith, S. M. General multilevel linear modeling for group analysis in FMRI. *Neuroimage* **20**, 1052–1063 (2003).

51. Poldrack, R. A. *et al.* The Neural Correlates of Motor Skill Automaticity. *J. Neurosci.* **25**, 5356–5364 (2005).

52. Rosenthal, C. R., Roche-Kelly, E. E., Husain, M. & Kennard, C. Response-dependent

contributions of human primary motor cortex and angular gyrus to manual and perceptual sequence learning. *J. Neurosci.* **29**, 15115–25 (2009).

53. de Bourbon-Teles, J. *et al.* Thalamic Control of Human Attention Driven by Memory and Learning. *Curr. Biol.* **24**, 993–999 (2014).

54. Endress, A. D., Nespor, M. & Mehler, J. Perceptual and memory constraints on language acquisition. *Trends Cogn. Sci.* **13**, 348–353 (2009).

55. Prince, S. E., Daselaar, S. M. & Cabeza, R. Neural Correlates of Relational Memory: Successful Encoding and Retrieval of Semantic and Perceptual Associations. *J. Neurosci.* **25**, 1203–1210 (2005).

56. Turk-Browne, N. B., Scholl, B. J., Chun, M. M. & Johnson, M. K. Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness. *J. Cogn. Neurosci.* **21**, 1934–1945 (2009).

57. Gennari, S. P., Millman, R. E., Hymers, M. & Mattys, S. L. Anterior paracingulate and cingulate cortex mediates the effects of cognitive load on speech sound discrimination. *Neuroimage* **178**, 735–743 (2018).

58. Ciaramelli, E., Grady, C. L. & Moscovitch, M. Top-down and bottom-up attention to memory: A hypothesis (AtoM) on the role of the posterior parietal cortex in memory retrieval. *Neuropsychologia* **46**, 1828–1851 (2008).

59. Gehring, W. J. & Fencsik, D. E. Functions of the Medial Frontal Cortex in the Processing of Conflict and Errors. *J. Neurosci.* **21**, 9430–9437 (2001).

60. Scargle, J. D., Norris, J. P., Jackson, B. & Chiang, J. STUDIES IN ASTRONOMICAL TIME SERIES ANALYSIS. VI. BAYESIAN BLOCK REPRESENTATIONS. *Astrophys. J.* **764**, 167 (2013).

61. Alloy, L. B. & Tabachnik, N. Assessment of covariation by humans and animals: the joint influence of prior expectations and current situational information. *Psychol. Rev.* **91**, 112–49 (1984).

62. Jessen, F. *et al.* Activation of human language processing brain regions after the presentation of random letter strings demonstrated with event-related functional magnetic resonance imaging. *Neurosci. Lett.* **270**, 13–6 (1999).

63. Mechelli, A. *et al.* Structural plasticity in the bilingual brain. *Nature* **431**, 757–757 (2004).

64. Catani, M., Jones, D. K. & ffytche, D. H. Perisylvian language networks of the human brain. *Ann. Neurol.* **57**, 8–16 (2005).

65. Frey, S., Campbell, J. S. W., Pike, G. B. & Petrides, M. Dissociating the Human Language Pathways with High Angular Resolution Diffusion Fiber Tractography. *J. Neurosci.* **28**, 11435–11444 (2008).

66. Kelly, C. *et al.* Broca's region: linking human brain functional connectivity data and non-human primate tracing anatomy studies. *Eur. J. Neurosci.* **32**, 383–398 (2010).

67. Seghier, M. L. The Angular Gyrus. *Neurosci.* **19**, 43–61 (2013).

68. Paraskevopoulos, E., Chalas, N. & Bamidis, P. Functional connectivity of the cortical network supporting statistical learning in musicians and non-musicians: an MEG study. *Sci. Rep.* **7**, 16268 (2017).

FIGURE CAPTIONS:

Figure 1. Example of a learning phase run (above), where 25.92-second pseudorandom blocks are interspersed with 181.44-second structured blocks; and the structure of the recognition trial (below).

Figure 2. Percentage of endorsed trials for words, non-words and phantoms. Error bars show 95% CI.

Figure 3. Mean confidence rating (weighted average) assigned to the trails with accepted and rejected words, phantoms and non-words. Error bars show 95%CI.

Figure 4. BOLD responses during the learning phase (Z >2.3, *P* = 0.05 corrected). Brain regions showing BOLD response increases in structured vs pseudorandom chunks in the course of training.

Figure 5. BOLD responses during the recognition phase (Z >2.3, *P* = 0.05 corrected. **(A)** Brain regions showing increased response for words accepted relative to nonword rejected (*w_acc > nonw_rej).* Correctly identified words elicit larger activation in the LIFG compared to correctly identified non-words. **(B)** Brain regions showing increased activity on trials with words accepted compared to phantoms rejected (*w_acc* > *ph*_rej). Endorsed words, relative to rejected phantoms, elicit larger activation in the angular gyrus and intra-parietal sulcus, the anterior division of the cingulate gyrus, and the posterior division of the right STG.

Figure 6. Illustration of the brain activity maps that showed increased activity during learning compared to the recognition phase (Z >2.3, *P* = 0.05 corrected).

Table 1. The list of words, phantoms and non-words used in the experiment

Words	**Phantoms**	**Non-words**
ROSENU	PASENU	ROTIMO
ROKAFA	LEKAFA	SEPAKO
PASETI	ROSETI	FALUSA
LEKATI	ROKATI	FOLERI
PAMONU	PAMOFA	TAMUPE
LEMOFA	LEMONU	NIKANU
PERIKO	MURIKO	NURIPE
MURIFO	LUTASA	FOLUKA
PETASA	PERIFO	NIMUKO
LUTAFO	PETAFO	MOPARO
MUNIKO	MUNISA	LESATI
LUNISA	LUNIKO	TASEFA

Table 2. Location of peaks related to the increase in the activation difference between structured and random chunks

Cluster	Extent (voxels)	Anatomical region	Z max	x	y	z
1	1008	Anterior division of cingulate gyrus, paracingulate gyrus, superior frontal gyrus	3.46	14	38	24
2	1002	Superior temporal gyrus, left (slightly extending to middle temporal gyrus)	3.97	-60	-14	-2
3	777	Superior temporal gyrus, right	3.57	64	-8	-2

Table 3. Location of peak activation differences (with NMI coordinates of the activation peaks) for the relevant contrasts in the recognition phase

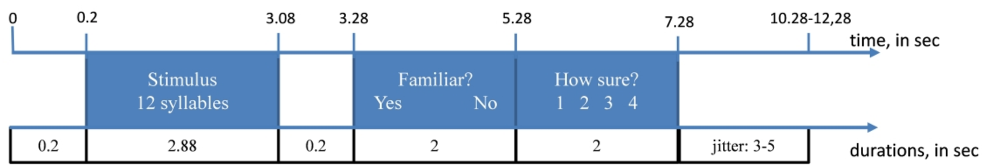
Contrast	Cluster	Extent (voxels)	Anatomical region	Z max	X	y	z
w_acc > nonw_rej	1	6207	LIFG (par opercularis extending to par triangularis), i.e., BA44 extending to BA45.	3.76	-38	18	22
w_acc > ph_rej	1	297	Anterior division of the cingulate gyrus, paracingulate gyrus	3.49	-2	26	36
	2	292	Posterior divisions of the right superior temporal and middle temporal gyri	3.7	64	-26	-2
	3	227	Anterior division of the left intra-parietal sulcus, angular gyrus	3.73	-34	-50	34

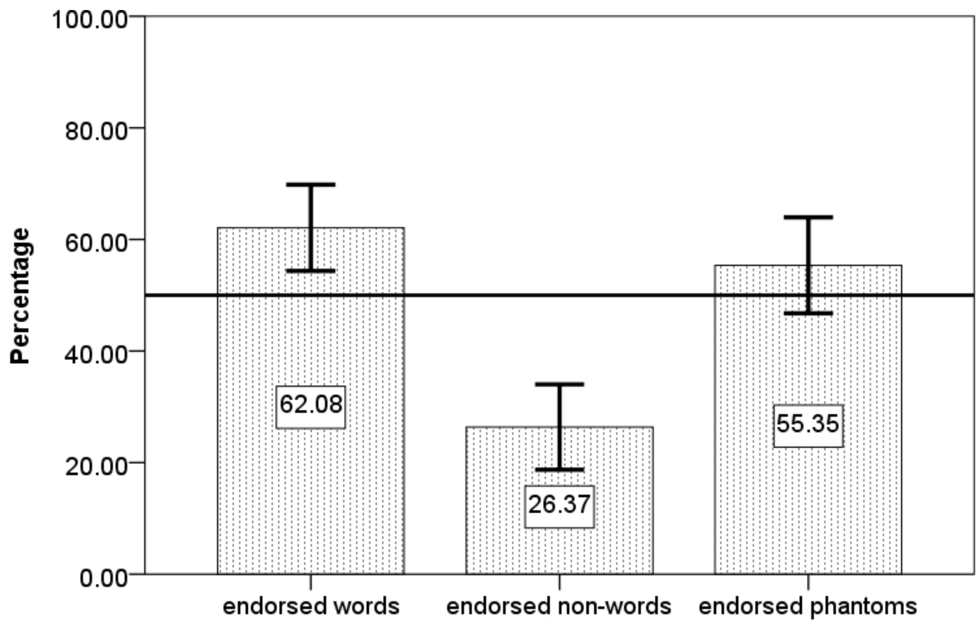
Table 4. Location of peak activation differences (with NMI coordinates of the activation peaks) for the *learning* > *recognition* contrast

Cluster	Extent (voxels)	Anatomical region	Z max	x	y	z
1	688	The superior temporal gyrus, slightly extending to the middle temporal gyrus, left.	3.88	-60	-14	-2
2	699	Anterior division of the cingulate gyrus, the paracingulate gyrus, right, the superior frontal gyrus.	3.51	14	40	24

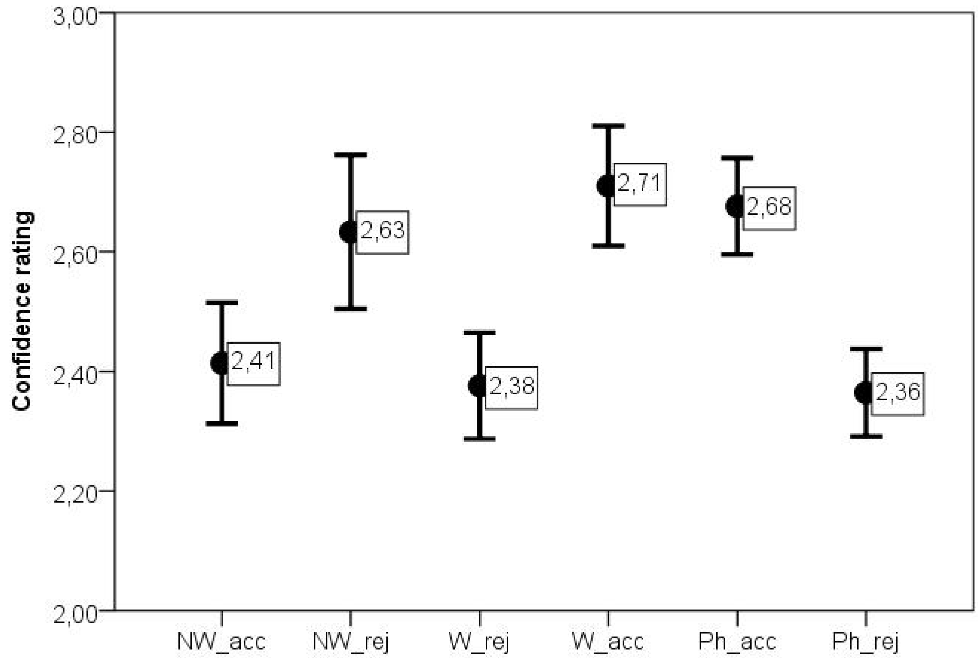
Example of a learning run

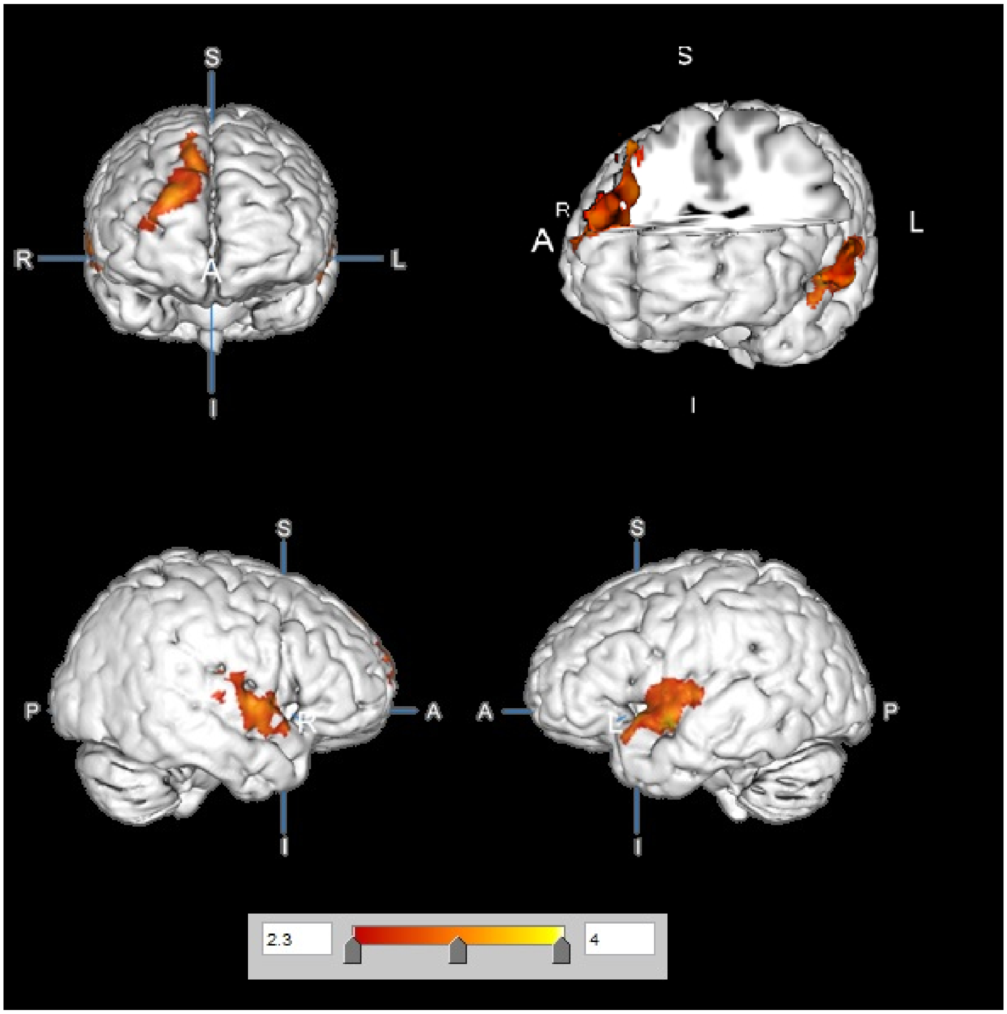
Structure of the recognition trial

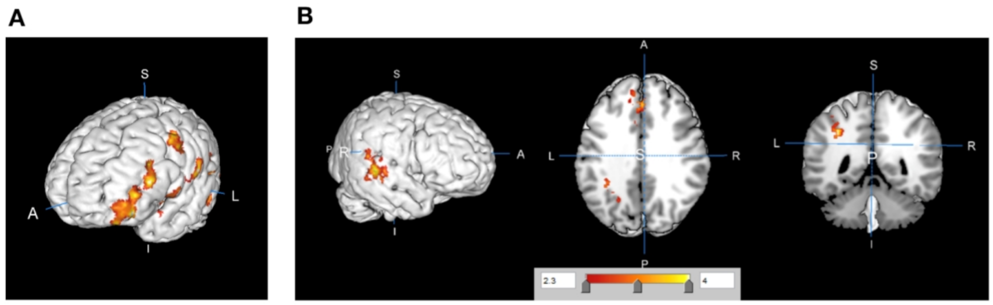




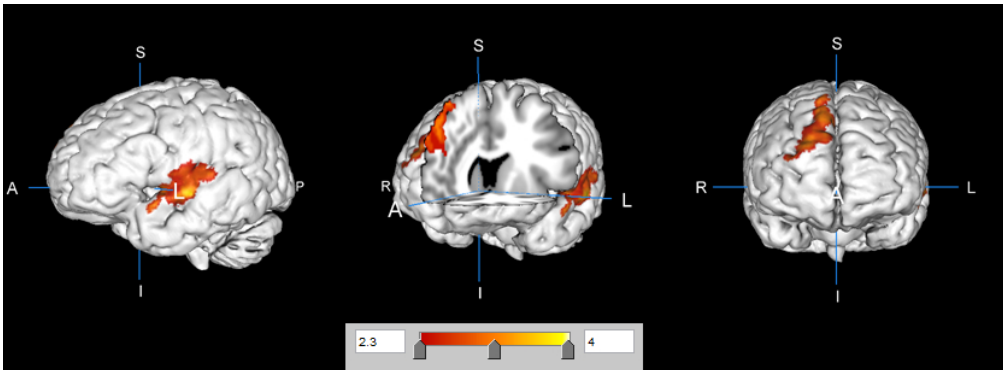
271x169mm (72 x 72 DPI)


Confidence rating
3,00
2,80
2,60
2,40
2,20
2,00
2,41
2,63
2,38
2,71
2,68
2,36
NW_acc
NW_rej
W_rej
W_acc
Ph_acc
Ph_rej


S
R
L
A
I
S
R
A
L
I
S
P
A
I
S
A
P
L
I
2.3
4



184x58mm (220 x 220 DPI)


S
A
P
S
R
A
L
I
S
R
L
A
I
I
2.3
4