The Journal of Neuroscience

https://jneurosci.msubmit.net

JN-RM-0144-17R2

Learning predictive statistics: strategies and brain mechanisms

Zoe Kourtzi, University of Cambridge
Rui Wang, Chinese Academy of Sciences
Yuan Shen, Univ of Birmingham
Peter Tino, univ of Birmingham
Andrew Welchman, University of Cambridge

Commercial Interest:

Learning predictive statistics: strategies and brain mechanisms

Rui Wang[1,2], Yuan Shen[3,4], Peter Tino[4], Andrew Welchman[1], Zoe Kourtzi*[1]

[1] Department of Psychology, University of Cambridge, Cambridge, United Kingdom
[2] Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[3] Department of Mathematical Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, China
[4] School of Computer Science, University of Birmingham, Birmingham, United Kingdom

Correspondence: Zoe Kourtzi, Department of Psychology, University of Cambridge, Cambridge, UK, Email: zk240@cam.ac.uk

- Abbreviated title: Brain mechanisms for learning temporal structure
- Number of pages: 38
- Number of figures: 8
- Number of tables: 1
- Number of words for Abstract: 194
- Number of words for Introduction: 513
- Number of words for Discussion: 837

Acknowledgements: We would like to thank Caroline di Bernardi Luft for help with data collection and Matthew Dexter for help with software development.

Funding: This work was supported by grants to PT from Engineering and Physical Sciences Research Council [EP/L000296/1], ZK from the Biotechnology and Biological Sciences Research Council [H012508], the Leverhulme Trust [RF-2011-378] and the [European Community's] Seventh Framework Programme [FP7/2007-2013] under agreement PITN-GA 2011-290011, AEW from the Wellcome Trust (095183/Z/10/Z)

Abstract

When immersed in a new environment we are challenged to decipher initially incomprehensible streams of sensory information. Yet, quite rapidly, the brain finds structure and meaning in these incoming signals, helping us to predict and prepare ourselves for future actions. This skill relies on extracting the statistics of event streams in the environment that contain regularities of variable complexity: from simple repetitive patterns to complex probabilistic combinations. Here, we test the brain mechanisms that mediate our ability to adapt to the environment's statistics and predict upcoming events. By combining behavioral training and multi-session fMRI in human participants (male and female), we track the cortico-striatal mechanisms that mediate learning of temporal sequences as they change in structure complexity. We show that learning of predictive structures relates to individual decision strategy; that is, selecting the most probable outcome in a given context (maximizing) vs. matching the exact sequence statistics. These strategies engage distinct human brain regions: maximizing engages dorsolateral prefrontal, cingulate, sensory-motor regions and basal ganglia (dorsal caudate, putamen), while matching engages occipito-temporal regions (including the hippocampus) and basal ganglia (ventral caudate). Our findings provide evidence for distinct cortico-striatal mechanisms that facilitate our ability to extract behaviorally-relevant statistics to make predictions.

Significance Statement

Making predictions about future events relies on interpreting streams of information that may initially appear incomprehensible. Past work has studied how humans identify repetitive patterns and associative pairings. However, the natural environment contains regularities that vary in complexity: from simple repetition to complex probabilistic combinations. Here, we

combine behavior and multi-session fMRI to track the brain mechanisms that mediate our ability to adapt to changes in the environment's statistics. We provide evidence for an alternate route for learning complex temporal statistics: extracting the most probable outcome in a given context is implemented by interactions between executive and motor cortico-striatal mechanisms compared to visual cortico-striatal circuits (including hippocampal cortex) that support learning of the exact temporal statistics.

Introduction

Making predictions about future events challenges us to extract structure from streams of sensory signals that initially appear incomprehensible. Typically, event structures in the natural environment contain regularities of variable complexity: from simple repetitive patterns to more complex probabilistic combinations. For example, when learning a new piece of music or a new language, we extract simple repetitive patterns (e.g. tones, syllables) as well as more complex contingencies (e.g., melodies or phoneme pairs) that determine the probability with which events occur. Learning to extract these statistics allows us to interpret incoming signals rapidly, and predict upcoming events. Despite the fundamental importance of this type of statistical learning for sensory interpretation and prediction, we know surprisingly little about its neural basis.

Previous work on statistical learning has focused on simple repetitive patterns or associative pairings. Behavioral studies provide evidence that mere exposure (i.e. without explicit feedback) to co-occuring stimuli can drive learning of contingencies (for reviews see: (Perruchet and Pacton, 2006; Aslin and Newport, 2012)). For example, observers become familiar with structured patterns after exposure to items (e.g. shapes, tones or syllables) that co-occur spatially or appear in a temporal sequence (Saffran et al., 1999; Chun, 2000; Fiser and Aslin, 2002; Turk-Browne et al., 2005). Here, we investigate the functional brain mechanisms that mediate our ability to adapt to changes in the environment's statistics and learn behaviorally-relevant structures for making predictions.

We combine behavioral measures with multi-session fMRI (before and after training) to examine the neural mechanisms that mediate learning of temporal sequences that change in their statistics: from repetitive patterns to more complex probabilistic contingencies. To do so unencumbered by past experience, we tested participants with sequences of unfamiliar

symbols, where the complexity of the sequence structure changed unbeknownst to the participants **(Figure 1)**. We increased sequence complexity by manipulating the memory order (i.e. context length) of the Markov model used to generate the sequences. In particular, we presented participants first with sequences that were determined by frequency statistics (i.e. occurrence probability per symbol), and then by more complex context-based statistics (i.e. the probability of a given symbol appearing depends on the preceding symbol). Participants performed a prediction task in which they indicated which symbol they expected to appear following exposure to a sequence of variable-length. Following previous statistical learning paradigms, participants were exposed to the sequences without trial-by-trial feedback.

Our behavioral results show that individuals adapt to the environment's statistics; that is, they are able to extract predictive structures of different complexity. Further, we show that learning of predictive structures relates to individual decision strategy; that is individuals differed in their decision strategies, favouring either probability maximization (i.e. extracting the most probable outcome in a given context) or matching the exact sequence statistics. We used this variability in decision strategy to interrogate fMRI activity. We find that distinct cortico-striatal mechanisms mediate the two strategies: matching engages occipito-temporal regions (including the hippocampus) and ventral caudate, while maximizing engages dorsolateral prefrontal, cingulate, sensory-motor regions and basal ganglia (dorsal caudate, putamen). This provides evidence for differentiated cortico-striatal mechanisms that support learning of behaviorally-relevant statistics for making predictions.

Figure 1

Methods

Observers

Thirty-four participants (mean age = 21.8 years, male and female) participated in the experiments (main experiment: n=23; control experiment: n=11). The data from two participants were excluded from further imaging analysis due to excessive head movement (greater than 3mm). All observers were naïve to the aim of the study, had normal or corrected-to-normal vision and gave written informed consent. This study was approved by the University of Birmingham Ethics Committee.

Stimuli

Stimuli comprised four symbols chosen from Ndjuká syllabary (Turk-Browne et al., 2009)(**Figure 1a**). These symbols were highly discriminable from each other and were unfamiliar to the observers. Each symbol subtended 8.5^{o} of visual angle and was presented in black on mid-grey background. Experiments were controlled using Matlab and the Psychophysics toolbox 3 (Brainard, 1997; Pelli, 1997). For the behavioral training sessions, stimuli were presented on a 21-inch CRT monitor (ViewSonic P225f 1280 x1024 pixel, 85 Hz frame rate) at a distance of 45 cm. For the pre- and post-training fMRI scans, stimuli were presented using a projector and a mirror set-up (1280 x 1024 pixel, 60 Hz frame rate) at viewing distance of 67.5 cm. The physical size of the stimuli was adjusted so that angular size was constant during behavioral and scanning sessions.

Sequence design

To generate probabilistic sequences of different complexity, we used a temporal Markov model and manipulated the memory order of the sequence, which we refer to as the context length.

The Markov model consists of a series of symbols, where the symbol at time *i* is determined

probabilistically by the previous 'k' symbols. We refer to the symbol presented at time i, $s(i)$, as the *target* and to the preceding k-tuple of symbols ($s(i-1)$, $s(i-2)$, …, $s(i-k)$) as the *context*. The value of 'k' is the order or level of the sequence:

$$P(s(i) \mid s(i-1), s(i-2), \ldots, s(1)) = P(s(i) \mid s(i-1), s(i-2), \ldots, s(i-k)),\ k<i$$

The simplest $k=0^{th}$ order model is a random memory-less source. This generates, at each time point i, a symbol according to symbol probability $P(s)$, without taking account of the previously generated symbols.

The order $k=1$ Markov model generates symbol $s(i)$ at each time i conditional on the previously generated symbol $s(i-1)$. This introduces a memory in the sequence; that is, the probability of a particular symbol at time i strongly depends on the preceding symbol $s(i-1)$. Unconditional symbol probabilities $P(s(i))$ for the case $k=0$ are replaced with conditional ones, $P(s(i) \mid s(i-1))$.

At each time point, the symbol that follows a given context is determined probabilistically, making the Markov sequences stochastic. The underlying Markov model can be represented through the associated context-conditional target probabilities. We used 4 symbols that we refer to as stimuli A, B, C and D. The correspondence between stimuli and symbols was counterbalanced across participants.

For level-0, the Markov model was based on the probability of symbol occurrence: one symbol had a high probability of occurrence, one low probability, while the remaining two symbols appeared rarely (**Figure 1b**). For example, the probabilities of occurrence for the four symbols A, B, C, and D were 0.18, 0.72, 0.05 and 0.05, respectively. Presentation of a given symbol was independent of the stimuli that preceded it.

For level-1, the target depended on the immediately preceding stimulus (**Figure 1b**). Given a context (the last seen symbol) only one of two targets could follow; one had a high probability of being presented and the other a low probability (e.g., 80% vs. 20%). For

example, when Symbol A was presented, only symbols B or C were allowed to follow, and B had a higher probability of occurrence than C.

Task design

We tested learning of temporal structures that differed in their complexity; that is, sequences determined by simple frequency statistics (level-0) and more complex sequences defined by context-based statistics (level-1). To define the complexity of our sequences, we quantified the average past-future mutual information in the sequences generated by stochastic sources (Grassberger, 1986), providing a statistic that has been applied in a number of probabilistic contexts (e.g., (Shaw, 1984; Li, 1991)). For Markov models of order 0 or 1, complexity is expressed as the difference between the entropy of the marginal symbol distribution and the entropy rate of the Markov chain (Li, 1991). This measure quantifies the average reduction in uncertainty of the next symbol in a sequence when the memory of the generating source is taken into account. For 0-order Markov models, the complexity is 0, as the source itself is memory-less. For Markov models of order 1, conditioning on the last symbol will reduce the uncertainty. For example, for the 1st order Markov model we used, the marginal symbol probabilities are equal, resulting in entropy close to the maximum value of 2 bits. However, conditional on the last symbol, only two symbols are allowed with unequal probabilities, resulting in lower entropy rate and therefore higher complexity (1.28).

To investigate whether participants adapt to changes in the temporal structure, we ensured that the sequences across levels were matched for properties (i.e. number or identify of symbols) other than complexity. Further, we designed the stochastic sources from which the sequences were generated so that the context-conditional uncertainty remained highly similar across levels. In particular, for the zero-order source, only two symbols were likely to occur most of the time; the remaining two symbols had very low probability (0.05); this was introduced to ensure that there was no difference in the number of symbols presented across

levels. Of the two dominant symbols, one was more probable (probability 0.72) than the other (probability 0.18). This structure is preserved in Markov chain of order 1, where conditional on the previous symbols, only two symbols were allowed to follow, one with higher probability (0.80) than the other (0.20). This ensures that the structure of the generated sequences across levels differed predominantly in memory order (i.e. context length) rather than context-conditional probability.

Procedure

Observers were initially familiarized with the task through a brief practice session (8 minutes) with random sequences (i.e. all four symbols were presented with equal probability 25% in a random order). Following this, observers participated in multiple behavioral training and fMRI scanning sessions that were conducted on different days **(Figure 1c).** Participants were trained with structured sequences and tested with both structured and random sequences to ensure that training was specific to the trained sequences.

In the first scanning session, participants were presented with zero- and first-order sequences and random sequences. Observers were then trained with zero-order sequences, and subsequently with first-order sequences. For each level, observers completed a minimum of 3 and a maximum of 5 training sessions (840-1400 trials). Training at each level ended when participants reached plateau performance (i.e. performance did not change significantly for two sessions). A post-training scanning session followed training per level (i.e. on the following day after completion of training) during which observers were presented with structured sequences determined by the statistics of the trained level and random sequences. The mean time interval (±standard error) between the pre-training session and the final test session was 23.5±0.5days.

Psychophysical training

Each training session comprised five blocks of structured sequences (56 trials per block) and

lasted one hour. To ensure that sequences in each block were representative of the Markov model order per level, we generated 10,000 Markov sequences per level comprising 672 stimuli per sequence. We then estimated the Kullback-Leibler divergence (KL divergence) between each example sequence and the generating source. In particular, for level-0 sequences this was defined as:

$$KL = \sum_{target} Q(target)\ log(\frac{Q(target)}{P(target)}),$$

and for level-1 sequences this was defined as:

$$KL = \sum_{context} Q(context)\ \sum_{target} Q(taget|context)\ log(\frac{Q(target|context)}{P(target|context)}),$$

where *P( )* refers to probabilities or conditional probabilities derived from the presented sequences, and *Q( )* refers to those specified by the source. We selected fifty sequences with the lowest KL divergence (i.e. these sequences matched closely the Markov model per level). The sequences presented to the participants during the experiments were selected randomly from this sequence set.

For each trial, a sequence of 8-14 stimuli appeared in the center of the screen, one at a time in a continuous stream, each for 300ms followed by a central white fixation dot (ISI) for 500ms (**Figure 1a**). This variable trial length ensured that observers maintained attention during the whole trial. Each block comprised equal number of trials with the same number of stimuli. The end of each trial was indicated by a red dot cue that was presented for 500ms. Following this, all four symbols were shown in a 2x2 grid. The positions of test stimuli were randomized from trial to trial. Observers were asked to indicate which symbol they expected to appear following the preceding sequence by pressing a key corresponding to the location of the predicted symbol. Observers learned a stimulus-key mapping during the familiarization phase: key '8', '9', '5' and '6' in the number pad corresponded to the four positions of the test stimuli - upper left, upper right, lower left and lower right, respectively. After the observer's response, a white circle appeared on the selected stimulus for 300ms to indicate

the observer's choice, followed by a fixation dot for 150ms (ITI) before the start of the next trial. If no response was made within 2s, a null response was recorded and the next trial started. Participants were given feedback (i.e. score in the form of Performance Index, see Data Analysis) at the end of each block – rather than per-trial error feedback– that motivated them to continue with training.

Scanning sessions

The pre-training scanning session (Pre) included six runs (i.e. three runs per level) the order of which was randomized across participants. Scanning sessions after training per level (denoted as Post 0, Post 1) included nine runs of structured sequences determined by the same statistics as the corresponding trained level and random sequences. Each run comprised five blocks of structured and five blocks of random sequences presented in a random counterbalanced order (2 trials per blocks; a total of 10 structured and 10 random trials per run), with an additional two 16-s fixation blocks, one at the beginning and one at the end of each run. Each run comprised 110 stimuli for structured sequences and 110 stimuli for random sequences. Each trial comprised a sequence of 10 stimuli that were presented for 250ms each, separated by a blank interval during which a white fixation dot was presented for 250ms. Following the sequence, a fixation screen (central red dot) appeared for 4s before the test display (comprising four test stimuli) appeared for 1.5s. Observers were asked to indicate which symbol they expected to appear following the preceding sequence by pressing a key corresponding to the location of the predicted symbol. A white fixation was then presented for 5.5s before the start of the next trial. In contrast to the training sessions, no feedback was given during scanning.

fMRI data acquisition

The experiments were conducted at the Birmingham University Imaging Centre using a 3-T Philips Achieva MRI scanner. T2*-weighted functional and T1-weighted anatomical (1 × 1 ×

1 mm resolution, slices=175) data were collected with a 32-channel SENSE head coil. Echo planar imaging data (gradient echo-pulse sequences) were acquired from 32 slices (whole brain coverage; TR = 2000 ms; TE = 35 ms; 2.5 × 2.5 × 4-mm resolution).

Behavioral data analysis

Performance index:

We assessed participant responses in a probabilistic manner. For each context, we computed the absolute Euclidean distance between the distribution of participant responses and the distribution of presented targets estimated across 56 trials per block:

$$\mathrm{AbDist}(\mathrm{context}) = \sum_{\mathrm{target}} |P_{\mathrm{resp}}(\mathrm{target}|\mathrm{context}) - P_{\mathrm{pres}}(\mathrm{target}|\mathrm{context})|$$

where the sum is over targets from the symbol set A, B, C and D. We estimate AbDist per context for each block. We quantified the minimum overlap between these two distributions by computing a performance index per context:

$$\mathrm{PI}(\mathrm{context}) = \sum_{\mathrm{target}} \min\left(P_{\mathrm{resp}}(\mathrm{target}|\mathrm{context}), P_{\mathrm{pres}}(\mathrm{target}|\mathrm{context})\right)$$

Note that $\mathrm{PI}(\mathrm{context}) = 1 - \mathrm{AbDist}(\mathrm{context})/2$. The overall performance index is then computed as the average of the performance indices across contexts, PI(context), weighted by the corresponding stationary context probabilities:

$$\mathrm{PI} = \sum_{\mathrm{context}} \mathrm{PI}(\mathrm{context}) \cdot P(\mathrm{context})$$

To compare across different levels, we defined a normalized PI measure that quantifies participant performance relative to random guessing. We computed a random guess baseline; i.e. performance index $\mathrm{PI}_{\mathrm{rand}}$ that reflects participant responses to targets with a) equal probability of 25% for each target per trial for level-0, ($\mathrm{PI}_{\mathrm{rand}} = 0.53$); b) equal probability for each target for a given context for level-1 ($\mathrm{PI}_{\mathrm{rand}} = 0.45$). To correct for differences in random-guess baselines across levels, we subtracted the random guess baseline from the performance index ($\mathrm{PI}_{\mathrm{normalized}} = \mathrm{PI} - \mathrm{PI}_{\mathrm{rand}}$).

Strategy choice and strategy Index: To quantify each observer's strategy, we compared individual participant response distributions (response-based model) to two baseline models: (i) probability matching, where probability distributions are derived from the Markov models that generated the presented sequences (Model-matching) and (ii) a probability maximization model, where only the single most likely outcome is allowed for each context (Model-maximization). We used Kullback-Leiber (KL) divergence to compare the response distribution to each of these two models. KL is defined as follows:

$$KL = \sum_{target} M(target)\ log(\frac{M(target)}{R(target)})$$

for level-0 model and

$$KL = \sum_{context} M(context) \sum_{target} M(target|context)\ log(\frac{M(target|context)}{R(target)|context})$$

for level-1 model where *R( )* and *M( )* denote the probability distribution or conditional probability distribution derived from the human responses and the models (i.e. probability matching or maximization) respectively, across all the conditions.

We quantified the difference between the KL divergence from Model-matching to the response-based model and the KL divergence from Model-maximization to the response-based model. We refer to this quantity as strategy choice indicated by ΔKL(Model-maximization, Model-matching). Negative strategy choice values indicate a strategy closer to matching, while positive values indicate a strategy closer to maximization. We computed strategy choice per training block, resulting in a strategy curve across training for each individual participant. We then derived an individual strategy index by calculating the integral of each participant's strategy curve and subtracting it from the integral of the exact matching curve, as defined by Model-matching across training. We defined the integral curve difference (ICD) between individual strategy and exact matching as the individual strategy

index. We used this index to investigate the relationship of individual strategy and fMRI signals.

fMRI data analysis

Data pre-processing: MRI data were processed using Brain Voyager QX (Brain Innovation). T1-weighted anatomical data were used for co-registration, three-dimensional cortex reconstruction, inflation, and flattening. Preprocessing of the functional data involved slice-scan time correction, head motion correction, temporal high-pass filtering (3 cycles), and removal of linear trends. Spatial smoothing (Gaussian filter; 5-mm FWHM kernel) was performed for group random-effect analysis. The functional images were aligned to anatomical data, and the complete data were transformed into Talairach space. For each observer, the functional imaging data between sessions were co-aligned by registering all volumes of each observer to the first functional volume acquired during the first session.

Whole-Brain General Linear Model: BOLD responses for each trial comprising structured or random sequences were modeled separately for each session using a general linear model (GLM). To search for brain regions that showed learning-dependent changes across sessions, we constructed a multiple regression design matrix that included the two stimulus conditions (structured vs. random sequences) for each of the scanning sessions (Pre, Post 0, Post 1) as regressors. Each regressor was time-locked to trial onset and included a range of volumes (Figures 3, 4: 5 volumes, Figure 5b: 3 volumes). To remove residual motion artefacts, the six zero-centered head movement parameters were also included as regressors. Serial correlations were corrected using a second order autoregressive model AR(2). The resulting parameter estimates (β value) were used in a voxel-wise mixed design ANOVA with sequence (structured vs. random) and scanning session (Pre, Post 0, Post 1). Statistical maps were cluster threshold corrected ($p<0.005$) using Monte Carlo simulations (5000 iterations) (Forman et al., 1995; Goebel et al., 2006) for multiple comparison correction that confirmed

FWE (family wise error) threshold of p=0.05. Note that our results also hold for a more conservative threshold ($p<0.001$)—as recommended by recent studies (Woo et al., 2014; Eklund et al., 2016)—but small volume correction is required for small structures (i.e. putamen) at this threshold.

Covariance analysis: To examine the relationship between brain activation and observers' performance, we conducted a voxel-wise covariance analysis. In particular, we used individual strategy index as covariate in a GLM model of fMRI responses. That is, for each voxel, we correlated fMRI signal difference between structured and random sequences before vs. after training with the strategy index. We calculated a Pearson correlation coefficient (R) for each voxel across the whole brain and identified voxel clusters showing significant correlations ($p < 0.05$, cluster threshold corrected). Positive correlations indicate increased activations after training that relate to maximization, while negative correlations indicate increased activations after training that relate to matching, as negative strategy index indicates matching.

Results

Behavioral results

Previous studies have compared learning of different spatiotemporal contingencies in separate experiments across different participant groups (Fiser and Aslin, 2002, 2005). Here, to investigate whether individuals extract changes in structure, we presented the same participants with sequences that changed in complexity unbeknownst to them (**Figure 1a**). We parameterized sequence complexity based on the memory-order of the Markov models used to generate the sequences (see Methods); that is, the degree to which the presentation of a symbol depended on the history of previously presented symbols (**Figure 1b**). We first presented participants with simple zero-order sequences (level-0) followed by more complex

first-order sequences (level-1) (**Figure 1c**), as previous work has shown that temporal dependencies are more difficult to learn as their length increases (van den Bos and Poletiek, 2008) and training with simple dependencies may facilitate learning of more complex contingencies (Antoniou et al., 2016). Zero-order sequences (level-0) were context-less; that is, the presentation of each symbol depended only on the probability of occurrence of each symbol. For first-order sequences (level-1), the presentation of a particular symbol was conditionally dependent on the previously presented symbol (i.e. context length of one).

As the sequences we employed were probabilistic, we developed a probabilistic measure to assess participants' performance in the prediction task. Specifically, we computed a Performance index (PI) that indicates how closely the distribution of participant responses matched the probability distribution of the presented symbols. This is preferable to a simple measure of accuracy because the probabilistic nature of the sequences means that the 'correct' upcoming symbol is not uniquely specified; thus, designating a particular choice as correct or incorrect is often arbitrary.

Comparing normalized performance (i.e. after subtracting performance based on random guessing) before and after training per level (**Figure 2**) showed that observers improved substantially and learnt the probabilistic structures (i.e. mean improvement higher than 20% for both levels). A repeated measures ANOVA with Session (pre, post) and Level (Level-0, Level-1) showed a significant effect of session ($F(1,20)=82.0$, $p<0.001$) but no significant effect of level ($F(1,20)<1$, $p=0.358$) nor a significant interaction ($F(1,20)<1$ $p=0.664$), indicating that observers improved similarly at both levels through training. Interestingly, performance during the pre-training test session was higher than random guessing ($F(1,20)=42.8$, $p<0.001$), suggesting fast learning of structured sequences consistent with the learning time course reported in previous perceptual learning studies (Karni and Sagi, 1993). However, improvement continued during training across blocks; that is mean performance for

the last two training blocks was significantly higher than mean performance for the first two training blocks ($F(1,20)=12.8$, $p=0.002$).

We then tested whether this learning-dependent improvement was specific to the trained structured sequences. First, we compared performance accuracy (i.e. proportion of correctly predicted trials based on the pre-defined sequences) for structured and random sequences. A repeated-measures ANOVA showed a significant interaction of Session (pre, post), and sequence type (structured vs. random) for level-0 ($F(1,20)=24.1$, $p<0.001$) and level-1 ($F(1,20)=54.5$, $p<0.001$), suggesting that learning improvement was specific to structured sequences. Second, we conducted a no-training control experiment, during which participants (n=11) were tested in two separate behavioral sessions but did not participate in any training sessions. The two test sessions were spaced apart by a period (mean of 27.9±1.9 days) comparable to the main experiment (23.3 days on average). Our results showed that there were no significant differences in performance between the two test sessions. In particular, a repeated measures ANOVA with Session (session 1, session 2) and Level (Level-0, Level-1) did not show any significant effect of session ($F(1,10)<1$, $p=0.736$) or level ($F(1,10)=1.84$, $p=0.205$), nor a significant interaction ($F(1,10)=1.16$, $p=0.308$). These results suggest that the improvement we observed in the main experiment was specific to training rather than simply due to repeating the test session twice (before and after training). Comparing performance index between experiments (main vs. no-training control experiment) showed a significant interaction between Experiment and Session (Level-0: $F(1,30)=15.1$, $p=0.001$, Level-1: $F(1,30)=7.95$, $p=0.008$), consistent with training-induced behavioral improvement.

Figure 2

fMRI analysis: Learning-dependent activation changes

To investigate the brain mechanisms that mediate our ability to adapt to changes in temporal statistics, we performed fMRI on participants before and after training on each level with structured and random sequences. To assess learning-dependent changes in fMRI signals, we conducted a whole brain voxel-wise GLM analysis (RFX group analysis). In particular, we tested for brain regions that showed a significant interaction ($p<0.005$, cluster threshold corrected) between sequence (structured vs. random) and scanning session (Pre, Post 0, Post 1). This analysis revealed a network of dorsal frontal, cingulate, posterior parietal and temporal regions, as well as subcortical (basal ganglia), and cerebellar regions (**Figure 3a, Table 1**).

We next asked whether functional signals in these regions change from learning frequency (level-0) to learning context-based statistics (level-1) over time. In particular, we compared fMRI responses for structured and random sequences before and after training for each level (level-0 vs. level-1) separately. For each participant and brain region identified by the GLM analysis, we calculated normalized fMRI responses (i.e. percent signal change (PSC) index); that is, we subtracted mean fMRI responses to random sequences from mean fMRI responses to structured sequences and divided by the average fMRI responses to random sequences. Note that this PSC analysis is complementary to the GLM analysis used to define regions of interest; it was conducted separately for each level, whereas the GLM tested for differences across sessions (i.e. Pre, Post 0, Post 1) rather than levels.

Comparing normalized fMRI responses before and after training for level-0 (**Figure 3b**) showed that bilateral dorsal frontal regions (medial: SFG: superior frontal gyrus; MeFG: medial frontal gyrus, lateral: MFG: middle frontal gyrus, PrG: precentral gyrus and IFG: inferior frontal gyrus) and right posterior parietal regions (IPL: inferior parietal lobule, AnG; Angular gyrus, SMG: supramarginal gyrus) were involved in learning frequency-based statistics. These regions showed increased fMRI responses to structured sequences during the

pre-training scanning session in contrast to decreased responses after training (i.e. post-training scanning session). In particular, a repeated measures ANOVA with session (pre, post) and ROI (regions of interest) showed a significant main effect of session in frontal ($F(1,20)=7.59$, $p=0.012$) and posterior parietal ($F(1,20)=6.58$, $p=0.018$) regions.

In contrast, learning context-based statistics (level-1) engaged dorsal medial frontal (SFG and MeFG), limbic (CG: cingulate gyrus, ACC: anterior cingulate cortex) and subcortical (Pu: putamen) areas (**Figure 3b**). Similar to the fMRI activation patterns for Level-0, dorsal frontal regions showed enhanced responses to structured compared to random sequences for the pre-training scan that decreased after training. This was supported by a repeated measures ANOVA that showed a significant session effect (frontal: $F(1,20)=6.36$, $p=0.020$; limbic: $F(1,20)=5.36$, $p=0.031$). In contrast, we observed the opposite pattern of results in putamen (paired t-test, $t(20)=-3.31$, $p=0.003$), that is, enhanced activations for structured sequences after training. Activation patterns differed significantly between putamen and frontal-limbic regions (i.e. significant interactions of Region and Session: frontal vs. putamen, $F(1,20)=16.22$, $p<0.001$; limbic vs. Putamen, $F(1,20)=16.34$, $p<0.001$). In a complementary analysis to the GLM analysis, comparing activations across levels showed significant differences in prefrontal regions (interaction of session and level, $F(1,20)=4.83$, $p=0.040$), right posterior parietal regions (main effect of level, $F(1,20)=7.41$, $p=0.013$) and putamen (main effect of level, $F(1,20)=4.56$, $p=0.045$). Consistent with the GLM analysis, these results support differential involvement of fronto-parietal and striatal regions in learning frequency compared to context-based statistics.

Interestingly, the GLM analysis showed activation changes across sessions in the visual cortex (IOG: inferior occipital gyrus, MOG: middle occipital gyrus, LiG: lingual gyrus). Comparing fMRI responses in these regions across sessions did not show any significant differences for either of the two levels (level-0: $F(1,20)<1$, $p=0.429$; level-1: $F(1,20)<1$,

p=0.531), suggesting that fMRI responses for structured sequences did not change significantly with training in the visual cortex. For learning frequency statistics (level-0) visual cortex showed stronger activations for random than structured sequences (i.e. negative PSC index values) both before (main effect of sequence, $F(1,20)=6.04$, $p=0.023$) and after ($F(1,20)=32.7$, $p<0.001$) training, suggesting decreased activation due to repetition (i.e. repetition suppression) of symbols that appeared more frequently in structured than random sequences (Summerfield and Egner, 2009). This effect was not observed for first-order sequences (level-1) (before training, $F(1,20)<1$, $p=0.981$; after training, $F(1,20)=1.87$, $p=0.187$), consistent with higher repetition of single symbols in zero-order than first-order sequences.

Figure 3

Next, we asked whether the differences we observed in the activation patterns between levels were due to differences in sequence predictability. To measure sequence predictability, we computed the entropy rate of the probability distribution of all possible sequences. For level-0, the entropy rate is defined as the entropy of the stationary distribution of symbols in the sequence. For level-1, the entropy rate is a weighted sum of the entropies of all context-conditional distributions where the weights are given by the stationary distribution of contexts. We calculated the entropy rate for each sequence; we then conducted the whole brain voxel-wise GLM analysis using entropy rate as regressor. This analysis showed significant interactions ($p<0.001$, cluster threshold corrected) between sequence (structured vs. random) and scanning session (Pre, Post 0, Post 1) in similar regions as the main analysis (**Figure 4a**), making it unlikely that our results were confounded by differences in sequence predictability between levels.

Comparing normalized fMRI responses before and after training (**Figure 4b**) for level-0, we observed increased fMRI responses to structured sequences before than after training

476 ($F(1,20)=5.18$, $p=0.034$) in bilateral frontal regions (SFG: superior frontal gyrus; PrG:
477 precentral gyrus and IFG: inferior frontal gyrus). In contrast, learning context-based statistics
478 (level-1) engaged dorsal frontal (SFG), limbic (ACC: anterior cingulate cortex) and
479 subcortical (Pu: putamen) areas. Dorsal frontal and limbic regions showed enhanced
480 responses to structured compared to random sequences for the pre-training scan that
481 decreased after training ($F(1,20)=5.76$, $p=0.026$). In contrast, putamen showed enhanced
482 activations for structured sequences after training (paired t-test, $t(20)=-2.78$, $p=0.012$).
483 Activation patterns differed significantly between putamen and frontal-limbic regions (i.e.
484 significant interactions of Region and Session: $F(1,20)=13.9$, $p<0.001$) in support of
485 differential involvement of frontal and striatal regions in learning temporal statistics.

486 **Figure 4**

487 Our results so far suggest that dorsal cortico-striatal mechanisms mediate learning of
488 behaviorally-relevant statistics. In particular, fronto-parietal and cingulate regions showed
489 higher fMRI responses for structured than random sequences during the pre-training scan.
490 This is consistent with the role of dorsal prefrontal cortex in decision making (Heekeren et
491 al., 2008; Rushworth and Behrens, 2008) and predictive coding (Monchi et al., 2001; Bar,
492 2009); that is, processes that are involved in both learning of frequency and context-based
493 statistics. Further, our results show that cingulate cortex is involved in learning more complex
494 context-based statistics that may relate to its involvement in learning under increased
495 uncertainty (Kahnt et al., 2011; Nastase et al., 2014). Higher fMRI responses for structured
496 sequences in these regions at the beginning of training may reflect processing of novel
497 structures (i.e. temporal regularities in the form of single- or paired-item repetition).
498 Significantly higher performance for structured sequences than random guessing during the
499 first scanning session suggests that participants extract these statistics early in the training.
500 Interestingly, fMRI responses for structured sequences decreased as these sequences became

familiar with training. This decreased signals can be understood in the context of repetition suppression previously observed for predictable events (Raichle et al., 1994; Den Ouden et al., 2009; Summerfield and Egner, 2009; Alink et al., 2010; Kok et al., 2012). In contrast, dorsal striatal regions (i.e. putamen) – that have been implicated in learning probabilistic associations (Rauch et al., 1997; Poldrack and Packard, 2003)– showed higher fMRI responses for structured compared to random sequences after training with first-order sequences, suggesting representations of context-target contingencies acquired through training.

Control analyses

We conducted a number of additional analyses and experiments to help rule out alternative explanations of our results.

First, we asked whether the differences we observed in fMRI responses between structured and random sequences were due to the participants attending more to the structured sequences either as the novel stimulus before training or the familiar stimulus after training. Comparing response times to structured and random sequences in the pre- and post-training session (3 way mixed design ANOVA: *session* X *sequence* X *level*) showed decreased response times after training (main effect of *session*: $F(1,20) = 8.63$, $p = 0.008$), but no significant differences between structured and random sequences (main effect of *sequence,* $F(1,20) = 0.152$, $p = 0.700$), suggesting that participants engaged with the task when both structured and random sequences were presented. Importantly, there was no significant interaction between session, sequence and level ($F(1,20) = 1.72$, $p = 0.205$), suggesting that differences in activation patterns across levels could not be simply due to differences in attention or task difficulty. Further, analysis of eye movement data collected during scanning did not show any significant differences between structured and random sequences for level-0

or level-1. There were no significant interactions observed ($p > 0.10$), suggesting that it is unlikely that our findings were significantly confounded by eye movements.

Second, we tested whether the learning-dependent fMRI changes we observed could be confounded by differences in the number of training sessions across participants. Training duration varied from 3-5 sessions per level across participants, with most participants completing four training sessions (level 0, n = 12; level 1, n = 17) before reaching plateau performance. An ANCOVA analysis on the behavioral data using the number of training sessions as covariate did not show any significant interactions between session and number of training sessions (level-0: $F(1,19) = 0.479$, $p = 0.497$; level-1: $F(1,19) = 0.089$, $p = 0.768$). Similar analysis on the fMRI data did not show any significant interaction between session and number of training sessions (level 0: frontal, $F(1,19) = 0.001$, $p = 0.874$, parietal, $F(1,19) = 0.447$, $p = 0.512$; level 1: frontal, $F(1,19) = 0.473$, $p = 0.500$, limbic, $F(1,19) = 0.705$, $p = 0.412$, subcortical regions, $F(1,19) = 3.53$, $p = 0.076$). Taken together these analyses suggest that it is unlikely that our fMRI results were confounded by differences in training duration across participants.

Third, we asked whether the activation patterns we observed relate to learning-dependent changes in the representation of the trained sequences or simply the participants' responses. In our design, the inter-stimulus interval jitter in each trial is too short to isolate the fMRI signal per stimulus in the sequence. However, the design of the paradigm allows us to analyze our fMRI data for sequence presentation separately from participant prediction. First, we compared PSC for the first two volumes related to the presented sequences and the fourth and fifth volume related to the participants' prediction (i.e. the third volume was not included in this analysis, as the sequences lasted 2.5 volumes). This analysis (**Figure 5a**) showed that activation patterns for fMRI signals related to the sequence presentation and the participants' prediction were similar to those observed in our main analysis **(Figure 3b, 4b).** In particular,

we observed a significant effect of Session (i.e. pre- vs. post-training) (level-0: frontal: $F(1,20)=4.97$, $p=0.037$; level-1: frontal-limbic: $F(1,20)=5.95$, $p=0.024$, putamen: $F(1,20)=7.29$, $p=0.014$), but no significant effect of processing stage (i.e. sequence vs. prediction) (level-0: frontal: $F(1,20)=0.004$, $p=0.951$: level-1: frontal-limbic: $F(1,20)=0.399$, $p=0.535$: putamen: $F(1,20)=3.29$, $p=0.085$). There was no significant interaction of session and processing stage (level-0: frontal: $F(1,20)=0.003$, $p=0.954$; level-1: frontal-limbic: $F(1,20)=0.496$, $p=0.490$; putamen: $F(1,20)=1.68$, $p=0.209$). Second, a whole brain voxel-wise GLM analysis using only the volumes that corresponded to the sequence presentation showed significant interactions ($p<0.001$, cluster threshold corrected) between sequence (structured vs. random) and scanning session (Pre, Post 0, Post 1) in similar regions as the main analysis (**Figure 5b**). Taken together, these analyses of fMRI signals related to the sequence presentation showed similar activation patterns as the main analysis (**Figure 3a**) that included fMRI signals from both the sequence presentation and the participant prediction. Thus, the learning-dependent changes we observed in the main analysis relate to the sequence structure and could not be simply driven by the participants' prediction or response, as fMRI signals related to the sequence presentation were recorded before the participants responded to the test stimulus.

Figure 5

Response strategies: matching vs. maximization

Previous work (Shanks et al., 2002; Rieskamp and Otto, 2006; Eckstein et al., 2013; Acerbi et al., 2014; Fulvio et al., 2014; Murray et al., 2015) on probabilistic learning and decision-making has proposed that individuals use two possible response strategies when making a choice: matching vs. maximization. Observers have been shown to either match their choices stochastically according to the underlying input statistics or to maximize their reward by

574 selecting the most probable positively rewarded outcomes. In the context of our task, as the
575 Markov models that generated stimulus sequences were stochastic, participants needed to
576 learn the probabilities of different outcomes to succeed in the prediction task. It is possible
577 that participants used probability maximization whereby they always select the most probable
578 outcome in a particular context. Alternatively, participants might learn the relative
579 probabilities of each symbol (e.g. p(A)=0.18; p(B)=0.72, p(C)=0.05; p(D)=0.05) and respond
580 so as to reproduce this distribution, a strategy referred to as probability matching.

581 To quantify the participants' strategies, we computed a strategy index that indicates
582 participant's preference (on a continuous scale) for responding using probability matching vs.
583 maximization. **Figure 6** illustrates individual strategy at the beginning (first two blocks) and
584 end of training (last two blocks). Comparing individual strategy across levels showed
585 significantly higher values after training for level-1 compared to level-0 ($F(1,20)=26.2$,
586 $p<0.001$). This shift in individual strategy was evident mainly after training ($F(1,20)=35.8$,
587 $p<0.001$); that is participants shifted more towards maximization when learning context-
588 based rather than frequency statistics. Note, that this relationship was not confounded by
589 differences in performance, as there were no significant correlations (Level 0: $r=0.31$,
590 $p=0.17$; Level 1: $r=0.22$, $p=0.34$) of performance index at the end of training (mean PI for the
591 last two blocks of training) and strategy index. Interestingly, despite greater maximization for
592 more complex structures than frequency statistics, we note that participants did not achieve
593 optimal maximization performance. Maximization is typically observed under supervised or
594 reinforcement learning paradigms (Shanks et al., 2002), so it is perhaps not surprising that
595 our participants did not achieve exact maximization as trial-by-trial feedback was not
596 provided.

597 **Figure 6**

598 ***fMRI covariance analysis with strategy***

599 To investigate the relationship between brain activations and individual strategy, we
600 conducted a voxel-wise GLM covariance analyses. In particular, we correlated learning-
601 dependent changes in fMRI signal (post-pre training PSC) for structured (compared to
602 random) sequences with individual strategy. We calculated a Pearson correlation coefficient
603 (R) for each voxel across the whole brain and identified voxel clusters showing significant
604 correlations ($p<0.05$) for learning frequency (level-0) and context-based statistics (level-1),
605 respectively. Positive correlations indicate increased activations after training that relate to
606 maximization, while negative correlations indicate increased activations after training that
607 relate to matching, as negative strategy values indicate strategy towards matching.

608 First, we observed negative correlations between learning-dependent fMRI changes and
609 strategy index in occipito-temporal (including hippocampal regions), basal ganglia (ventral
610 caudate) and thalamic regions (**Figure 7**). These correlations indicate that increased
611 activations for structured sequences after training in these regions relate to matching. Further,
612 these correlations were observed for both levels suggesting that learning frequency or
613 context-based statistics by matching involves regions in visual cortico-striatal circuits that
614 have been previously implicated in the implicit learning of temporal sequences (Hindy et al.,
615 2016; Rosenthal et al., 2016) and novel categories (Ashby and Maddox, 2005; Seger, 2013).
616 In particular, previous work has implicated the striatum and the medial temporal lobe (i.e.
617 hippocampus) (Rauch et al., 1997; Poldrack and Packard, 2003; Schendan et al., 2003; Cools
618 et al., 2004; Gheysen et al., 2011; Rose et al., 2011; Schapiro et al., 2012; Hsieh et al., 2014)
619 in learning probabilistic associations. Further, medial temporal cortex has been implicated in
620 explicit rule-based categorization, whereas caudate in categorization based on information
621 integration (Nomura et al., 2007).

Figure 7

In contrast, we observed positive correlations between learning-dependent fMRI changes and strategy index, indicating that increased activations for structured sequences after training relate to maximizing (**Figure 8**). In particular, for Level 0, we observed positive correlations in dorsolateral prefrontal areas (MFG/IFG), the dorsal caudate and the cingulate (including anterior cingulate) cortex. For Level 1, we observed positive correlations in dorsolateral prefrontal (MFG/IFG), and posterior parietal regions, as well as cingulate and temporal cortex. Interestingly, we also observed positive correlations for sensory-motor cortex (pre-central and post-central gyrus) and basal ganglia (putamen). Our results are consistent with the role of prefrontal and cingulate cortex in decision-making, monitoring performance, correcting errors, and switching between associations and strategies. Previous work on humans and animals emphasizes the role of the caudate in switching between strategies (Monchi et al., 2001; Cools et al., 2004; Seger and Cincotta, 2006), and learning after a rule reversal (Cools et al., 2002; Pasupathy and Miller, 2005). This tonic and fast learning in the caudate is thought to train slower learning mechanisms in the frontal cortex that may facilitate generalization and abstraction of learned associations. Finally, putamen—known to be involved in skilled and habitual performance (Daw et al., 2005; Balleine and O'Doherty, 2010)—may facilitate learning by maximizing. That is, once participants have extracted the most probable outcome for a given context they may then habitually select it as the predicted outcome.

Figure 8

Discussion

Here, we investigate the brain mechanisms that medicate our ability to adapt to changes in the environment's statistics and make predictions. To test how individuals extract structure

changes, we manipulate the complexity of temporal sequences during training from simple frequency to context-based statistics. Our results provide evidence for dissociated cortico-striatal mechanisms that mediate our ability to extract behaviorally-relevant statistics. We find that fronto-parietal activity decreases for frequency-based learning, while context-based learning is associated with decreased fronto-cingulate activity and increased striatal activity. Decreased fMRI signals in fronto-parietal circuits can be understood in the context of predictive coding as repetition suppression for predictable events (Raichle et al., 1994; Den Ouden et al., 2009; Summerfield and Egner, 2009; Alink et al., 2010; Kok et al., 2012). In contrast, increased fMRI signals in putamen –that is implicated in learning probabilistic associations (Rauch et al., 1997; Poldrack and Packard, 2003)– suggest representations of predictive structures acquired through training.

Importantly, our approach allows us to track participants' predictions and their decision strategies during training. We demonstrate that learning predictive structures relates to decision strategies; that is, learning complex structures relates to extracting the most probable target per context (i.e. maximizing) than matching the exact sequence statistics. Importantly, these decision strategies engage distinct cortico-striatal circuits: performance based on probability matching engages occipito-temporal and basal ganglia (ventral caudate) regions, while performance based on maximizing engages dorsolateral prefrontal, cingulate, sensory-motor regions and basal ganglia (dorsal caudate, putamen). Recent work has focused on the role of hippocampus in learning temporal sequences (Hsieh et al., 2014; Rosenthal et al., 2016) and predictive associations (Hindy et al., 2016). Our findings suggest an alternate route to learning via maximizing that is implemented by interactions between executive and motor cortico-striatal mechanisms rather than visual cortico-striatal circuits (including hippocampal cortex) that support learning by matching.

Previous studies have implicated these cortico-striatal circuits in reinforcement learning (e.g. for reviews (Robbins, 2007; Balleine and O'Doherty, 2010)). We show that learning predictive statistics may proceed without explicit trial-by-trial feedback and involve interactions between cortico-striatal circuits similar to those known to support reward-based learning (Alexander et al., 1986; Lawrence et al., 1998). In particular, we show that dorsal fronto-parietal regions are involved in extracting novel regularities, monitoring and adjusting strategy throughout training. In contrast, striatal regions represent context-based statistics learned through bootstrap training (i.e. multiple sessions of exposure to structured sequences) that may optimize the selection of the most probable outcome in a given context. Previous work investigating learning of sequential contingencies in the context of the serial reaction time task suggests that striatal vs. hippocampal circuits relate to distinct error-driven learning processes and operate at different learning rates (Bornstein and Daw, 2012). In particular, fast learning was shown to engage striatal regions (i.e. putamen), whereas slow learning the hippocampus. Although our paradigm does not dissociate learning time course from structure complexity, it is possible that learning of temporal structures proceeds from cortico-striatal to hippocampal circuits.

Further, we considered whether the learning we observed occurred in an incidental manner or involved explicit knowledge of the underlying sequence structure. Previous studies have suggested that learning of regularities may occur implicitly in a range of tasks: visuomotor sequence learning (Nissen and Bullemer, 1987; Seger, 1994; Schwarb and Schumacher, 2012), artificial grammar learning (Reber, 1967), probabilistic category learning (Knowlton et al., 1994), and contextual cue learning (Chun and Jiang, 1998). This work has focused on implicit measures of sequence learning, such as familiarity judgments or reaction times. In contrast, our paradigm allows us to directly test whether exposure to temporal sequences facilitates the observers' ability to explicitly predict the identity of the next stimulus in a

sequence. Although, our experimental design makes it unlikely that the participants memorized specific stimulus positions or the full sequences, debriefing the participants suggests that most extracted some high probability symbols or context-target combinations. Thus, it is possible that prolonged exposure to probabilistic structures (i.e. multiple sessions in contrast to single exposure sessions typically used in statistical learning studies) in combination with prediction judgments (Dale et al., 2012) may evoke some explicit knowledge of temporal structures, in contrast to implicit measures of anticipation typically used in statistical learning studies.

Finally, previous work has implicated additional brain regions related to learning modality-specific regularities (Nastase et al., 2014); that is visual cortex is implicated in learning visual statistical regularities (Aizenstein et al., 2004; Turk-Browne et al., 2010; Meyer and Olson, 2011), while inferior frontal and temporal regions in learning temporal regularities related to music and language (Bahlmann et al., 2009; Leaver et al., 2009; Karuza et al., 2013; Koelsch et al., 2013). Our results provide evidence for cortico-striatal mechanisms that mediate learning of predictive statistics. We speculate that these mechanisms may mediate domain-general learning of complex structures that can be specialized to support higher cognitive functions such as, learning music or language.

Figures

Figure 1. Trial and sequence design. (a) The trial design: 8-14 stimuli were presented sequentially followed by a cue and the test display. (b) Sequence design: Markov models

718 comprising two levels of complexity. For the zero-order model (level-0): different states (A,
719 B, C, D) are assigned to four symbols with different frequencies. For the first order model
720 (level-1), a diagram indicates states (circles) and transitional probabilities (black arrow: high
721 probability (e.g. 80%); grey arrow: low probability (e.g. 20%)). Transitional probabilities are
722 shown in a four-by-four conditional probability matrix, with rows indicating temporal context
723 and columns indicating the corresponding target. (c) Experimental protocol. Observers
724 underwent multiple days of behavioral training first with zero-order sequences and then with
725 first-order sequences. For each level, observers completed 3-5 training sessions (an average
726 of 4 sessions is shown for illustration purposes). Three fMRI scanning sessions were
727 conducted before (Pre) and immediately following training per level (Post0, Post1).

728 **Figure 2.** Behavioral performance. (a) Mean performance index (PI) across participants for
729 test (open symbols) and training (solid symbols) blocks for level-0 and level-1. Data is fitted
730 (least squares non-linear fit) across training blocks. Random guess baseline is indicated by
731 dotted lines. (b) Normalized performance index during scanning. Data is shown before (grey
732 bars) and after (black bars) training for each level. Error bars indicate standard error of the
733 mean.

734 **Figure 3.** fMRI results. (a) GLM maps for the 2-way interaction between Scanning session
735 (Pre, Post0, Post1) and Sequence (structured vs. random), at $p < 0.005$ (cluster threshold
736 corrected). Only the first 5 volumes were included in the analysis that correspond to the
737 presentation of sequence, the participants' prediction and the test display presentation, to
738 avoid confounding the results by the participants' response. Similar results were observed at a
739 more conservative threshold ($p<0.001$) but small volume correction was necessary for small
740 structures (i.e. putamen) at this threshold. (b) PSC index (percent signal change for structured
741 sequences compared to random sequences) before and after training for level-0 and level-1.
742 Data is shown for ROIs that showed a significant interaction between Session (pre- vs. post-

training) and Sequence (structured vs. random). SFG: superior frontal gyrus; MeFG: medial frontal gyrus; MFG: middle frontal gyrus; IFG: inferior frontal gyrus; PrG: precentral gyrus; IPL: inferior parietal lobule; AnG: Angular gyrus; SMG: supramarginal gyrus;CG: cingulate gyrus; ACC: anterior cingulate cortex; Pu: putamen; IOG: inferior occipital gyrus; MOG: middle occipital gyrus; LiG: lingual gyrus. Error bars denote SEM. Note that different number of runs were scanned before and after training (i.e. pre-training scan comprised 3 runs per level while post-training scans comprised nine runs per level). To compare equal amount of data before and after training, we selected three out of the nine runs from each post-training scan; that is, we divided each session into two time periods and selected randomly one run per time period to match the order in which data was collected during the pre-training scan. Whole brain voxel-wise GLM analysis showed significant interactions for sequence (structured vs. random) and scanning session (Pre, Post0, Post1) in frontal, parietal and subcortical regions, consistent with our main result.

Figure 4. fMRI results controlled for differences in sequence entropy across levels. (a) GLM maps ($p < 0.001$, cluster threshold corrected) for 2-way interaction between Scanning session (Pre, Post0, Post1) and Sequence (structured vs. random) including entropy rate as regressor. (b) PSC index before and after training for level-0 and level-1. Error bars denote SEM. Data is shown for ROIs that showed a significant interaction between Session (pre- vs. post-training) and Sequence (structured vs. random). SFG: superior frontal gyrus; IFG: inferior frontal gyrus; PrG: precentral gyrus; ACC: anterior cingulate cortex; Pu: putamen.

Figure 5. fMRI results for sequence presentation and participants' prediction. (a) PSC index for sequence presentation (volumes 1-2), and participant prediction (volumes 4-5) before and after training for level-0 and level-1. Data are shown for the representative ROIs from Figure 4b. Error bars denote SEM. (b) GLM maps for the 2-way interaction between Scanning

767 session and Sequence, at $p < 0.005$ (cluster threshold corrected) using only the volumes that
768 correspond to sequence presentation.

769 **Figure 6.** Strategy choice. Strategy choice is shown at the beginning (first two runs) and end
770 (last two runs) of training for level-0 (squares) and level-1 (circles). Open symbols indicate
771 individual participant data; closed symbols indicate mean date per level. Strategy choice was
772 measured by comparing participant responses to two possible strategies: matching (i.e.
773 predicting the presented target distribution) vs. maximization (i.e. predicting the high
774 probability targets per context). Negative values indicate a strategy closer to matching,
775 whereas positive values indicate a strategy closer to maximization. Error bars indicate
776 standard error of the mean.

777 **Figure 7.** Brain activations correlating with matching. Covariance analysis showing
778 significant ($p< 0.05$, cluster threshold corrected) negative correlations (R correlation
779 coefficient) between individual strategy index and learning-dependent fMRI change (i.e. after
780 vs. before training) for (a) level-0 and (b) level-1. Whole brain maps and plots showing
781 negative correlations between strategy index and PSC (percent signal change) index change
782 (post- vs. pre-training) for representative ROIs, as derived from the covariance analysis (note
783 that these correlation plots are only presented for demonstration purposes; no additional
784 statistical analysis was performed in these ROIs following the covariance analysis to avoid
785 circularity). Cd: caudate: (b) and (t) indicating body and tail, respectively; Th: thalamus;
786 PHG: parahippocampal gyrus; Hipp: hippocampus; LiG: lingual gyrus.

787 **Figure 8.** Brain activations correlating with maximization. Covariance analysis showing
788 significant ($p< 0.05$, cluster threshold corrected) positive correlations (R correlation
789 coefficient) between individual strategy index and learning-dependent fMRI change (i.e. after
790 vs. before training) for (a) level-0 and (b) level-1. Whole brain maps and plots showing

positive correlations between strategy index and PSC (percent signal change) index change (post- vs. pre-training) for representative ROIs, as derived from the covariance analysis (note that these correlation plots are only presented for demonstration purposes; no additional statistical analysis was performed in these ROIs following the covariance analysis to avoid circularity). MFG/IFG: middle/inferior frontal gyrus; CG: cingulate gyrus; ACC: anterior cingulate cortex; PrG: precentral gyrus; IPL: inferior parietal lobule; MTG: middle temporal gyrus; Cd: caudate; Pu: putamen.

References

Acerbi L, Vijayakumar S, Wolpert DM (2014) On the origins of suboptimality in human probabilistic inference. PLoS Comput Biol 10:e1003661.

Aizenstein HJ, Stenger VA, Cochran J, Clark K, Johnson M, Nebes RD, Carter CS (2004) Regional brain activation during concurrent implicit and explicit sequence learning. Cerebral Cortex 14:199-208.

Alexander GE, DeLong MR, Strick PL (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annual review of neuroscience 9:357-381.

Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L (2010) Stimulus predictability reduces responses in primary visual cortex. The Journal of Neuroscience 30:2960-2966.

Antoniou M, Ettlinger M, Wong PCM (2016) Complexity, Training Paradigm Design, and the Contribution of Memory Subsystems to Grammar Learning. PLOS ONE 11:e0158812.

Ashby FG, Maddox WT (2005) Human category learning. Annu Rev Psychol 56:149-178.

Aslin RN, Newport EL (2012) Statistical learning from acquiring specific items to forming general rules. Current directions in psychological science 21:170-176.

Bahlmann J, Schubotz RI, Mueller JL, Koester D, Friederici AD (2009) Neural circuits of hierarchical visuo-spatial sequence processing. Brain research 1298:161-170.

Balleine BW, O'Doherty JP (2010) Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology 35:48-69.

Bar M (2009) The proactive brain: memory for predictions. Philosophical Transactions of the Royal Society B: Biological Sciences 364:1235-1243.

Bornstein AM, Daw ND (2012) Dissociating hippocampal and striatal contributions to sequential prediction learning. European Journal of Neuroscience 35:1011-1023.

Brainard DH (1997) The psychophysics toolbox. Spatial vision 10:433-436.

Chun MM (2000) Contextual cueing of visual attention. Trends in Cognitive Sciences 4:170-178.

Chun MM, Jiang Y (1998) Contextual cueing: Implicit learning and memory of visual context guides spatial attention. Cognitive psychology 36:28-71.

Cools R, Clark L, Robbins TW (2004) Differential responses in human striatum and prefrontal cortex to changes in object and rule relevance. The Journal of Neuroscience 24:1129-1135.

Cools R, Clark L, Owen AM, Robbins TW (2002) Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. The Journal of Neuroscience 22:4563-4567.

Dale R, Duran ND, Morehead JR (2012) Prediction during statistical learning, and implications for the implicit/explicit divide. Advances in Cognitive Psychology 8:196.

Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature neuroscience 8:1704-1711.

Den Ouden HE, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009) A dual role for prediction error in associative learning. Cerebral Cortex 19:1175-1185.

Eckstein MP, Mack SC, Liston DB, Bogush L, Menzel R, Krauzlis RJ (2013) Rethinking human visual attention: Spatial cueing effects and optimality of decisions by honeybees, monkeys and humans. Vision Res 85:5-19.

Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proceedings of the National Academy of Sciences of the United States of America 113:7900-7905.

Fiser J, Aslin RN (2002) Statistical learning of higher-order temporal structure from visual shape sequences. Journal of Experimental Psychology: Learning, Memory, and Cognition 28:458.

Fiser J, Aslin RN (2005) Encoding multielement scenes: statistical learning of visual feature hierarchies. Journal of Experimental Psychology: General 134:521.

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved Assessment of Significant Activation in Functional Magnetic Resonance Imaging (fMRI): Use of a Cluster-Size Threshold. Magnetic Resonance in Medicine 33:636-647.

Fulvio JM, Green CS, Schrater PR (2014) Task-Specific Response Strategy Selection on the Basis of Recent Training Experience. PLoS computational biology 10:e1003425.

Gheysen F, Van Opstal F, Roggeman C, Van Waelvelde H, Fias W (2011) The neural basis of implicit perceptual sequence learning. Front Hum Neurosci 5:137.

Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. Human Brain Mapping 27:392-401.

Grassberger P (1986) Toward a quantitative theory of self-generated complexity. International Journal of Theoretical Physics 25:907-938.

Heekeren HR, Marrett S, Ungerleider LG (2008) The neural systems that mediate human perceptual decision making. Nature reviews neuroscience 9:467-479.

Hindy NC, Ng FY, Turk-Browne NB (2016) Linking pattern completion in the hippocampus to predictive coding in visual cortex. Nature Neuroscience.

Hsieh L-T, Gruber MJ, Jenkins LJ, Ranganath C (2014) Hippocampal activity patterns carry information about objects in temporal context. Neuron 81:1165-1178.

Kahnt T, Grueschow M, Speck O, Haynes J-D (2011) Perceptual learning and decision-making in human medial frontal cortex. Neuron 70:549-559.

Karni A, Sagi D (1993) The time course of learning a visual skill. Nature 365:250-252.

Karuza EA, Newport EL, Aslin RN, Starling SJ, Tivarus ME, Bavelier D (2013) The neural correlates of statistical learning in a word segmentation task: An fMRI study. Brain and language 127:46-54.

Knowlton BJ, Squire LR, Gluck MA (1994) Probabilistic classification learning in amnesia. Learning & Memory 1:106-120.

Koelsch S, Rohrmeier M, Torrecuso R, Jentschke S (2013) Processing of hierarchical syntactic structure in music. Proceedings of the National Academy of Sciences 110:15443-15448.

Kok P, Jehee JF, de Lange FP (2012) Less is more: expectation sharpens representations in the primary visual cortex. Neuron 75:265-270.

Lawrence AD, Sahakian BJ, Robbins TW (1998) Cognitive functions and corticostriatal circuits: insights from Huntington's disease. Trends in cognitive sciences 2:379-388.

Leaver AM, Van Lare J, Zielinski B, Halpern AR, Rauschecker JP (2009) Brain activation during anticipation of sound sequences. The Journal of Neuroscience 29:2477-2485.

Li W (1991) On the relationship between complexity and entropy for Markov chains and regular languages. Complex Systems 5:381-399.

Meyer T, Olson CR (2011) Statistical learning of visual transitions in monkey inferotemporal cortex. Proceedings of the National Academy of Sciences 108:19401-19406.

Monchi O, Petrides M, Petre V, Worsley K, Dagher A (2001) Wisconsin Card Sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. The Journal of Neuroscience 21:7733-7741.

Murray RF, Patel K, Yee A (2015) Posterior Probability Matching and Human Perceptual Decision Making. PLoS Comput Biol 11:e1004342.

Nastase S, Iacovella V, Hasson U (2014) Uncertainty in visual and auditory series is coded by modality-general and modality-specific neural systems. Human brain mapping 35:1111-1128.

Nissen MJ, Bullemer P (1987) Attentional requirements of learning: Evidence from performance measures. Cognitive psychology 19:1-32.

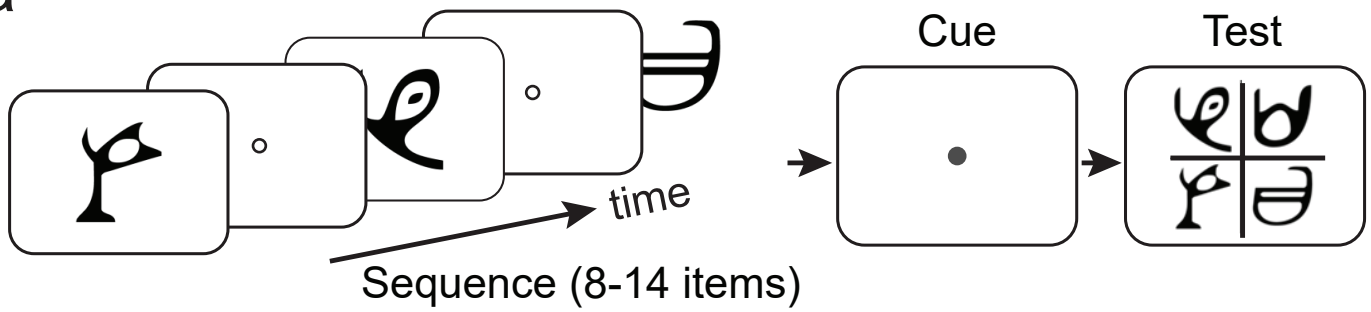
Nomura E, Maddox W, Filoteo J, Ing A, Gitelman D, Parrish T, Mesulam M, Reber P (2007) Neural correlates of rule-based and information-integration visual category learning. Cerebral Cortex 17:37-43.
Pasupathy A, Miller EK (2005) Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature 433:873-876.
Pelli DG (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial vision 10:437-442.
Perruchet P, Pacton S (2006) Implicit learning and statistical learning: one phenomenon, two approaches. Trends Cogn Sci 10:233-238.
Poldrack RA, Packard MG (2003) Competition among multiple memory systems: converging evidence from animal and human brain studies. Neuropsychologia 41:245-251.
Raichle ME, Fiez JA, Videen TO, MacLeod A-MK, Pardo JV, Fox PT, Petersen SE (1994) Practice-related changes in human brain functional anatomy during nonmotor learning. Cerebral cortex 4:8-26.
Rauch SL, Whalen PJ, Savage CR, Curran T, Kendrick A, Brown HD, Bush G, Breiter HC, Rosen BR (1997) Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. Human brain mapping 5:124-132.
Reber AS (1967) Implicit learning of artificial grammars. Journal of verbal learning and verbal behavior 6:855-863.
Rieskamp J, Otto PE (2006) SSL: a theory of how people learn to select strategies. Journal of Experimental Psychology: General 135:207.
Robbins TW (2007) Shifting and stopping: fronto-striatal substrates, neurochemical modulation and clinical implications. Philosophical Transactions of the Royal Society of London B: Biological Sciences 362:917-932.
Rose M, Haider H, Salari N, Büchel C (2011) Functional dissociation of hippocampal mechanism during implicit learning based on the domain of associations. The Journal of Neuroscience 31:13739-13745.
Rosenthal CR, Andrews SK, Antoniades CA, Kennard C, Soto D (2016) Learning and Recognition of a Non-conscious Sequence of Events in Human Primary Visual Cortex. Current Biology.
Rushworth MF, Behrens TE (2008) Choice, uncertainty and value in prefrontal and cingulate cortex. Nature neuroscience 11:389-397.
Saffran JR, Johnson EK, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. Annual Review of Psychology 70:27-52.
Schapiro AC, Kustner LV, Turk-Browne NB (2012) Shaping of object representations in the human medial temporal lobe based on temporal regularities. Current biology : CB 22:1622-1627.
Schendan HE, Searl MM, Melrose RJ, Stern CE (2003) An FMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. Neuron 37:1013-1025.
Schwarb H, Schumacher EH (2012) Generalized lessons about sequence learning from the study of the serial reaction time task. Advances in cognitive psychology 8:165.
Seger CA (1994) Implicit learning. Psychol Bull 115:163-196.
Seger CA (2013) The visual corticostriatal loop through the tail of the caudate: Circuitry and function. Frontiers in systems neuroscience 7.
Seger CA, Cincotta CM (2006) Dynamics of frontal, striatal, and hippocampal systems during rule learning. Cerebral Cortex 16:1546-1555.
Shanks DR, Tunney RJ, McCarthy JD (2002) A Re-examination of Probability Matching and Rational Choice. J Behav Dec Making 15:233–250.
Shaw R (1984) The Dripping Faucet as a Model Chaotic System: Aerial Press.
Summerfield C, Egner T (2009) Expectation (and attention) in visual cognition. Trends in cognitive sciences 13:403-409.
Turk-Browne NB, Jungé JA, Scholl BJ (2005) The automaticity of visual statistical learning. Journal of Experimental Psychology: General 134:552.

946 Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK (2009) Neural evidence of statistical learning:
947 Efficient detection of visual regularities without awareness. Journal of Cognitive
948 Neuroscience 21:1934-1945.
949 Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM (2010) Implicit perceptual anticipation triggered
950 by statistical learning. The Journal of Neuroscience 30:11177-11187.
951 van den Bos E, Poletiek FH (2008) Effects of grammar complexity on artificial grammar learning.
952 Memory & Cognition 36:1122-1131.
953 Woo CW, Krishnan A, Wager TD (2014) Cluster-extent based thresholding in fMRI analyses: pitfalls
954 and recommendations. Neuroimage 91:412-419.

955

Figure 1

a

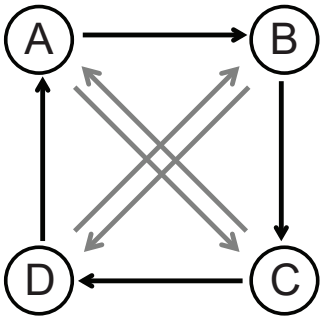


b

Level-0: Zero-order model

A	B	C	D
0.18	0.72	0.05	0.05

Level-1: First-order model



Level-1		Target			
		A	B	C	D
Context	A		0.8	0.2	
	B			0.8	0.2
	C	0.2			0.8
	D	0.8	0.2		

c

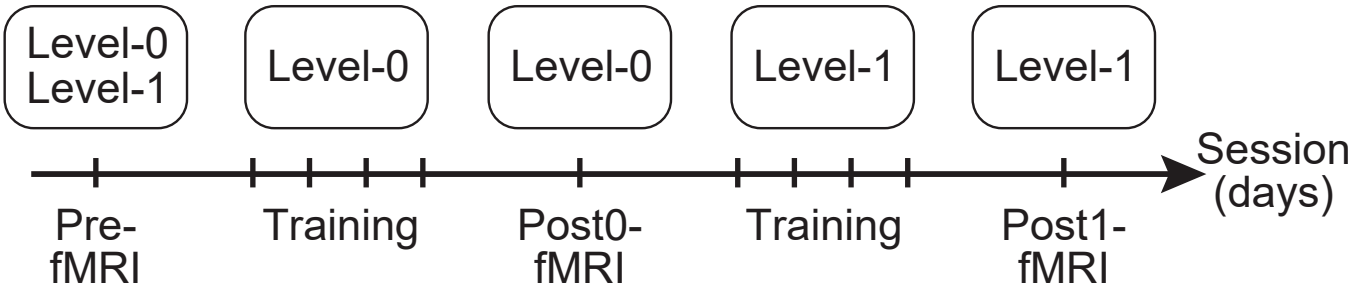
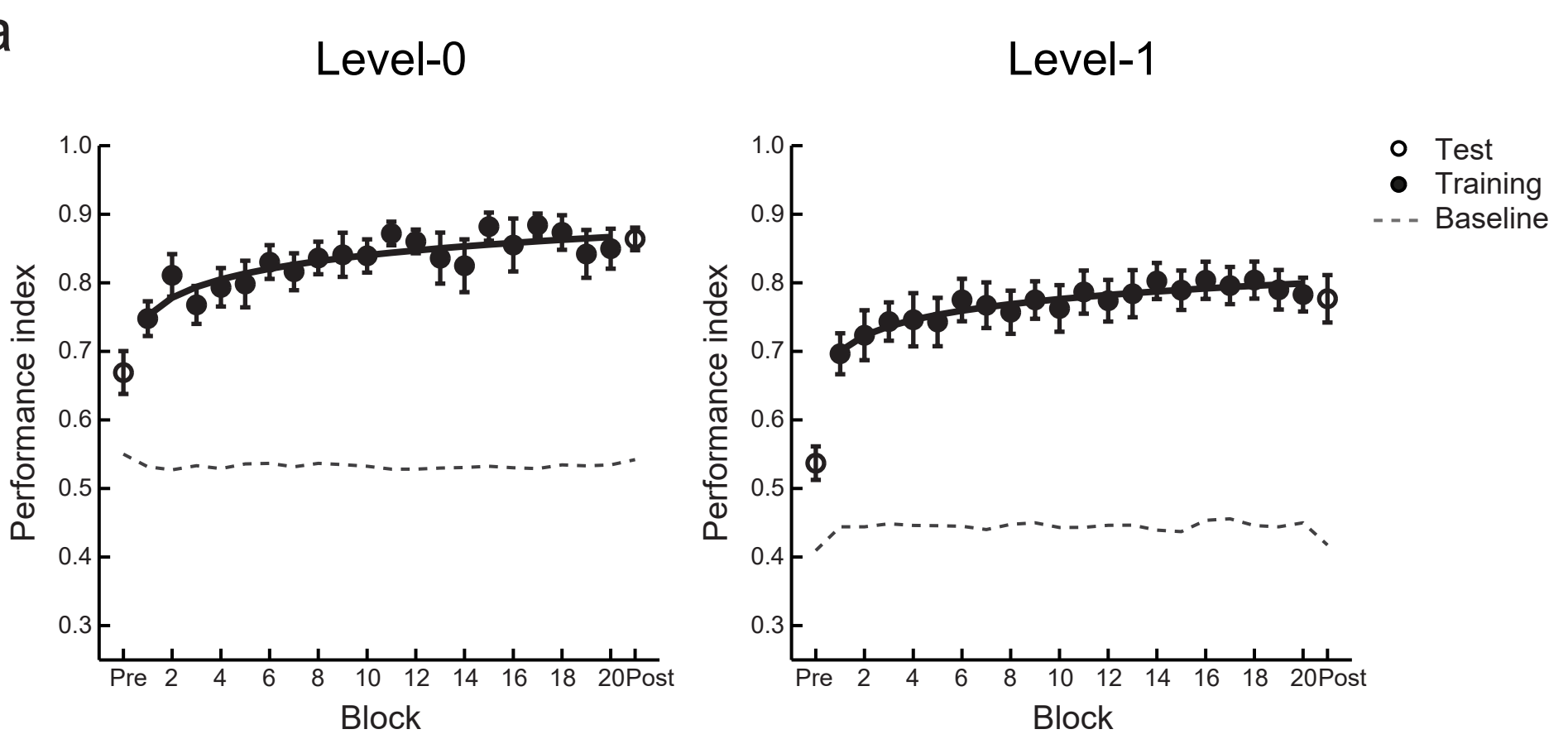


Figure 2


a
Level-0
Level-1
Performance index
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
Pre 2 4 6 8 10 12 14 16 18 20Post
Block
Test
Training
Baseline

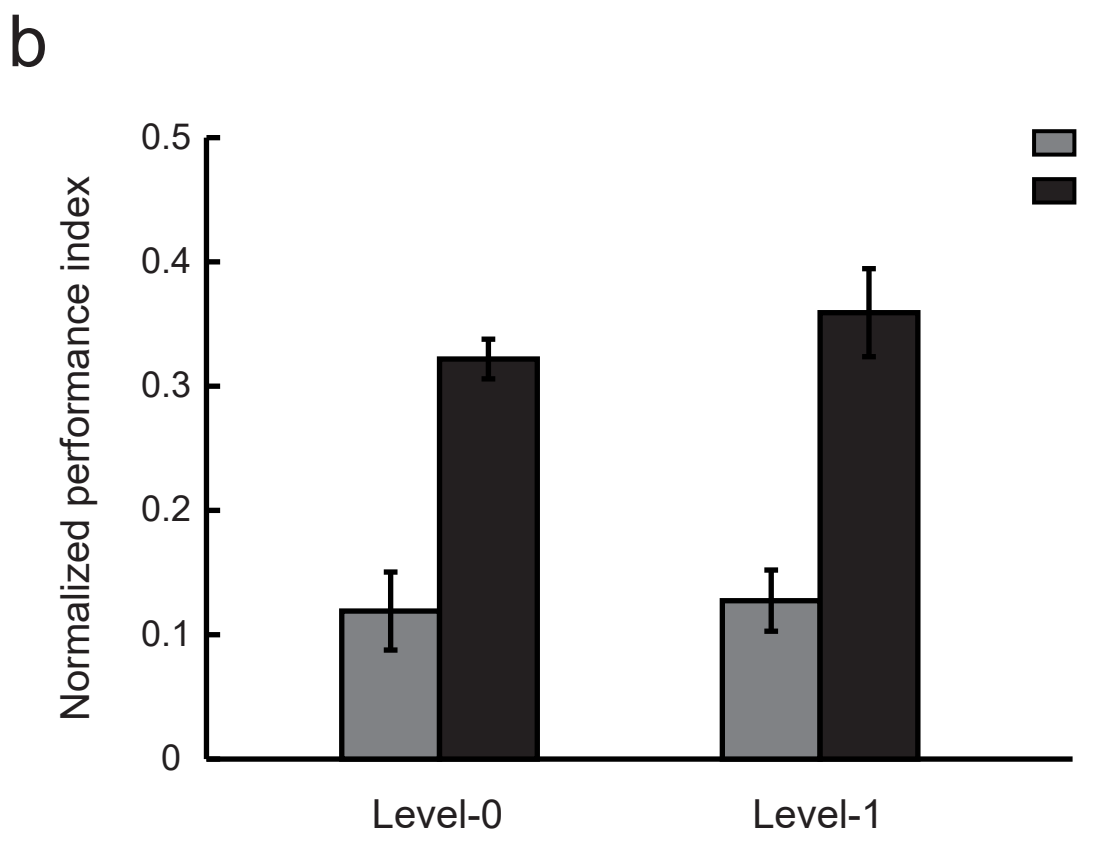
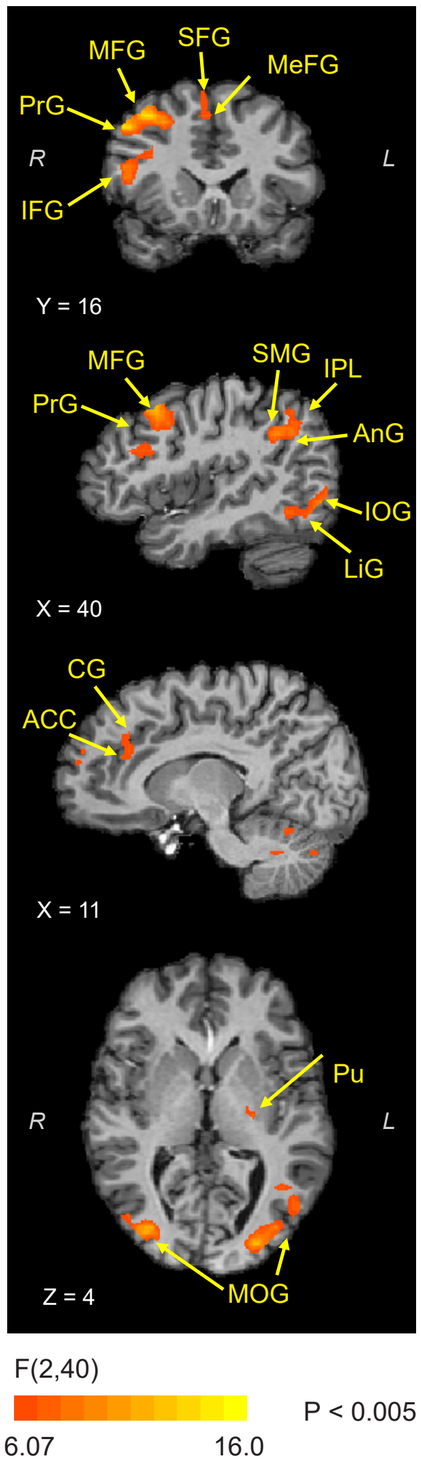

b
Normalized performance index
0.5
0.4
0.3
0.2
0.1
0
Pre-training
Post-training
Level-0
Level-1

Figure 3

a


Interaction
Scanning session × Sequence
MFG
SFG
MeFG
PrG
R
L
IFG
Y = 16
MFG
SMG
IPL
PrG
AnG
IOG
LiG
X = 40
CG
ACC
X = 11
Pu
R
L
Z = 4
MOG
F(2,40)
P < 0.005
6.07
16.0

b

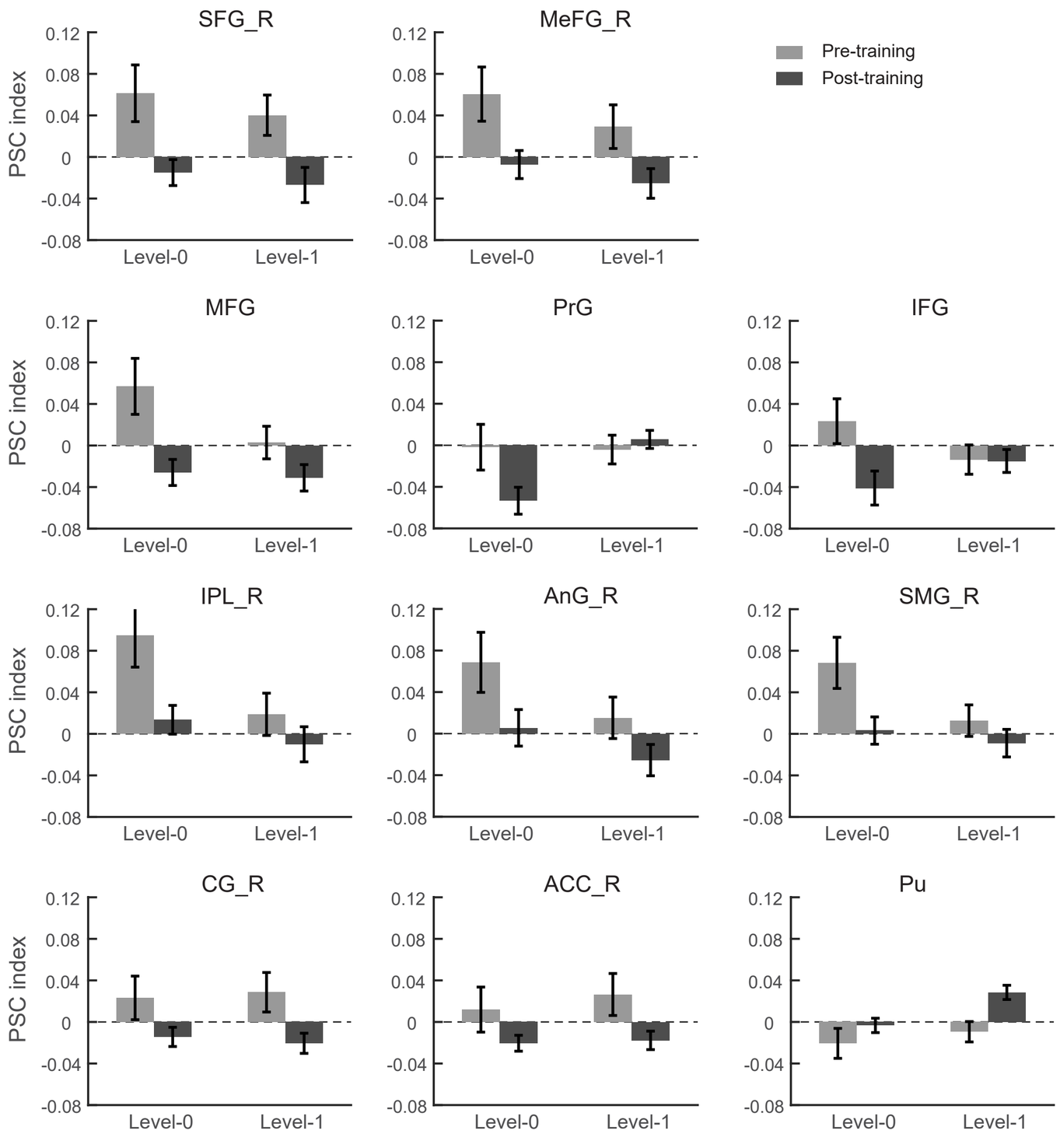
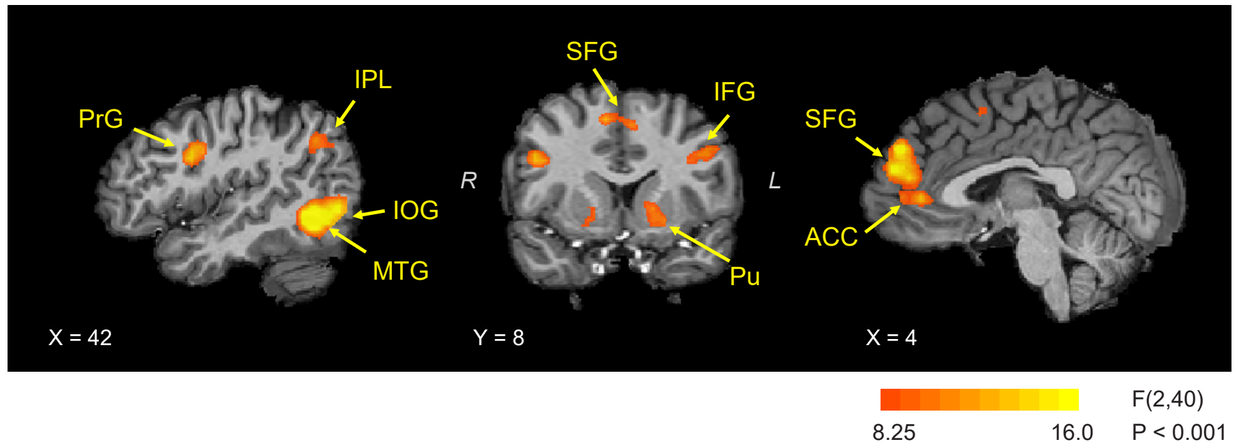

SFG_R
MeFG_R
Pre-training
Post-training
MFG
PrG
IFG
IPL_R
AnG_R
SMG_R
CG_R
ACC_R
Pu
PSC index
0.12
0.08
0.04
0
-0.04
-0.08
Level-0
Level-1

Figure 4

a


PrG
IPL
IOG
MTG
SFG
IFG
Pu
R
L
SFG
ACC
X = 42
Y = 8
X = 4
F(2,40)
8.25
16.0
P < 0.001

b

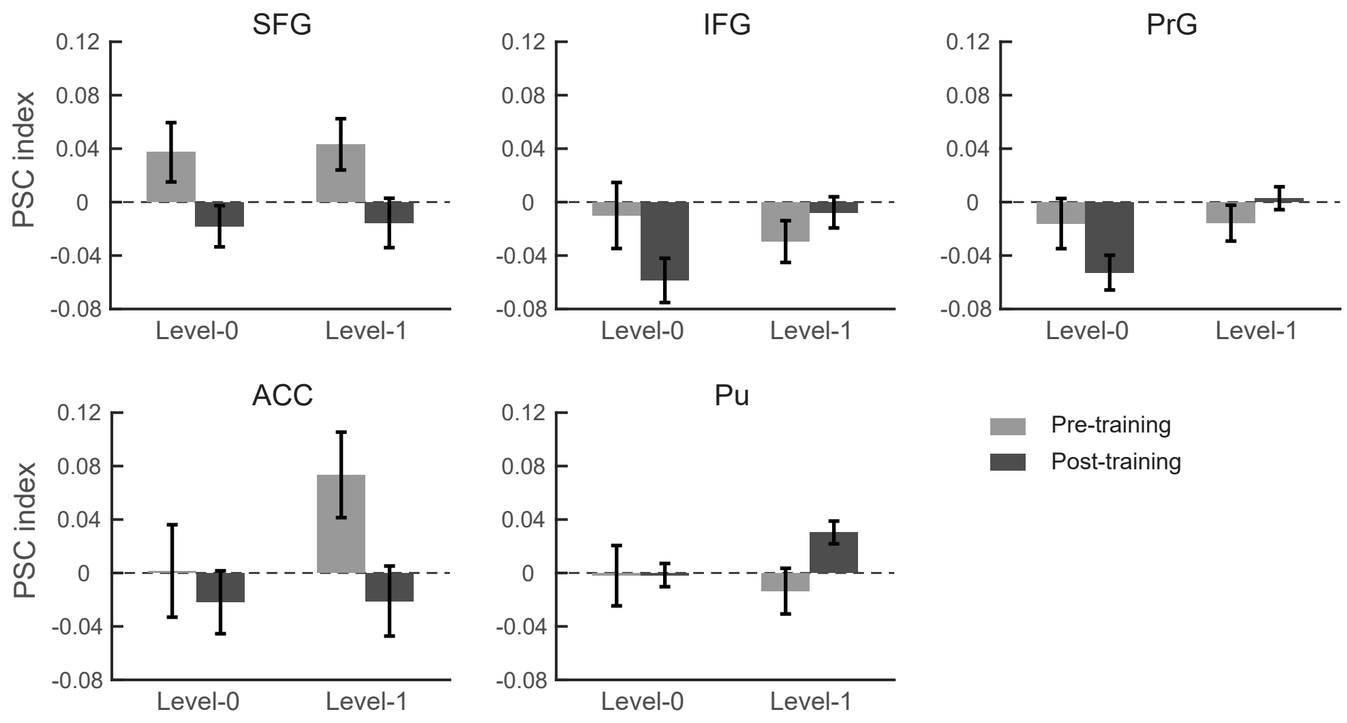
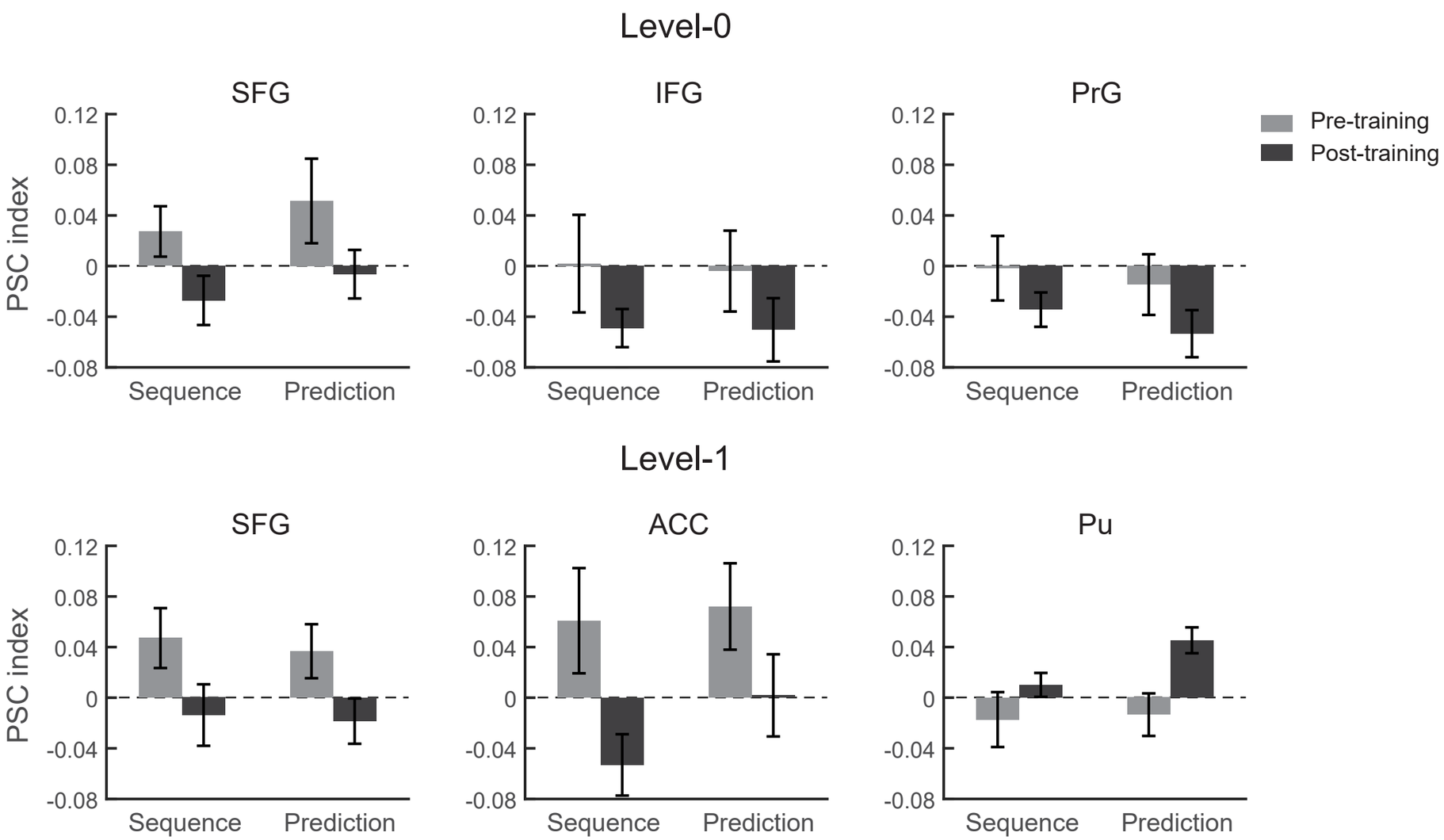

SFG
IFG
PrG
ACC
Pu
PSC index
0.12
0.08
0.04
0
-0.04
-0.08
Level-0
Level-1
Pre-training
Post-training

Figure 5

a


Level-0
SFG
IFG
PrG
Level-1
SFG
ACC
Pu
PSC index
0.12
0.08
0.04
0
-0.04
-0.08
Sequence
Prediction
Pre-training
Post-training

b

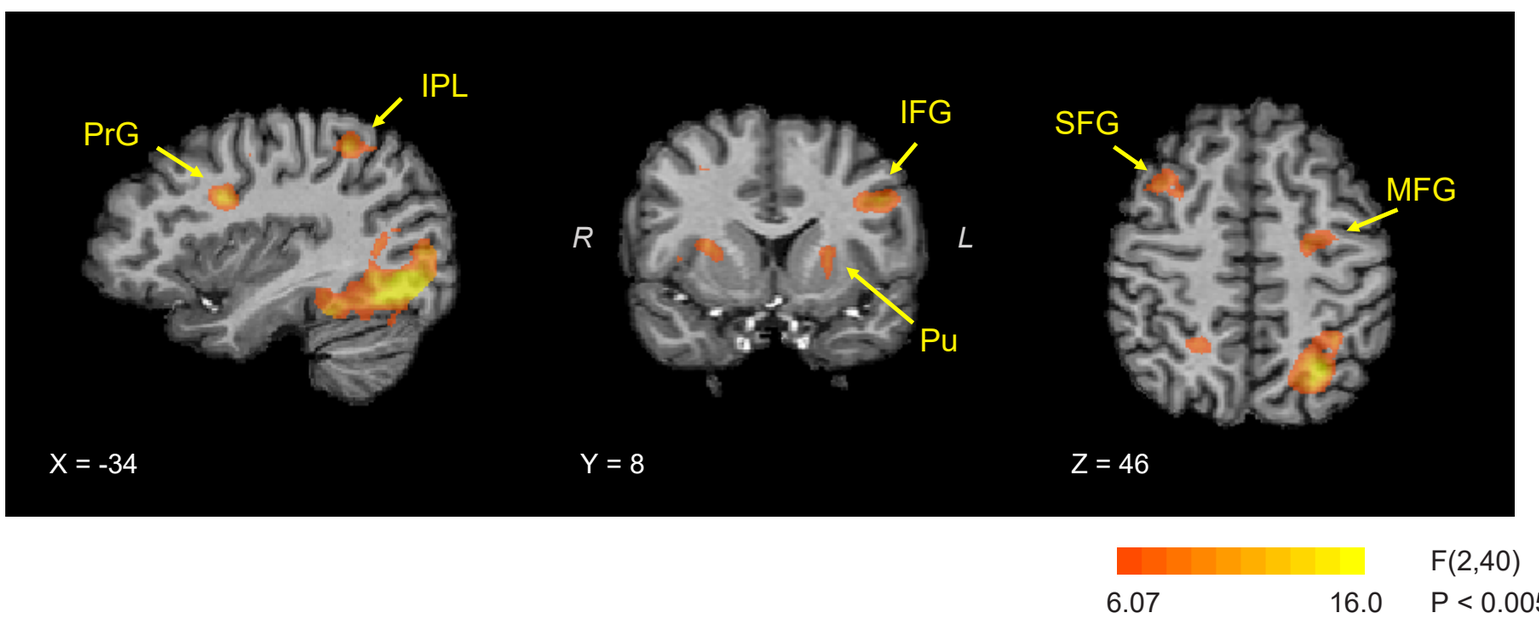

PrG
IPL
IFG
Pu
SFG
MFG
R
L
X = -34
Y = 8
Z = 46
F(2,40)
6.07
16.0
P < 0.005

Figure 6

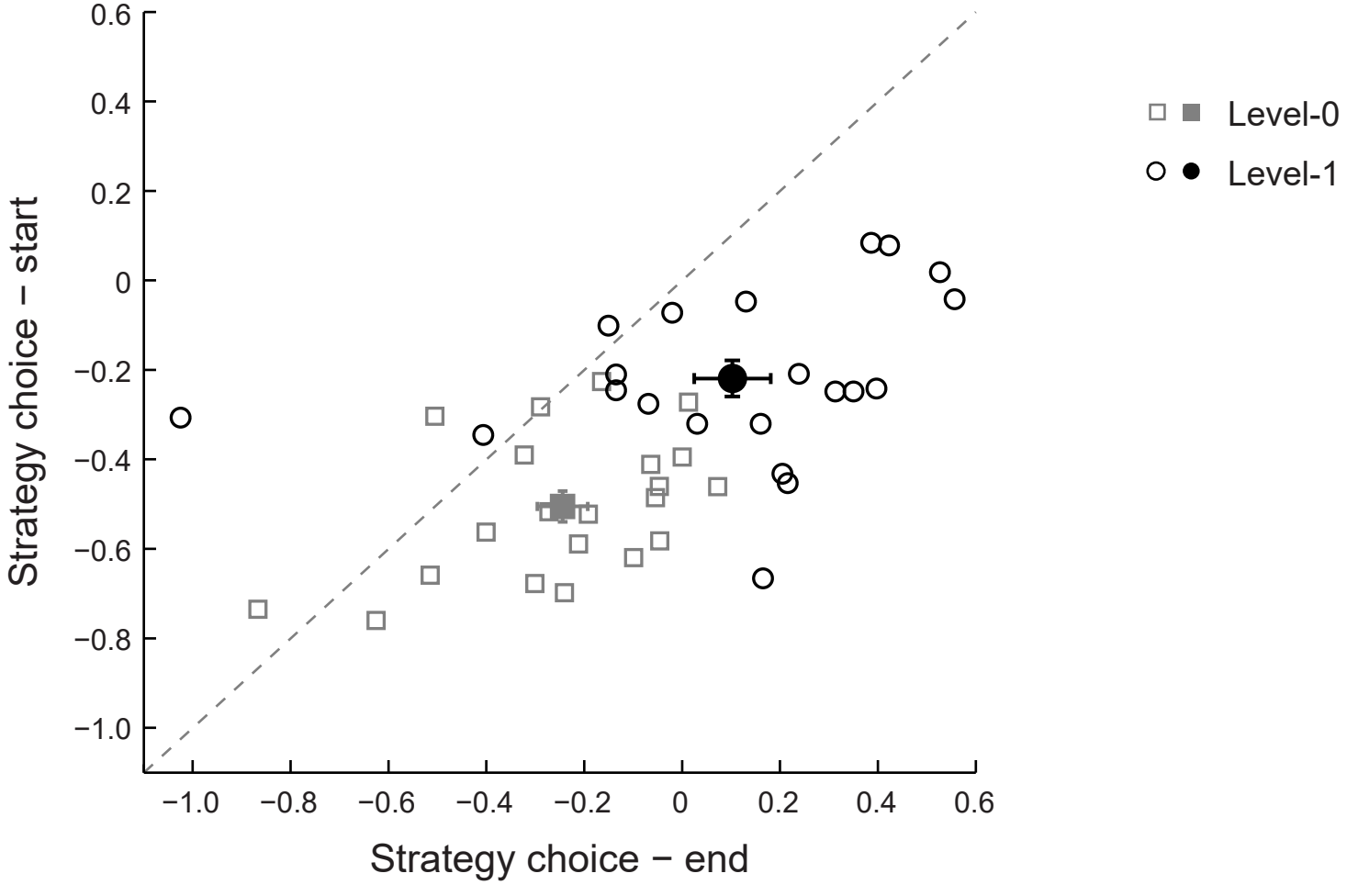
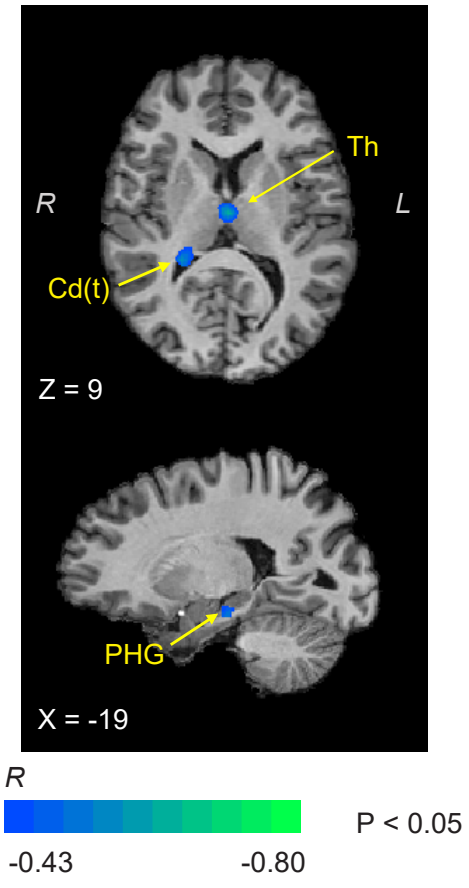
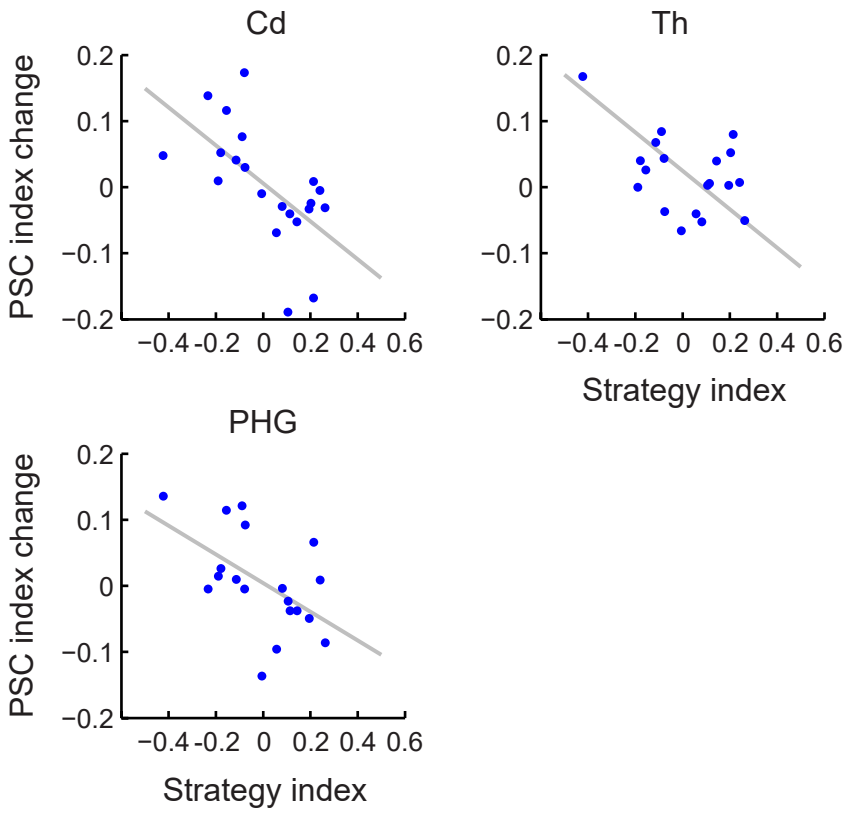

0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1.0
Strategy choice − start
−1.0
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
Strategy choice − end
Level-0
Level-1

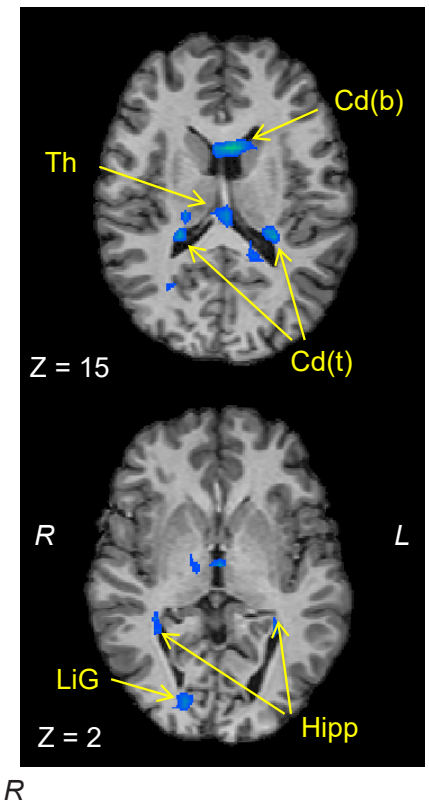
Figure 7

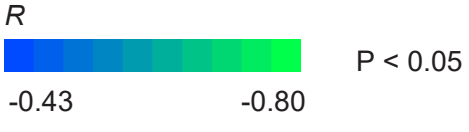
a Level-0





b Level-1





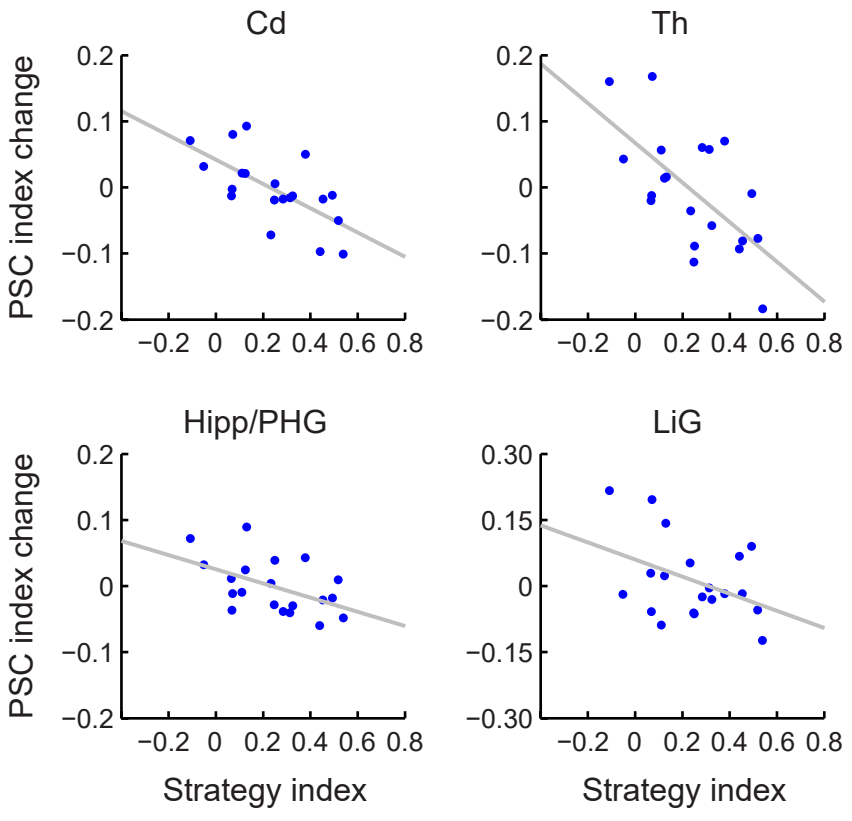
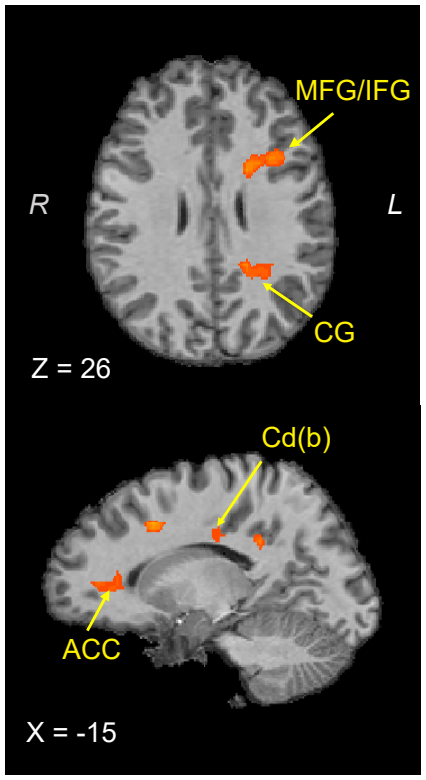
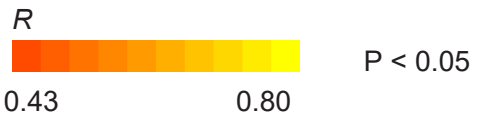
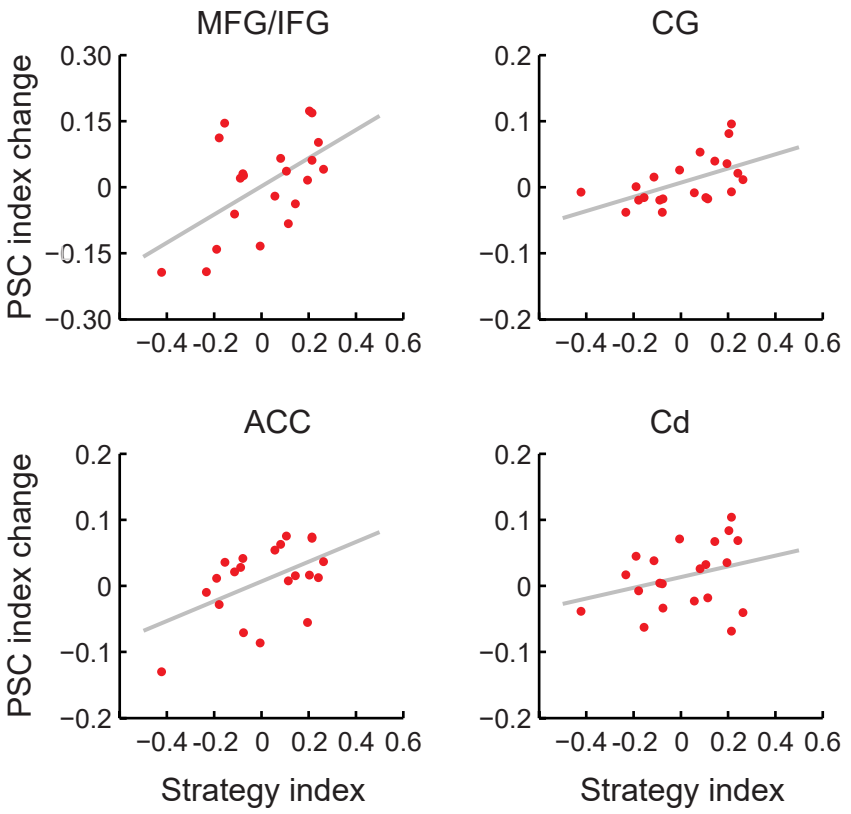


Figure 8

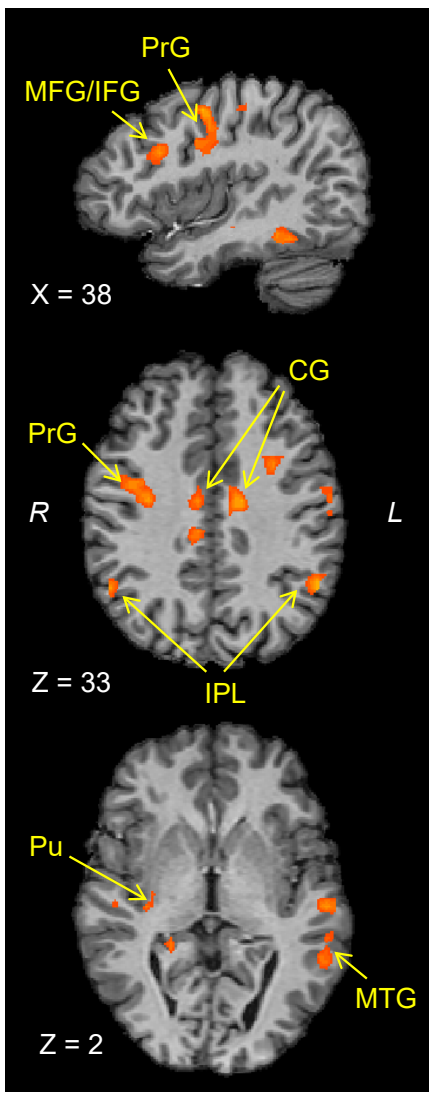
a Level-0

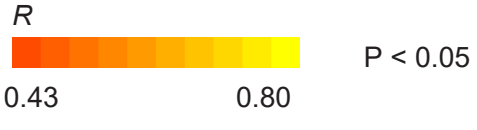






b Level-1





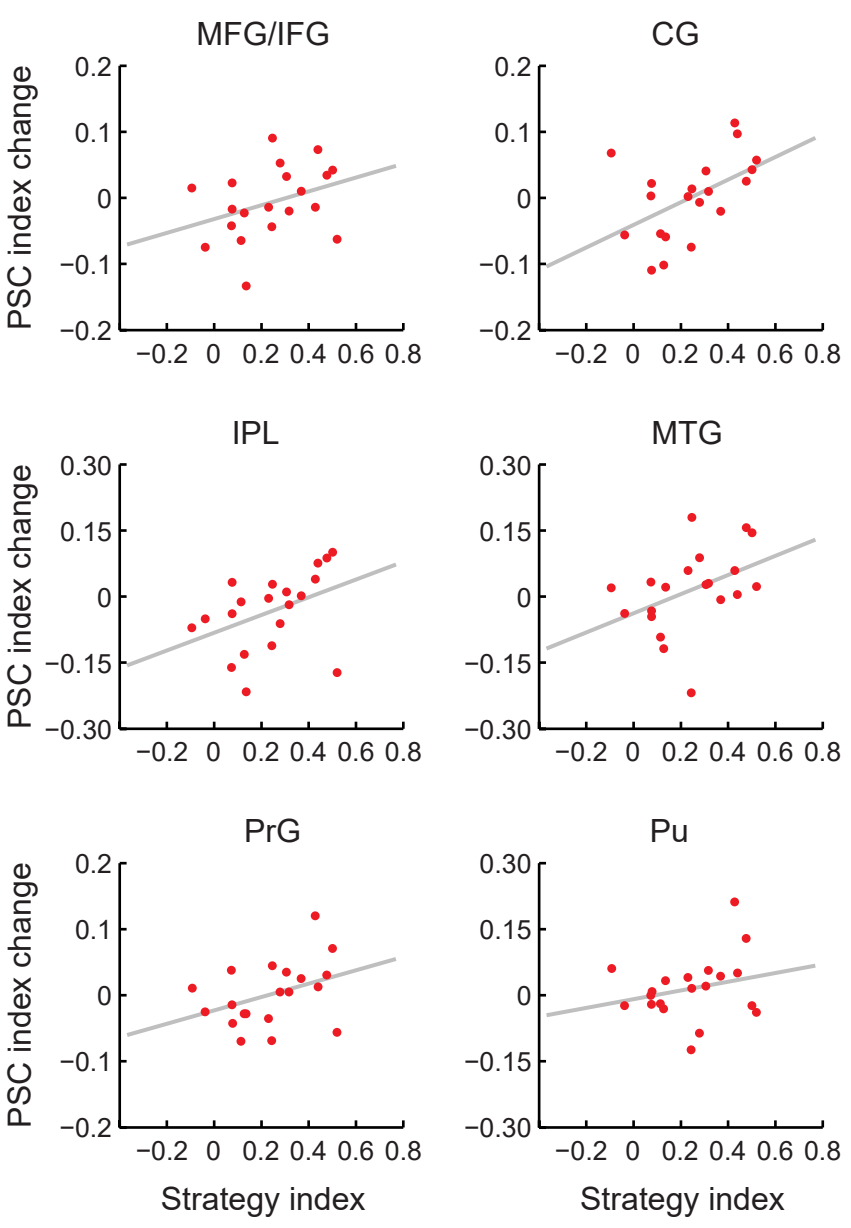


Table 1. Brain regions showing significant interaction between scanning session (pre, post-0, post-1) and sequence (structured vs. random), $p < 0.005$, cluster corrected.

ROI	Hem	Vol(mm^3)	X	Y	Z	F	P
Frontal							
Superior Frontal Gyrus (SFG)	R	1633	36	16	46	15.49539	0.00001
Medial Frontal Gyrus (MeFG)	R	922	6	32	37	9.33239	0.00047
Middle Frontal Gyrus (MFG)	L	251	-45	0	37	13.73743	0.00003
Middle Frontal Gyrus (MFG)	R	4352	45	14	40	17.07472	0.00000
Inferior Frontal Gyrus (IFG)	L	273	-45	2	31	11.74197	0.00010
Inferior Frontal Gyrus (IFG)	R	510	48	14	19	10.29143	0.00025
Precentral Gyrus (PrG)	L	1462	-45	-4	40	17.85552	0.00000
Precentral Gyrus (PrG)	R	272	43	15	40	12.12258	0.00008
Insula (Ins)	L	182	-39	-4	-2	13.93199	0.00003
Insula (Ins)	R	81	44	14	17	7.47606	0.00174
Parietal							
Precuneus (PCu)	L	1381	-21	-64	40	9.97693	0.00031
Superior Parietal Lobule (SPL)	L	506	-24	-58	40	11.28717	0.00013
Inferior Parietal Lobule (IPL)	R	859	39	-50	34	11.30387	0.00013
Angular Gyrus (AnG)	R	365	39	-58	34	10.92339	0.00016
Supramarginal Gyrus (SMG)	R	148	39	-49	34	11.47595	0.00012
Occipital							
Middle Occipital Gyrus (MOG)	L	2574	-27	-82	-5	19.95821	0.00000
Middle Occipital Gyrus (MOG)	R	1263	35	-80	1	12.63784	0.00006
Inferior Occipital Gyrus (IOG)	L	929	-36	-73	-8	21.95450	0.00000
Inferior Occipital Gyrus (IOG)	R	497	37	-79	1	13.67147	0.00003
Lingual Gyrus (LiG)	L	1346	-35	-70	-6	17.45473	0.00000
Lingual Gyrus (LiG)	R	759	30	-76	1	11.94279	0.00009
Cuneus (Cun)	L	293	-24	-82	10	9.99664	0.00030
Cuneus (Cun)	R	154	24	-79	16	8.28265	0.00098

Fusiform Gyrus (FG)	L	1901	-36	-73	-9	21.95450	0.00000
Fusiform Gyrus (FG)	R	650	36	-63	-5	12.01979	0.00008
Temporal							
Middle Temporal Gyrus (MTG)	L	662	-41	-58	-4	17.05987	0.00000
Inferior Temporal Gyrus (ITG)	L	516	-44	-58	-5	15.70175	0.00001
Sub Gyral (SGL)	L	81	-42	-51	-3	8.82149	0.00067
Parahippocampal Gyrus (PHG)	L	149	-39	-50	1	9.78701	0.00035
Parahippocampal Gyrus (PHG)	R	98	33	-55	-5	11.96719	0.00008
Limbic							
Cingulate Gyrus (CG)	R	188	24	11	43	9.27872	0.00049
Anterior Cingulate (ACC)	R	160	15	32	22	8.12427	0.00109
Subcortical							
Claustrum (Cl)	L	132	-37	-4	-2	11.50993	0.00011
Putamen (Pu)	L	93	-24	-16	4	8.27780	0.00098
Thalamus (Th)	L	266	-12	-19	7	8.50720	0.00084
Cerebellum (Cb)							
Culmen	L	61	-1	-61	-22	7.37876	0.00187
Culmen	R	611	19	-58	-19	11.72096	0.00010
Nodule	L	505	0	-53	-26	16.10464	0.00001
Nodule	R	582	0	-52	-26	16.60786	0.00001
Pyramis	L	197	0	-67	-26	12.81382	0.00005
Pyramis	R	252	6	-70	-26	14.53668	0.00002
Declive	L	586	-36	-61	-11	15.01214	0.00001
Declive	R	1752	18	-58	-17	12.49729	0.00006
Uvula	L	266	0	-68	-27	12.27454	0.00007
Uvula	R	372	6	-70	-29	14.59243	0.00002
Cerebellar Tonsil	L	195	-6	-52	-32	9.38866	0.00045
Cerebellar Tonsil	R	113	3	-59	-31	8.41985	0.00089