Articles in PresS. J Neurophysiol (March 4, 2015). doi:10.1152/jn.00753.2014

# Decoding the future from past experience: learning shapes predictions in early visual cortex.

Caroline Di Bernardi Luft[1], Alan Meeson[2], Andrew E Welchman[3], Zoe Kourtzi[3]
[1]Department of Psychology, Goldsmiths, University of London, London UK
[2]School of Psychology, University of Birmingham, Birmingham, B15 2TT, UK
[3]Department of Psychology, University of Cambridge, Cambridge UK

Running title: Learning to predict in the visual cortex

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

Number of pages: 35
Number of figures: 6

*Keywords*: visual learning, prediction, fMRI, visual cortex, sensory processing

## Acknowledgements

We are grateful to Dr Hiroshi Ban and Nuno Reis Gonçalves for helpful discussions. This work was supported by a Wellcome Trust Senior Research Fellowship to AEW (095183/Z/10/Z) and grants to ZK from the Biotechnology and Biological Sciences Research Council [H012508], a Leverhulme Trust Research Fellowship (RF-2011-378) and the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement no. PITN-GA-2011-290011.

**Abstract**

Learning the structure of the environment is critical for interpreting the current scene and predicting upcoming events. However, the brain mechanisms that support our ability to translate knowledge about scene statistics to sensory predictions remain largely unknown. Here, we provide evidence that learning of temporal regularities shapes representations in early visual cortex that relate to our ability to predict sensory events. We tested the participants' ability to predict the orientation of a test stimulus following exposure to sequences of leftwards or rightwards orientated gratings. Using fMRI decoding, we identified brain patterns related to the observers' visual predictions rather than stimulus-driven activity. Decoding of predicted orientations following structured sequences was enhanced after training, while decoding of cued orientations following exposure to random sequences did not change. These predictive representations appear to be driven by the same large-scale neural populations that encode actual stimulus orientation and to be specific to the learned sequence structure. Thus, our findings provide evidence that learning temporal structures supports our ability to predict future events by reactivating selective sensory representations as early as in primary visual cortex.

## Introduction

Successful everyday interactions entail that we exploit information about the structure of the environment to interpret the current scene and predict upcoming events. Recent theoretical work (Geisler 2008; Petrov et al. 2005) suggests that the brain achieves this challenge by learning through exposure to the environment's statistics. There is accumulating evidence that mere exposure to stimuli that co-occur in the environment facilitates our ability to extract spatial and temporal regularities (for reviews: Aslin and Newport 2012; Perruchet and Pacton 2006). However, the brain mechanisms that mediate our ability to predict upcoming events based on previous knowledge about the environment's statistics remain largely unknown.

Previous neuroimaging work has implicated subcortical and medial temporal lobe regions in the learning of temporal statistics. In particular, the striatum and hippocampus have been implicated in learning of probabilistic associations (Poldrack et al. 2001; Shohamy and Wagner 2008) and temporal sequences (Gheysen et al. 2011; Hsieh et al. 2014; Rauch et al. 1997; Rose et al. 2011; Schapiro et al. 2014; Schapiro et al. 2012; Schendan et al. 2003a). While these brain regions are thought to be involved at the initial stages of training, prefrontal regions have been shown to engage at later learning stages (Leaver et al. 2009; Pasupathy and Miller 2005). Despite accumulating evidence for neural circuits involved in learning temporal regularities, it remains unknown whether this knowledge of temporal statistics facilitates sensory predictions. Here, we test whether learning of temporal regularities shapes processing in primary visual cortex and mediates our ability to predict the identity of upcoming visual stimuli.

We devised a novel paradigm to measure behavioral performance and brain activity related to visual predictions. First, we tested the participants' ability to predict the identity of a visual stimulus (i.e. grating orientation) following learning of temporal sequences. Our behavioral results demonstrate that observers learn to exploit temporal regularities and improve their

ability to predict the identity of upcoming stimuli. Second, using fMRI we tested whether processing in visual cortex is altered after learning of temporal sequences and reflects the observers' improved ability to predict the identity of upcoming stimuli. To ensure that we measured activity related to the observers' predictions rather than the presented stimuli, we introduced a long blank interval between the presentation of temporal sequences and the test stimulus (Figure 1a). Despite the low BOLD signal during this period of no-stimulation, we were able to decode the orientation predicted by the observers in each trial after training using multi-voxel pattern classification methods. Further, to test whether decoding accuracy reflected knowledge of temporal structure, we tested brain activity before and after training when observers were presented with a random sequence and asked whether a cued orientation at the end of the sequence matched the test stimulus. Decoding of cued orientations following a random sequence did not change after training and was weaker than decoding of predicted orientations following structured sequences, suggesting that learning of temporal structure shapes predictive representations in primary visual cortex.

## Materials and Methods

### Participants

Sixteen healthy students from the University of Birmingham (mean age=21±2.6) took part in the study. All participants were naïve to the aims of the study, were right-handed, had normal or corrected-to-normal vision, had no history of neurological disorders and gave written informed consent. This study was approved by the University of Birmingham Ethics Committee.

### Stimuli

Stimuli comprised grayscale sinusoidal gratings that were presented at $9^o$ visual angle, spatial frequency that ranged from 0.85 to 1 cycles per degree across trials, 100% contrast and randomized phase. These gratings were rotated +/- $45^o$ from vertical orientation ($90^o$), resulting in gratings oriented at either $135^o$ (left) or $45^o$ (right). To avoid local position adaptation, we randomized the phase and jittered the orientation within a range of $2^o$ across trials. We used these stimuli to generate two sequences, each comprising of 8 gratings, as shown below (1 refers to a leftwards oriented grating and 2 refers to a rightwards oriented grating):

Sequence A: 2 1 2 1 1 2 1 2

Sequence B: 1 1 2 1 2 2 1 2

As the two sequences predicted different orientations (Sequence A predicted a rightwards oriented grating; Sequence B predicted a leftwards oriented grating), we generated two more sequences by replacing leftwards with rightwards orientations and vice versa while keeping the sequence structure the same. These sequences were as follows:

Sequence A' = 1 2 1 2 2 1 2 1

Sequence B' = 2 2 1 2 1 1 2 1

This manipulation allowed us to counterbalance for the predicted orientations as it resulted in sequences that had the same structure but predicted different orientations (e.g. A and A') and sequences that had different structure but predicted the same orientation (e.g. A and B'). This ensured that decoding of the predicted orientation was not confounded by the specific orientations used at each temporal position but related to knowledge of the sequence structure. Analysis of the behavioral data showed that the participants were equally accurate across all sequences. This was confirmed by a 2 (*sequence type: A vs. B*) x 2 (*sequence version*: A/B vs. A'/B') repeated measures ANOVA, which showed that there was no

significant effect for *sequence type* ($F_{1,11}$ = 0.08, $P$ = .783), *sequence version* ($F_{1,11}$ = 2.17, $P$ = .169) nor a significant interaction ($F_{1,11}$ = 0.06, $P$ = .816).

Each sequence comprised four leftwards and four rightwards oriented gratings. As all gratings were presented at the same rate, participants could not use stimulus duration to group elements together or segment the sequences. Further, to ensure that participants did not perform the task simply by memorizing the first stimulus in the sequence, the orientation of the first stimulus was randomized in each trial during scanning; that is, for each of the four sequences half of the trials started with leftwards and the rest with rightwards oriented gratings. Further, to ensure that participants did not learn the task simply by memorizing the last orientations in the sequence, the last three stimuli in each sequence pair (A and B; A' and B') were the same across all sequences. These manipulations preserved equal frequency of appearance for the two orientations across trials. Finally, as the frequency of occurrence was matched for the two grating orientations in the sequence and the participants did not know how many items each sequence contained, to perform the task participants were required to learn the order of the elements in the sequence (i.e. temporal order associations among pairs or triplets of oriented gratings).

Stimuli were generated and presented using Psychtoolbox-3 (Brainard 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 64 cm. In order to keep the same visual angle for both training and scanning sessions, the stimulus size was adjusted according to the viewing distance.

**Design**

All participants (n=16) took part in two pre-training fMRI sessions, 3 to 5 behavioral training sessions, and two post-training fMRI sessions. Participants were tested in both the prediction task (1 pre-training and 1 post-training scan) and the control task (1 pre-training and 1 post-training scan). The order of the scans (prediction vs. control task) was counterbalanced across participants. After the pre-training scans, all participants were trained for 3 to 5 sessions in the lab. The number of training sessions was determined by each participant's performance: the training stopped when the participant reached performance higher than 80% correct in all four runs comprising a training session. The post-training scans took place in subsequent days following the last behavioral training session (one scan in the first and the other in the second day following the last training session). In addition, the participants completed two scanning runs of an *orientation decoding experiment* and two retinotopy localizer scans.

**Behavioral training**

Participants were trained on the prediction task without feedback for 3-5 sessions. Participants viewed 16 gratings (each sequence of 8 gratings was repeated twice in a trial) presented sequentially on a grey background at the centre of the screen. All stimuli were presented at the same rate; that is, each grating was presented for 0.3 s followed by a fixation interval of 0.3 s. Participants were asked to respond to a test grating that appeared for 0.3 s surrounded by a red circle (0.3 s). The test stimulus was preceded by a cue (red dot presented for 1 s) and was followed by a white fixation dot (1700 ms). Participants were instructed to respond (the maximum response time was 2000 ms), indicating whether the test image had the same orientation (left vs. right) as the grating they expected to appear in that position in the sequence. The test stimulus appeared only in the second repeat of the sequence and its position was randomized across trials. The test stimulus could appear in any position in the sequence except the last three positions; stimuli in these positions were the same across trials. For each run, 50% of the test stimuli were presented at the correct orientation for their

position in the sequence. After the participant's response, the remaining gratings in the sequence were presented until all 16 stimuli had been presented, ensuring that all trials had the same length. A black cross (1 s) indicated the end of the sequence and the start of a new trial. There was no feedback across all training sessions. In each training session, participants performed the prediction task for 4 runs of 40 trials each (20 per sequence type) with a minimum two-minute break between runs. The number of training sessions was determined based on performance; the participants stopped training after reaching consistent session accuracy above 80% (all training runs within one session had to be above 80%).

After each training session, the participants were asked to complete a debriefing questionnaire with the following questions: 1) Please describe any strategies you may have used when responding to this task 2) How many of your responses do you think that were correct? (1 to 5, from 'few correct' to 'most correct'); 3) How did you find the task? (1 to 5, from 'very difficult' to 'very easy'); 4) How tired did you feel at the end of each run? (1 to 5, from 'very tired' to 'not tired at all'); 5) How many different sequences of stimuli do you think that were presented? In addition, after the last training session, the participants were asked to write down the sequences they thought that were presented during the experiment."

**fMRI design**

The participants took part in 2 pre- and 2 post-training scans: one pre- and one post-training scan for the prediction task and one pre- and one post-training scan for the control task. The order of the scans was counterbalanced: half of the participants did the prediction task in the first pre- and post-training scan session, whereas the other half started with the control task. In addition, the participants completed two runs of an *orientation decoding experiment* and two retinotopy localizer scans (Polar angle and Eccentricity).

193 *Prediction task scan* (Figure1a): Participants completed seven to nine runs (12 trials per run)
194 of the prediction task per scan session. Each run followed an event-related design comprising
195 12 trials and a fixation block (15 s) at the beginning and end of the run. Participants were
196 presented with all four sequences used for training. Each sequence was repeated once per trial
197 (comprising 8 stimuli) followed by a test stimulus. For each trial (28.5 s long), a sequence of
198 eight leftwards ($135^{o}$) or rightwards ($45^{o}$) oriented gratings was presented. Each grating was
199 presented for 0.25 s followed by fixation for 0.2 s. The sequence of gratings was followed by
200 a fixation period (11.6 s), a cue (black square, 0.5 s), a test grating (0.5 s) and a red dot (2 s)
201 before the start of the next trial. The participants were instructed to pay attention to the
202 sequence, and respond whether the orientation of the test grating matched the orientation they
203 expected to follow from the preceding sequence. To ensure that all participants viewed the
204 test grating for the same duration and there were no differences in reaction time across
205 participants, participants were instructed to delay their response until the red dot appeared
206 after the test grating. After each trial, there was a fixation period of 10.5 s.

207 To acquire adequate data (i.e. number of trials) for the fMRI analysis within the time
208 constraints of the scanning sessions, we used shorter sequences (single sequence comprising
209 8 stimuli) during scanning (instead of two repeats of the same 8-item sequence comprising in
210 total 16 stimuli during training). That is, during scanning participants were required to predict
211 the orientation of the stimulus in the $9^{th}$ temporal position of the trained sequences. Our
212 experiential design during training (i.e. same presentation duration across stimuli, variable
213 temporal test position during training) made it unlikely that the participants had explicit
214 knowledge of the sequence length or number of sequence repeats, as also indicated by
215 debriefing. Further, to ensure that participants did not simply memorize the first stimulus in
216 the sequence during scanning, we randomized (across trials) the orientation of the first
217 stimulus that was then followed by the remaining seven items in the sequence. Thus, it is

unlikely that participants memorized the orientation of stimuli presented at individual temporal positions for each of the four trained sequences. In contrast, it is more likely that participants learned temporal associations between sequence items (i.e. pairs or triplets) during training that remained the same in the test sequences and facilitated their predictions.

*Control task scan* (Figure1b): Participants completed 7 to 9 runs (12 trials per run) of a control task per scan session. To test learning improvement specific to structured sequences we presented participants with a random sequence of leftwards and rightwards oriented gratings that were presented equally often but at randomised positions within the sequence. To ensure that performance in the control task was comparable to the prediction task after training, we asked participants to compare the orientation of the test grating with a cued orientation presented after the sequence. This allowed us to compare fMRI activations between tasks (prediction vs. control) with comparable levels of behavioural performance. Each run followed an event-related design comprising 12 trials and a fixation block (15 s) at the beginning and end of the run. The per-trial design for the control task matched that of the prediction task. Each trial (28.5 s long) comprised a random sequence of eight leftwards ($135^{o}$) or rightwards ($45^{o}$) oriented gratings (i.e. gratings were presented at random order in the sequence). Each grating was presented for 0.25 s followed by fixation for 0.2 s. This random sequence of gratings was followed by a cue: 'R' or 'L' (0.25 s) indicating whether the participants should remember a rightwards or leftwards oriented grating respectively. This cue was followed by a fixation period (11.35 s), a cue (black square, 0.5 s), a test grating (0.5 s) and a red dot (2 s) before the start of the next trial. The control and prediction tasks were matched on a trial-by-trial basis for the orientation of expected and remembered items; that is, the orientation indicated by the cue in the control task matched the expected orientation in the prediction task on a per-trial basis. The participants were instructed to pay attention to

the sequence, remember a grating rotated leftwards or rightwards as indicated by the cue and indicate whether the test grating matched the cued orientation.

*Orientation decoding scan*: All participants completed two runs of an orientation decoding experiment following procedures described previously (Harrison and Tong 2009). Participants were presented with the same gratings as in the prediction and control task (100% contrast and oriented either leftwards or rightwards). Leftwards vs. rightwards oriented gratings were presented in separate 15 s long blocks. Similar to the prediction task scans, to avoid adaptation due to stimulus repetition, we randomized the phase and jittered the orientation of the gratings within a range of $2^{o}$ across trials. Each block comprised 30 gratings. Each grating was presented for 0.25 s followed by a blank interval of 0.25 s. Each run comprised 20 blocks (10 per orientation) and two fixation blocks: one in the beginning and one at the end of the run. The order of the blocks was randomized across runs. Participants were asked to perform a contrast change detection task on the fixation. That is, participants were instructed to press a button when they detected a contrast change at fixation (twice per block at random time points).

*Retinotopic mapping scans*: For each participant we independently localized regions in the early (V1, V2) and higher (V3v, V3d & hV4) visual areas following standard retinotopic mapping procedures (e.g. Sereno et al. 1995). Data from polar and eccentricity scans were collected either during the pre- or post-training scan session. hV4 comprises the ventral but not the dorsal sub-region of V4.

**fMRI Data Acquisition**

fMRI data were acquired in a 3T Achieva Philips scanner at the Birmingham University Imaging Centre using an eight-channel head coil. Anatomical images were obtained using a sagittal three-dimensional T1-weighted sequence (voxel size=1 × 1 × 1 mm, slices=175).

Functional EPI images were acquired using a high-resolution gradient echo-pulse sequence covering the occipital and posterior temporal cortex (20 slices at 1.5 × 1.5 × 2 mm resolution; matrix size 128 x 128; slice thickness: 2mm with no gap between slices; FOV: 192 x 192; TR: repetition time, 1500-ms; TE: time to echo, 35 ms).

**Eye movement recordings**

We recorded eye-movements (n=6) using the ASL 6000 Eye-tracker (Applied Science Laboratories, Bedford, MA, sampling rate: 60Hz) in the scanner. Eye tracking data were pre-processed using the EyeNal Data Analysis software (Applied Science Laboratories, Bedford, MA) and analyzed using custom toolbox based on Matlab (Mathworks, MA) software. Due to poor signal quality, data from two participants were excluded from the analysis. Runs with more than 10% signal loss were removed from the analysis. We computed (A) horizontal eye position, (B) vertical eye position, (C) proportion of saccades for each condition at different saccade amplitude ranges, and (D) number of saccades per trial per condition during the blank interval following the sequence presentation, separately for each pre-training (I) and post-training (II) sessions. Histograms of the horizontal and vertical eye positions peaked and were centered on the fixation at zero degrees suggesting that participants could fixate well both before and after training when predicting leftwards or rightwards oriented gratings.

**Data analysis**

**Behavioral data analysis**

The performance on the task was assessed by the accuracy in correctly predicting whether the next grating in the sequence was left or right. For the training sessions, we averaged the accuracy for each run of the sequential sessions and estimated a learning rate by fitting a logarithmic function to the data (Figure 1c). Data across runs were fitted (least squares non-

linear fit) using the following equation: y=k*log (x) + c; where k is the value of the curve tangent at x=1, and c is the value of y for x=1.

**fMRI data pre-processing**

Neuroimaging data was analyzed using Brain Voyager QX (Brain Innovation, Maastricht, Netherlands). Anatomical data was used for 3D cortex reconstruction, inflation and flattening. Pre-processing of functional data included slice scan time correction, three-dimensional motion correction, linear trend removal, and temporal high-pass filtering (3 cycles). fMRI data was recorded at high resolution (1.5. x 1.5 mm in plane) and interpolated to 2 x 2 x 2 mm using trilinear interpolation. Trials with head motion larger than 1 mm of translation or 1º of rotation or sharp motion above 0.5 mm (on average 25 trials per session across areas and tasks) were excluded from the analysis. Runs whose motion analysis resulted in the exclusion of more than 50% of the trials were excluded from further analysis. The functional images were manually aligned to anatomical data and the complete data were transformed into Talairach space. For each observer, the functional imaging data between the four sessions were co-aligned, registering all the volumes for each observer to the first functional volume of the first run and session. The retinotopic mapping scans (polar and eccentricity) were also co-aligned with the first volume of the first run and session. This procedure ensured a cautious co-registration across sessions.

**fMRI decoding**

We used a linear support vector machine (SVM) with a leave-one-run-out cross-validation procedure for pattern classification. To investigate the link between fMRI activity and the participants' responses, we tested the classifier's performance in decoding the participant's prediction (leftwards vs. rightwards); that is, if the participant responded that a leftwards oriented test grating was 'correct', they predicted 'left', if they indicated 'incorrect', then they

313 predicted ‘right’ and vice-versa. That is, we trained the classifier to associate fMRI signals
314 with a label (Predicted Left vs. Predicted Right) as indicated by the participant’s response to
315 the test grating in each trial. To control for potential bias in the classification due to the
316 unequal numbers of trials responded to as ‘correct’ or ‘incorrect’ by the participants, we used
317 a cost-factor and weighed the error term during SVM training by the ratio of fMRI patterns
318 related to ‘correct response’ over fMRI patterns related to ‘incorrect response’.

319 To select voxels for the pattern classification analysis, we used the retinotopic
320 mapping scans for each participant. We selected voxels that corresponded to stimulus area of
321 $8^{o}$ of visual angle and were significantly more activated by the grating stimuli than the
322 fixation ($p<0.05$, uncorrected). This procedure allowed us to avoid voxels corresponding to
323 the edges of the grating stimulus (Harrison and Tong 2009). Each voxel time course was z-
324 score normalized for each experimental run separately. The data pattern for each trial was
325 generated by shifting the fMRI time series by 3 volumes (4.5 s) to account for the
326 hemodynamic delay. That is, volumes corresponding to the no-stimulation period following
327 the sequence (volumes 4 to 10) were shifted to 7 to 13 (Figure 3 shows fMRI volumes before
328 shifting to account for the hemodynamic delay). We trained and tested the classifier during
329 this no-stimulation period (volumes 8 to 13), excluding the first one (volume 7) to avoid
330 interference from BOLD responses related to the preceding sequence. We performed this
331 analysis on each of these volumes separately as well as on the average signal across these
332 volumes. For each cross-validation (from 5 to 9 depending on the number of runs per
333 participant), we used 60 to 108 patterns (for averaged signals across no-stimulation volumes)
334 for training the classifier and 12 independent patterns for testing the classifier’s accuracy. We
335 plotted classifier accuracy across voxels - starting with voxels that have the highest t-value
336 for gratings vs. fixation -and selected the 300 most activated voxels in each ROI for each
337 participant, as pattern classification accuracy had saturated at this pattern size across areas.

We then averaged the classifier's accuracy for this pattern size across cross-validations for each participant.

**Generalization of classifier accuracy**

To evaluate the correspondence between neural representations for physical and predicted orientations we followed two different approaches. First, we identified common voxels across experiments (i.e. informative voxels for both the classification of the physical and the predicted orientation). That is, using a *recursive feature elimination* (RFE) procedure (Ban et al. 2012; De Martino et al. 2008) we identified voxels across visual areas that contributed (i.e. as indicated by the classifier's linear weights) to the decoding of (1) the physical stimulus, and (2) the predicted stimulus orientations using data from the orientation decoding experiment and the prediction task respectively. These two RFE analyses were conducted separately (i.e. the classifiers were trained using a cross validation procedure either on the physical or the predicted orientations) resulting in two sets of voxels: voxels informative for the classification of the physical stimulus and voxels informative for the classification of the predicted stimulus. We then ranked voxels in each visual area resulting from the two RFE analyses and chose voxels that were informative in both analyses (i.e. 300 most informative voxels). We used these voxels to train an SVM classifier on fMRI signals related to physical orientations and tested the accuracy of this classifier in decoding predicted orientations from fMRI data collected when observers performed the prediction task (i.e. no-stimulation interval following the sequence presentation). It is important to note that we run the RFE analysis and selected voxels separately for each cross-validation of the MVPA analysis (i.e. decoding of predicted orientations) to avoid circularity; that is we excluded the test data from both the RFE and the decoding analysis. The results from this analysis are presented in Figure 5. Second, using the same *recursive feature elimination* (RFE) procedure we identified the top 300 voxels that contributed (i.e. as indicated by the classifier's linear weights) to the

363 decoding of the physical stimulus using data from the orientation decoding experiment only.
364 We then used these voxels to train and test the classifier (using a leave-one-run-out cross-
365 validation procedure) in decoding predicted orientations from fMRI data collected when
366 observers performed the prediction task (i.e. no-stimulation interval following the sequence
367 presentation). The results from the second procedure were similar to those reported in Figure
368 3, suggesting that the generalization of classification accuracy that we observed could not be
369 simply due to the voxel selection or MVPA procedure used.
370

371 **Results**

372 **Behavioural results**

373 We presented participants with a sequence of leftwards and rightwards oriented gratings
374 (Figure 1a) and asked them to predict the next grating in the sequence. We trained
375 participants on this prediction task without feedback for 3 to 5 sessions (as determined by
376 individual performance). To control for the possibility that observers memorized specific
377 items in the sequence or full sequences rather than learning the temporal structure, we trained
378 participants with four different sequences and presented all stimuli at the same rate and in a
379 continuous stream. Further, the position of the test stimulus was randomized across trials, the
380 last three items were the same across sequences, and for half of the trials the incorrect test
381 stimulus was presented.

382 Before and after training, we tested participants while performing the prediction task and a
383 control task. In particular, for the prediction task, we asked participants to indicate whether
384 the orientation of a test grating matched the orientation they anticipated following a
385 preceding structured sequence of gratings (Figure 1a). For the control task participants were
386 asked to indicate whether the orientation of a test stimulus matched the orientation of a cued
387 grating presented following a random sequence of oriented gratings (Figure 1b).

Performance on the prediction task improved for most participants (12/16 participants improved; 4 participants did not improve during training showing 57% mean performance at the last training session) as they gained more exposure to the temporal sequences (Figure 1c). We further focus on the analysis of the behavioral and fMRI data for the 12 participants that showed improvement during training and post-training performance higher than 80% correct; the data of the weaker learners (post-training performance lower than 65% correct) are considered later in a separate control analysis. Comparing performance in the prediction and control task during scanning (Figure 1d) showed that observers' performance improved after training in the prediction task, while performance in the control task remained high both before and after training. These results were confirmed by a 2 (*task*: prediction vs. control task) x 2 (*session*: pre- vs. post-test) repeated measures ANOVA showing a significant interaction between *task* and *session* ($F_{1,11}$ = 76.08, $P$ < .001), consistent with enhanced performance after training in the prediction ($t_{11}$ = -8.50, $P$ < .001) but not in the control task ($t_{11}$ = 1.42, $P$ = .183). The comparable behavioural performance after training for the prediction and control task, that involved structured vs. random sequences respectively, ensured that comparing fMRI activation patterns between the two tasks reflected learning specific to the sequence structure that was not confounded by differences in behavioural performance.

Improvement in the prediction task after training indicates that participants acquired knowledge of the sequence structure. Debriefing the participants suggests that this knowledge was most likely implicit and it was unlikely that the participants memorized the sequences explicitly. In particular, participants were significantly more confident in their responses after training ($t_{11}$ = -5.03, $P$ < .001) and found the task easier ($t_{11}$ = -4.31, $P$ = .001), but did not feel significantly more or less tired ($t_{11}$ = -1.48, $P$ = .166). Interestingly, this was not the case for weaker learners that did not report any substantial changes in their confidence (pre-

training mean = 1.5, standard deviation = 0.58; post-training mean = 1.75, standard deviation = 0.96) or task difficulty (pre-training mean = 3.50, standard deviation = 1.29; post-training mean = 2.25, standard deviation = 0.5). Finally, we asked participants to estimate the number of sequences presented during the experiment; this number did not change significantly after training ($t_{11}$ = -0.41, $P$ = .689). Only three participants indicated that there were four sequences in total, but none of the participants could explicitly report the sequences correctly. Taken together, the debriefing data suggest that it was unlikely that the participants improved in the task by explicitly memorizing the sequences.

FIGURE 1 ABOUT HERE

**fMRI results: decoding predictions in visual cortex**

To test whether learning of temporal regularities shapes sensory predictions in visual cortex, we scanned participants before and after training. For each participant we identified retinotopic visual areas using high-resolution fMRI and standard mapping procedures. To test for fMRI signals related to the participants' predictions rather than the stimuli per se, we extracted activity during a blank interval between the temporal sequences and the test stimulus. Analysis of univariate fMRI signals (percent signal change from mean BOLD response across trials) confirmed that BOLD responses during this no-stimulation period were low and did not differ before vs. after training (paired t-test for all ROIs, $P > 0.05$; Figure 2a). However, previous studies have shown that neural preference for orientation can be decoded from no-stimulated intervals in the visual cortex using multi-voxel pattern (MVPA) classification analysis (Harrison and Tong 2009; Kamitani and Tong 2005; Serences et al. 2009; Tong and Pratte 2012). Using this approach, we trained an SVM classifier to associate responses from each fMRI volume to the participants' prediction in each trial and tested the accuracy of the classifier in predicting the participants' responses (leftwards vs.

rightwards predicted orientation) using an independent data set. In contrast to the univariate signals, this analysis allowed us to successfully decode the orientation predicted by the participants from fMRI signals during periods of no-stimulation in V1. Importantly, classification accuracy increased after compared to before training (Figure 2b). Inspection of the MVPA accuracy time course across volumes showed that the improvement in decoding performance after training peaked during the blank interval following the presentation of the temporal sequence during which participants predicted the orientation of the upcoming stimulus (Figure 2b). This result reflects learning-dependent changes specific to processing of predicted orientations in the primary visual cortex.

FIGURE 2 ABOUT HERE

To quantify and compare decoding accuracies across visual areas, we selected and averaged fMRI responses from all volumes (8-13) that corresponded to the no-stimulation period during which participants predicted the orientation of the upcoming stimulus. Consistent with the behavioral results for the prediction task, decoding of predicted orientations improved significantly after training (Figure 3). In contrast for the control task, orientation decoding did not change with training, consistent with the participants' behaviour in this task.

Comparing decoding accuracies between tasks showed higher decoding accuracies after training for the prediction than the control task. In particular, a 2 (*session: pre- vs. post-training* scan) X 2 (*task: prediction vs. control*) X 6 (*ROI: V1,V2,V3d,V3a,V3v and hV4)* repeated-measures ANOVA showed a significant interaction between *task* and *session* ($F_{1,11}$ = 9.77, $P$ = .010). Enhanced decoding accuracy after training for the prediction task was primarily observed in early visual areas V1 and V2. In particular, the interaction between *task* and *session* was significant in V1 ($F_{1,11}$ = 6.683, $P$ = .025) and V2 ($F_{1,11}$ = 6.55, $P$ = .027), as decoding accuracy increased after training only for the prediction task (*paired t-test:* $t_{11}$ = -

5.03, $P < .001$), but not for the control task ($P = .783$). No significant interactions were observed in V3d and V3v, hV4 and V3a ($F < 1$). Furthermore, there was no three-way interaction (*session* X *task* X ROI: $F_{5,75} = 0.545$, $P = .741$), as indicated by higher classification accuracy in V1 and V2 compared to higher visual areas for both the prediction and the control task. These weaker effects in higher compared to early (V1, V2) visual areas (effect of *ROI:* $F_{5,55} = 4.44$, $P = .002$) have been previously observed in fMRI studies testing responses in non-stimulated visual cortex (Harrison and Tong 2009; Kok et al. 2013; Kok et al. 2012; Smith and Muckli 2010) and are potentially due to stronger orientation selective responses in early visual areas (Kamitani and Tong 2005).

Further, additional analysis after removing more volumes at the beginning (volumes 8 and 9) and the end (volume 13) -to avoid activity due to stimulation from the sequences or the test grating- showed the same pattern of results. As in the main analysis, a 2 (*session: pre- vs. post-training* scan) X 2 (*task: prediction vs. control*) X 6 (*ROI: V1,V2,V3d,V3a,V3v and hV4)* repeated-measures ANOVA showed a significant interaction between *task* and *session* ($F_{1,11} = 7.80$, $P = .017$). Enhanced decoding accuracy after training for the prediction task was primarily observed in early visual areas V1 and V2. That is, we observed significant effects for *session* in V1 ($F_{1,11} = 17.77$, $P = .001$) and V2 ($F_{1,11} = 7.92$, $P = .017$), but not in V3d, V3v, V3a and hV4. There was a significant interaction between *session* and *task* in both V1 ($F_{1,11} = 17.3$, $P = .002$) and V2 ($F_{1,11} = 11.39$, $P = .006$), consistent with enhanced classification accuracy after training only for the prediction task.

FIGURE 3 ABOUT HERE

These results provide evidence for visual cortex representations that are specific to the learned sequence structure - rather than random sequences as presented in the control task- and reflect participants' predictions. Increased decoding accuracies after training for the

prediction task could not be simply explained by (1) general familiarity with the stimuli or the task after training, as these remained the same across tasks, (2) differences in the task design. In particular, as both tasks require a coarse (left vs. right) rather than fine orientation matching of the predicted or cued orientation to the test stimulus, participants may hold in memory and potentially mentally imagine a label or visual image of the predicted or cued stimulus. The cue in the control task may encourage participants to keep a label in memory that can be visualized for comparison to the visual test stimulus, while in the prediction task there is no cue and the participants may verbalize or visualize their prediction. However this difference in task design did not result in significant differences in average fMRI responses (main *effect for task:* $F_{1,11} = 0.01$, $P = .905$; *interaction between session* and *task*: $F_{1,11} = 0.41$, $P = .537$) during the non-stimulation period, suggesting similar effect of working memory or imagery processes in early visual cortex across tasks. In contrast, the key difference between tasks is in the content (predicted stimulus following a structured sequence vs. cued physical stimulus following a random sequence) of the representation accessed by the participants. Thus, higher decoding accuracies after training in the prediction compared to the control task–despite similar behavioral performance after training in both tasks (paired t-test: $t_{11} = -1.92$, $P = 0.082$)- suggest that our results indicate predictive representations specific to the trained structured sequences rather than differences in working memory or visual imagery processes across tasks.

Finally, significant correlation of the participants' performance with decoding accuracy in V1 after training ($r = .609$, $P = .036$) suggest that selective representations for predicted orientation relate to the observers' enhanced ability to predict upcoming stimuli after training on temporal sequences (Figure 4a). Interestingly, decoding accuracy did not improve significantly after training in any of the visual areas ($F < 1$) for participants (n=4) who did not show improved performance during training (57% mean performance at the last training

session). A 2 (*session: pre- vs. post-training* scan) X 6 (*ROI: V1,V2,V3d,V3a,V3v,hV4)* repeated-measures ANOVA showed no significant main effects for session ($F_{1,3}$ = 0.028, $P$ = .878), ROI nor a significant interaction ($F_{1,3}$ = 0.073, $P$ = .804) (Figure 4b). Taken together these results suggest a strong link between the observers' ability to predict the identity of upcoming stimuli following training on temporal sequences and orientation representations in early visual areas.

FIGURE 4 ABOUT HERE

**Comparing physical vs. predicted orientation representations**

Finally, we asked whether neural populations –as recorded at a large scale by fMRI- that encode physical stimulus orientation may also represent predicted orientation. To this end, we tested whether activity patterns for predicted orientations resemble stimulus-driven activity elicited during viewing of oriented gratings. We collected an independent set of fMRI data while the participants viewed leftwards vs. rightwards oriented gratings in separate blocks (45$^{o}$ or 135$^{o}$). Consistent with previous studies (Harrison and Tong 2009; Haynes and Rees 2005; Kamitani and Tong 2005) decoding of orientation from stimulus-driven activity was successful across visual areas (Figure 5a). To evaluate the correspondence between neural representations for physical and predicted orientations, we trained an SVM on fMRI signals related to physical orientations and tested the accuracy of the classifier in decoding predicted orientations from fMRI data collected when observers performed the prediction task. Despite stimulus and task differences between these experiments, we observed generalization of classifier performance in V1 and V2 after training but not in higher areas where orientation selectivity is known to be weaker. Importantly, we observed improved classification accuracy for predicted orientations after training, suggesting that learning temporal sequences modulates neural populations in early visual cortex involved in the selective processing of orientation (Figure 5b). In particular, a 2 (*session: pre- vs. post-training*) X 6 (*ROI: V1, V2,*

535 *V3d, V3a, V3v and hV4*) repeated-measures ANOVA showed higher classification accuracy
536 after training (i.e. main effect for *session:* $F_{1,11}$ = 18.768, $P$ = .001). Although there was no
537 significant interaction between *session* and *ROI* ($F_{1,11}$ = 1.05, $P$ = .400), improved decoding
538 accuracy after compared to before training was more prominent in V1 (paired t-test: $t_{11}$ = -
539 3.62, $P$ = .004) and V2 (paired t-test: $t_{11}$ = -4.16, $P$ = .002), but only marginally significant
540 for V3d (paired t-test: $t_{11}$ = -1.93, $P$ = .080). Similar analysis for the control task (Figure 5c)
541 did not show any significant differences after vs. before training in predicting cued
542 orientations across ROIs (paired t-tests: $t_{11}$ <1, n.s.). Comparing generalization of classifier
543 performance between the prediction and control tasks showed a significant interaction
544 between task and session ($F_{1,11}$ = 5.17, $P$ = .044), consistent with our main findings suggesting
545 enhanced predictive representations that are specific to training with structured than random
546 sequences of items.

547 FIGURE 5 ABOUT HERE

548 **Control analyses**

549 We conducted the following additional analyses to control for possible alternative
550 explanations of the results.

551 First, to control for the possibility that our results are due to random correlations in the data,
552 we conducted the same decoding analysis using randomly permuted fMRI patterns (i.e. we
553 randomized the correspondence between fMRI data and labels and performed MVPA for
554 10000 iterations). This analysis resulted in classification accuracies that did not differ from
555 chance for both pre- (mean=50%, standard deviation = 0.05) and post-training (mean =
556 49.8%, standard deviation = 0.054) data ($t$ < 1, n.s.) in V1. In particular, decoding accuracy
557 for the prediction task was higher than the 97.5th percentile of the random distribution for 10
558 out of 12 participants after training but only for 2 participants before training. Similar results

were observed across ROIs. This analysis suggests that our decoding results could not be simply accounted for by random variations in the data'

Second, we focused on fMRI decoding based on the participants' responses in the prediction task, as our goal was to understand whether the visual cortex contains information that relates to behavioral predictions. However, this analysis may be confounded by differences in the number of correct trials between sessions; that is, larger number of correct trials after compared to before training. To control for this, we first conducted the same analysis using fMRI signals only for correct trials for both scan sessions. A 2 (*session: pre- vs. post-training* scan) X 2 (*task: prediction vs. control*) X 6 (*ROI: V1,V2,V3d,V3a,V3v and hV4)* repeated-measures ANOVA, showed similar improvement of classification accuracies after training. That is, we observed significantly higher classification accuracy after than before training in the main task as indicated by significant *session* x *task* interactions for V1 ($F_{1,11}$ = 20.59, $P$ = .001), V2 ($F_{1,11}$ = 15.43, $P$ = .002), V3a ($F_{1,11}$ = 8.83, $P$ = .013) and V3d ($F_{1,11}$ = 5.26, $P$ = .043), but not for V3v ($F_{1,11}$ = 3.056, $P$ = .108) and hV4 ($F_{1,11}$ = 0.356, $P$ = .563). Second, we performed an additional analysis (Figure 6) that decoded the expected orientation as determined by the presented sequence rather than the participants' prediction. We found a similar pattern of results, with significantly higher classification accuracies after than before training across ROIs (*session* x *task:* $F_{1,11}$ = 10.76, $P$ = .007). Interestingly, this was not the case for weaker learners: prediction accuracies did not change across sessions, consistent with the results presented in Figure 4b. These control analyses suggest that our results could not be simply explained by differences in task difficulty between scanning sessions.

FIGURE 6 ABOUT HERE

Third, is it possible that the difference in the decoding accuracy for the prediction task before and after training was due to different strategies adopted by the observers? That is, before training the participants found the task too difficult and responded randomly when the test

stimulus appeared instead of predicting the upcoming stimulus. However, low prediction accuracies before training for the expected orientation as determined by the presented sequence (Figure 6) suggests that fMRI patterns in visual cortex contain information about structured sequences after but not before training. As this analysis does not rely on the participants' responses, low accuracies before training could not be due to the participants' response strategy (e.g. guessing).

Fourth, our results could not be simply explained by differences in participants' attention between scan sessions, as analysis of univariate fMRI signals for leftwards and rightwards oriented gratings did not differ before and after training. Further, this analysis justifies our choice of multi-voxel pattern classification methods for the analysis of predictive representations that have been shown to be more sensitive than univariate approaches in extracting selective signals from brain patterns. Finally, analysis of eye-movement recordings did not show any differences between scanning sessions and between left vs. rightwards predicted orientations across sessions, suggesting that it is unlikely that eye movements could explain our results. In particular, a 2 (*session:* pre- vs. post-training) X 2 (*orientation:* left- vs. right) repeated-measures ANOVA revealed that there was no significant effects of *session* or *orientation* for the mean eye position (session: $F_{1,3}$=0.086, $P$ = .789; orientation: $F_{1,3}$) = 0.29, $P$ = .625), mean saccade amplitude ($F_{1,3}$ = 2.2, $P$ = .230; $F_{1,3}$ = 0.48, $P$ = .536), or number of saccades per trial per condition ($F_{1,3}$ = 0.57, $P$ = .504; $F_{1,3}$ = 0.42, $P$ = .564). Also, there was no significant interaction between *session* and *orientation* for the mean eye position ($F_{1,3}$ = 1.89, $P$ = .263), the number of saccades ($F_{1,3}$ = 0.55, $P$ = .512) or the saccade amplitude ($F_{1,3}$ = 0.44, $P$ =.55).

**Discussion**

Our results provide evidence that learning of temporal regularities supports our ability to predict future events by reactivating selective sensory representations in primary visual

609 cortex. Interestingly, these predictive representations appear to be driven by the same large-
610 scale neural populations that encode physical stimulus properties (i.e. orientation) and to be
611 specific to the learned sequence structure. Further, these representations reflect our ability to
612 predict future events as indicated by a significant correlation between fMRI decoding and
613 behavioral improvement in the prediction task after training.

614 Consistent with our previous behavioral work (Baker et al. 2013), we demonstrate that
615 exposure to temporal regularities in a scene allows us to accumulate information about its
616 structure and predict future events. Although we used deterministic sequences, we ensured
617 that observers learned the global sequence structure (i.e. temporal order statistics across items
618 rather than each item position in the sequence) by matching the frequency of occurrence of
619 each item (i.e. grating orientation) in the sequence. Previous studies have suggested that
620 learning of regularities may occur implicitly in a range of tasks: visuomotor sequence
621 learning (Nissen and Bullemer 1987), artificial grammar learning (Reber 1967), probabilistic
622 category learning (Knowlton et al. 1994), and contextual cue learning (Chun and Jiang 1998).
623 In our study, participants were exposed to the sequences without feedback but were asked to
624 make an explicit judgement about the identity of the upcoming test stimulus (leftward vs.
625 rightward oriented grating) making them aware of the dependencies between the stimuli
626 presented in the sequence. However, debriefing the participants showed that it was unlikely
627 that the participants explicitly memorized the sequences, suggesting that they made
628 predictions based on implicit knowledge of temporal structure.

629 Our fMRI findings advance our understanding of the brain mechanisms that support our
630 ability to translate previous knowledge to future predictions in four main respects. First,
631 previous imaging and physiology studies suggest that extensive training on visual detection
632 or discrimination tasks may modulate processing in primary visual cortex (Bao et al. 2010;
633 Furmanski et al. 2004; Jehee et al. 2012; Schoups et al. 2001). Our findings extend beyond

634 this work showing that mere exposure to the statistics of the environment alters selectively
635 orientation representations in primary visual cortex to reflect the observers' prediction.
636 Further, recent animal physiology studies provide evidence for reactivation of neural
637 responses in visual cortex when neurons are activated in a temporal sequence (Eagleman and
638 Dragoi 2012; Gavornik and Bear 2014). Our study provides novel evidence that such
639 experience-dependent neural reactivation correlates and may facilitate the ability of human
640 observers to make predictions of upcoming sensory events. Although the nature of the
641 signals that support orientation decoding has been recently debated (Freeman et al. 2011),
642 here we demonstrate that the same large-scale neural populations that encode physical
643 orientations in primary visual cortex encode also the predicted orientations. Thus, our work
644 provides novel evidence that previous knowledge alters processing in primary visual cortex
645 that mediates our ability to make sensory predictions.

646 Second, learning of spatial and temporal regularities has been suggested to engage the
647 striatum and medial temporal lobe regions (Gheysen et al. 2011; Hsieh et al. 2014; Rauch et
648 al. 1997; Rose et al. 2011; Schapiro et al. 2014; Schapiro et al. 2012; Schendan et al. 2003a).
649 For example, studies (Foerde et al. 2006; Poldrack et al. 2001; Poldrack et al. 1999) using the
650 weather prediction paradigm have implicated these regions in implicit learning of
651 probabilistic associations. Previous work has implicated mainly striatal regions (e.g. caudate
652 and putamen) in implicit learning (Hazeltine et al. 1997; Rauch et al. 1995), while the medial
653 temporal lobe in both implicit and explicit learning (Schendan et al. 2003a; b). Our results
654 suggest interactions between these memory circuits and visual cortex; that is, learning of
655 temporal structures that is known to engage this circuit may shape representations in primary
656 visual cortex that relate to our ability to make sensory predictions. This is consistent with
657 recent work implicating the hippocampus in memory of temporal order (for a review see:
658 Eichenbaum 2013; Howard and Eichenbaum 2013), and prospective memory of simulated

659 future events (Ingvar 1985; Szpunar et al. 2013). Further, related work has suggested that
660 associative learning engaging medial temporal lobe regions modulates processing in
661 inferotemporal (Meyer and Olson 2011; Miyashita 1993) and area MT (Schlack and Albright
662 2007) and early visual (Bosch et al. 2014) cortex. While this previous work has focused on
663 paired associations, we propose that learning higher order regularities in the context of
664 temporal sequences may employ similar brain circuits to translate knowledge about temporal
665 structure in medial temporal lobe to predictions in sensory areas.

666 Third, there is accumulating evidence for the role of primary visual cortex in predictive
667 coding. In particular, recent fMRI (Alink et al. 2010; Harrison et al. 2007; Kok et al. 2014;
668 Kok et al. 2012; Murray et al. 2002; Schoups et al. 2001; Smith and Muckli 2010;
669 Summerfield and Egner 2009), and neurophysiology studies (Guo et al. 2007; Kim et al.
670 2012; Meyer and Olson 2011; Perrett et al. 2009) have shown that responses in visual cortex
671 are modulated by spatio-temporal context. These findings have been observed in the context
672 of tasks involving stimulus anticipation based on paired associations or short-term history
673 (i.e. probability of occurrence for a preceding stimulus). Further, higher responses are
674 observed for unexpected than expected stimuli, consistent with increased prediction error
675 when sensory signals and top-down expectations do not correspond (Bastos et al. 2012;
676 Friston 2005; Summerfield and Egner 2009). Our study extends beyond these findings in
677 several respects. First, our paradigm allows us to test how longer-term knowledge acquired
678 through several training sessions rather than short-term stimulus history affects prediction in
679 primary visual cortex. Second, our study is the first to test the role of sequence learning on
680 predictions related to visual recognition. Previous work on learning temporal sequences has
681 focused on implicit measures of sequence learning, such as familiarity judgments or reaction
682 times (Nissen and Bullemer 1987); for review see (Schwarb and Schumacher 2012).
683 Although such paradigms implicate that implicit learning of temporal sequences facilitates

684 the anticipation of upcoming events, they do not test whether this knowledge can be used to
685 explicitly predict the identity of upcoming stimuli. In contrast, our design allows us to test for
686 neural representations related to explicit predictions about the identity of an upcoming
687 stimulus (i.e. grating orientation) rather than anticipation as revealed typically by implicit
688 measures (e.g. reaction times, familiarity) of visual recognition. Third, using multi-voxel
689 pattern classification methods allows us to test how previous knowledge affects selective
690 processing of sensory features (i.e. orientation) related to the observers' response (i.e. per-
691 trial prediction) rather than simply changes in the overall fMRI magnitude related to
692 expectation. Decoding predicted orientation during a no-stimulation period before the test
693 stimulus appears allows us to investigate the processes involved in predicting upcoming
694 sensory events, in contrast to previous work investigating predictive coding based on the
695 error generated when unexpected stimuli are presented.

696 Finally, recent imaging work has highlighted the role of primary visual cortex in cognitive
697 functions such as working memory and visual imagery (Albers et al. 2013; Harrison and
698 Tong 2009; Serences et al. 2009). The prediction task used in our study involves these
699 processes, as it entails that participants hold in memory and/or imagine the predicted stimulus
700 in order to match it to the test stimulus. However, comparing the prediction task with a
701 control task using random sequences and a similar design involving the same processes (i.e.
702 holding in memory and/or imagining a grating orientation) demonstrates predictive
703 representations in primary visual cortex that are specific to the knowledge of structured
704 sequences. In particular, comparing average fMRI responses for the no-stimulation period
705 between tasks did not show any significant differences, suggesting that higher orientation
706 decoding accuracy in the prediction than the control task cannot be simply due to differences
707 in the task design. In contrast, the critical difference between tasks is in the content of the
708 representation accessed by the participants (predicted orientation following a structured

sequence vs. cued physical orientation following a random sequence) which we decode using multivoxel pattern classification of fMRI data. . This result is further supported by significant correlation of the participants' performance in the prediction task with decoding accuracy in V1 after training. Taken together these results suggest predictive representations in early visual cortex following learning of structured sequences that cannot not be simply explained by differences in working memory or visual imagery processes across tasks.

In sum, our findings provide evidence that knowledge of temporal regularities alters processing in primary visual cortex to support our ability for sensory predictions. The high resolution imaging adopted in our study afforded us the signal quality necessary to reveal activity patterns related to predictive representations, but it restricted brain coverage to visual cortex. Given the complex nature of the BOLD signal, it is possible that the fMRI selectivity that we observed for predicted orientations is enhanced by feedback from other cortical circuits. Possible candidates include: a) medial temporal lobe and subcortical areas that are known to be involved in associative learning and temporal memory, b) prefrontal circuits that support rule-based behaviours and prediction of future events (Bar 2009; Leaver et al. 2009; Pasupathy and Miller 2005). It is also important to note that - despite the enhanced sensitivity of our methodology- decoding reveals neural preferences at the scale of large neural populations rather than tuning of individual neurons. Therefore, understanding the cortical circuits that support our ability to translate previous knowledge to sensory predictions requires further whole brain connectivity studies combining advanced imaging and neurophysiological techniques.

## References


**Albers AM, Kok P, Toni I, Dijkerman HC, and de Lange FP**. Shared representations for working memory and mental imagery in early visual cortex. *Curr Biol* 23: 1427-1431, 2013.

**Alink A, Schwiedrzik CM, Kohler A, Singer W, and Muckli L**. Stimulus predictability reduces responses in primary visual cortex. *J Neurosci* 30: 2960-2966, 2010.

**Aslin RN, and Newport EL**. Statistical Learning: From Acquiring Specific Items to Forming General Rules. *Current Directions in Psychological Science* 21: 170-176, 2012.

**Baker R, Dexter M, Hardwicke T, Goldstone A, and Kourtzi Z**. Learning to predict: Exposure to temporal sequences facilitates prediction of future events. *Vision Res* 2013.

**Ban H, Preston TJ, Meeson A, and Welchman AE**. The integration of motion and disparity cues to depth in dorsal visual cortex. *Nat Neurosci* 15: 636-643, 2012.

**Bao M, Yang L, Rios C, He B, and Engel SA**. Perceptual learning increases the strength of the earliest signals in visual cortex. *J Neurosci* 30: 15080-15084, 2010.

**Bar M**. The proactive brain: memory for predictions. *Philos Trans R Soc Lond B Biol Sci* 364: 1235-1243, 2009.

**Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, and Friston KJ**. Canonical microcircuits for predictive coding. *Neuron* 76: 695-711, 2012.

**Bosch SE, Jehee JF, Fernandez G, and Doeller CF**. Reinstatement of associative memories in early visual cortex is signaled by the hippocampus. *J Neurosci* 34: 7493-7500, 2014.

**Brainard DH**. The Psychophysics Toolbox. *Spat Vis* 10: 433-436, 1997.

**Chun MM, and Jiang YH**. Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology* 36: 28-71, 1998.

**De Martino F, Valente G, Staeren N, Ashburner J, Goebel R, and Formisano E**. Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage* 43: 44-58, 2008.

**Eagleman SL, and Dragoi V**. Image sequence reactivation in awake V4 networks. *PNAS* 109: 19450–19455, 2012.

**Eichenbaum H**. Memory on time. *Trends Cogn Sci* 17: 81-88, 2013.

**Foerde K, Knowlton BJ, and Poldrack RA**. Modulation of competing memory systems by distraction. *Proc Natl Acad Sci U S A* 103: 11778-11783, 2006.

**Freeman J, Brouwer GJ, Heeger DJ, and Merriam EP**. Orientation decoding depends on maps, not columns. *J Neurosci* 31: 4792-4804, 2011.

**Friston K**. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360: 815-836, 2005.

**Furmanski CS, Schluppeck D, and Engel SA**. Learning strengthens the response of primary visual cortex to simple patterns. *Curr Biol* 14: 573-578, 2004.

**Gavornik JP, and Bear MF**. Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nature Neuroscience* in press: 2014.

**Geisler WS**. Visual perception and the statistical properties of natural scenes. *Annu Rev Psychol* 59: 167-192, 2008.

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

**Guo K, Robertson RG, Pulgarin M, Nevado A, Panzeri S, Thiele A, and Young MP**. Spatio-temporal prediction and inference by V1 neurons. *Eur J Neurosci* 26: 1045-1054, 2007.

**Harrison LM, Stephan KE, Rees G, and Friston KJ**. Extra-classical receptive field effects measured in striate cortex with fMRI. *Neuroimage* 34: 1199-1208, 2007.

**Harrison SA, and Tong F**. Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458: 632-635, 2009.

**Haynes JD, and Rees G**. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci* 8: 686-691, 2005.
**Hazeltine E, Grafton ST, and Ivry R**. Attention and stimulus characteristics determine the locus of motor-sequence encoding. A PET study. *Brain* 120 ( Pt 1): 123-140, 1997.
**Howard MW, and Eichenbaum H**. The hippocampus, time, and memory across scales. *J Exp Psychol Gen* 142: 1211-1230, 2013.
**Hsieh LT, Gruber MJ, Jenkins LJ, and Ranganath C**. Hippocampal Activity Patterns Carry Information about Objects in Temporal Context. *Neuron* 81: 1165-1178, 2014.
**Ingvar DH**. "Memory of the future": an essay on the temporal organization of conscious awareness. *Hum Neurobiol* 4: 127-136, 1985.
**Jehee JF, Ling S, Swisher JD, van Bergen RS, and Tong F**. Perceptual learning selectively refines orientation representations in early visual cortex. *J Neurosci* 32: 16747-16753a, 2012.
**Kamitani Y, and Tong F**. Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8: 679-685, 2005.
**Kim T, Kim HR, Kim K, and Lee C**. Modulation of V1 spike response by temporal interval of spatiotemporal stimulus sequence. *PLoS One* 7: e47543, 2012.
**Knowlton BJ, Squire LR, and Gluck MA**. Probabilistic classification learning in amnesia. *Learning & memory (Cold Spring Harbor, NY)* 1: 106-120, 1994.
**Kok P, Brouwer GJ, van Gerven MA, and de Lange FP**. Prior expectations bias sensory representations in visual cortex. *J Neurosci* 33: 16275-16284, 2013.
**Kok P, Failing MF, and de Lange FP**. Prior expectations evoke stimulus templates in the primary visual cortex. *J Cogn Neurosci* 26: 1546-1554, 2014.
**Kok P, Jehee JF, and de Lange FP**. Less is more: expectation sharpens representations in the primary visual cortex. *Neuron* 75: 265-270, 2012.
**Leaver AM, Van Lare J, Zielinski B, Halpern AR, and Rauschecker JP**. Brain activation during anticipation of sound sequences. *J Neurosci* 29: 2477-2485, 2009.
**Meyer T, and Olson CR**. Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc Natl Acad Sci U S A* 108: 19401-19406, 2011.
**Miyashita Y**. Inferior temporal cortex: where visual perception meets memory. *Annu Rev Neurosci* 16: 245-263, 1993.
**Murray SO, Kersten D, Olshausen BA, Schrater P, and Woods DL**. Shape perception reduces activity in human primary visual cortex. *Proc Natl Acad Sci U S A* 99: 15164-15169, 2002.
**Nissen MJ, and Bullemer P**. Attentional requirements of learning - evidence from performance measures. *Cognitive Psychology* 19: 1-32, 1987.
**Pasupathy A, and Miller EK**. Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433: 873-876, 2005.
**Perrett DI, Xiao D, Barraclough NE, Keysers C, and Oram MW**. Seeing the future: Natural image sequences produce "anticipatory" neuronal activity and bias perceptual report. *Q J Exp Psychol (Hove)* 62: 2081-2104, 2009.
**Perruchet P, and Pacton S**. Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences* 10: 233-238, 2006.
**Petrov AA, Dosher BA, and Lu Z-L**. The dynamics of perceptual learning: an incremental reweighting model. *Psychological review* 112: 715, 2005.
**Poldrack RA, Clark J, Pare-Blagoev EJ, Shohamy D, Creso Moyano J, Myers C, and Gluck MA**. Interactive memory systems in the human brain. *Nature* 414: 546-550, 2001.
**Poldrack RA, Prabhakaran V, Seger CA, and Gabrieli JD**. Striatal activation during acquisition of a cognitive skill. *Neuropsychology* 13: 564-574, 1999.

**Rauch SL, Savage CR, Brown HD, Curran T, Alpert NM, Kendrick A, Fischman AJ, and Kosslyn SM**. A PET investigation of implicit and explicit sequence learning. *Human Brain Mapping* 3: 271-286, 1995.
**Rauch SL, Whalen PJ, Savage CR, Curran T, Kendrick A, Brown HD, Bush G, Breiter HC, and Rosen BR**. Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. *Hum Brain Mapp* 5: 124-132, 1997.
**Reber AS**. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 6: 855-&, 1967.
**Rose M, Haider H, Salari N, and Buchel C**. Functional dissociation of hippocampal mechanism during implicit learning based on the domain of associations. *J Neurosci* 31: 13739-13745, 2011.
**Schapiro AC, Gregory E, Landau B, McCloskey M, and Turk-Browne NB**. The Necessity of the Medial Temporal Lobe for Statistical Learning. *Journal of Cognitive Neuroscience* 1-12, 2014.
**Schapiro AC, Kustner LV, and Turk-Browne NB**. Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr Biol* 22: 1622-1627, 2012.
**Schendan HE, Searl MM, Melrose RJ, and Stern CE**. An FMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron* 37: 1013-1025, 2003a.
**Schendan HE, Searl MM, Melrose RJ, and Stern CE**. Sequence? What Sequence?: the human medial temporal lobe and sequence learning. *Mol Psychiatry* 8: 896-897, 2003b.
**Schlack A, and Albright TD**. Remembering visual motion: neural correlates of associative plasticity and motion recall in cortical area MT. *Neuron* 53: 881-890, 2007.
**Schoups A, Vogels R, Qian N, and Orban G**. Practising orientation identification improves orientation coding in V1 neurons. *Nature* 412: 549-553, 2001.
**Schwarb H, and Schumacher EH**. Generalized lessons about sequence learning from the study of the serial reaction time task. *Advances in cognitive psychology / University of Finance and Management in Warsaw* 8: 165-178, 2012.
**Serences JT, Ester EF, Vogel EK, and Awh E**. Stimulus-specific delay activity in human primary visual cortex. *Psychol Sci* 20: 207-214, 2009.
**Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, and Tootell RB**. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268: 889-893, 1995.
**Shohamy D, and Wagner AD**. Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron* 60: 378-389, 2008.
**Smith FW, and Muckli L**. Nonstimulated early visual areas carry information about surrounding context. *Proc Natl Acad Sci U S A* 107: 20099-20103, 2010.
**Summerfield C, and Egner T**. Expectation (and attention) in visual cognition. *Trends Cogn Sci* 13: 403-409, 2009.
**Szpunar KK, Addis DR, McLelland VC, and Schacter DL**. Memories of the future: new insights into the adaptive value of episodic memory. *Front Behav Neurosci* 7: 47, 2013.
**Tong F, and Pratte MS**. Decoding patterns of human brain activity. *Annu Rev Psychol* 63: 483-509, 2012.

## Figures

**Figure 1. Task design and behavioural performance. (a)** Prediction task: in each trial, participants were presented with a structured sequence of 8 gratings followed by a blank interval (11.6 s). After this period, a test grating, preceded by a brief cue (black square, 0.5 s), was presented and the participants had to indicate whether the orientation of the test grating matched their expectation or not. **(b)** Control task: the participants were instructed to attend to the sequence and indicate whether a cued grating ('R' or 'L') matched the orientation of the test stimulus or not. The timing for this task was matched to the prediction task. **(c)** Mean proportion of correct responses for each run across training sessions. Data is shown for 12/16 participants that showed improvement after training, excluding participants (n=4) who did not show improvement in the task after training (57% mean performance at the last training session). Data across runs were fitted using least squares non-linear fit. Data is shown for 16 blocks (4 training sessions); one participant completed only three sessions, as performance had already saturated above 80%; the rest of the participants completed either 4 (n = 4) or 5 (n = 7) training sessions. **(d)** Mean proportion of correct responses during scanning before and after training for both the prediction and control task. Error bars indicate the S.E.M. Behavioral data for all 16 participants showed a similar pattern of results; that is, we observed a significant effect of *session* ($F_{1,11} = 144.13$, $P < .001$) and *task* and *session* interaction ($F_{1,11} = 76.08$, $P < .001$)

**Figure 2. Univariate vs. multivariate analysis of BOLD signals. (a)** Univariate mean trial time course (across voxels and participants) of BOLD responses in V1 (percent signal change calculated in relation to the average signal across the whole run) for the pre- (light blue) and post-training scanning sessions (dark blue). The shaded grey area (volumes 8 to 13) indicates the volumes used to decode the participants' prediction after accounting for the hemodynamic lag. Volume 7 was not used to avoid confounding activation from the sequence presentation.

**(b)** Mean decoding SVM accuracy (proportion correct) per fMRI volume of the participants' predictions before (light blue) and after (dark blue) training. Note that per-volume fMRI signals are noisier than signals averaged across volumes resulting in lower MVPA accuracy. Error bars indicate the S.E.M.

**Figure 3. Comparing MVPA for the prediction and control tasks.** Mean classification accuracy (across volumes 8 to 13 and participants) of the predicted orientation before (light colours) and after (dark colours) training in early visual areas for the prediction (blue) and control task (red). Error bars indicate the S.E.M.

**Figure 4. Linking classification accuracy and behavioral performance**. (a) Correlation between decoding accuracy for predicted orientations in V1 and performance (percentage of correct predictions) in the prediction task. No significant correlations ($P > 0.05$) were observed for other visual areas; (b) Decoding accuracy across visual areas for the weak-learners (n=4) in the prediction task before and after training.

**Figure 5. Classifier generalization performance.** (a) Classification accuracy for decoding the two grating orientations (rightwards and leftwards) in an independent block design experiment. Decoding (mean classification accuracy) of (b) predicted (prediction task) and (c) cued (control task) orientations before (light blue) and after (dark blue) training using a classifier trained on physical stimulus orientations.

**Figure 6. Decoding the expected orientation as determined by the preceding sequence.** Mean classification accuracy (across volumes 8 to 13 and participants) of decoding the expected orientation as determined by the presented sequence before (light blue) and after (dark blue) training in early visual areas. Error bars indicate the S.E.M.

Figure 1

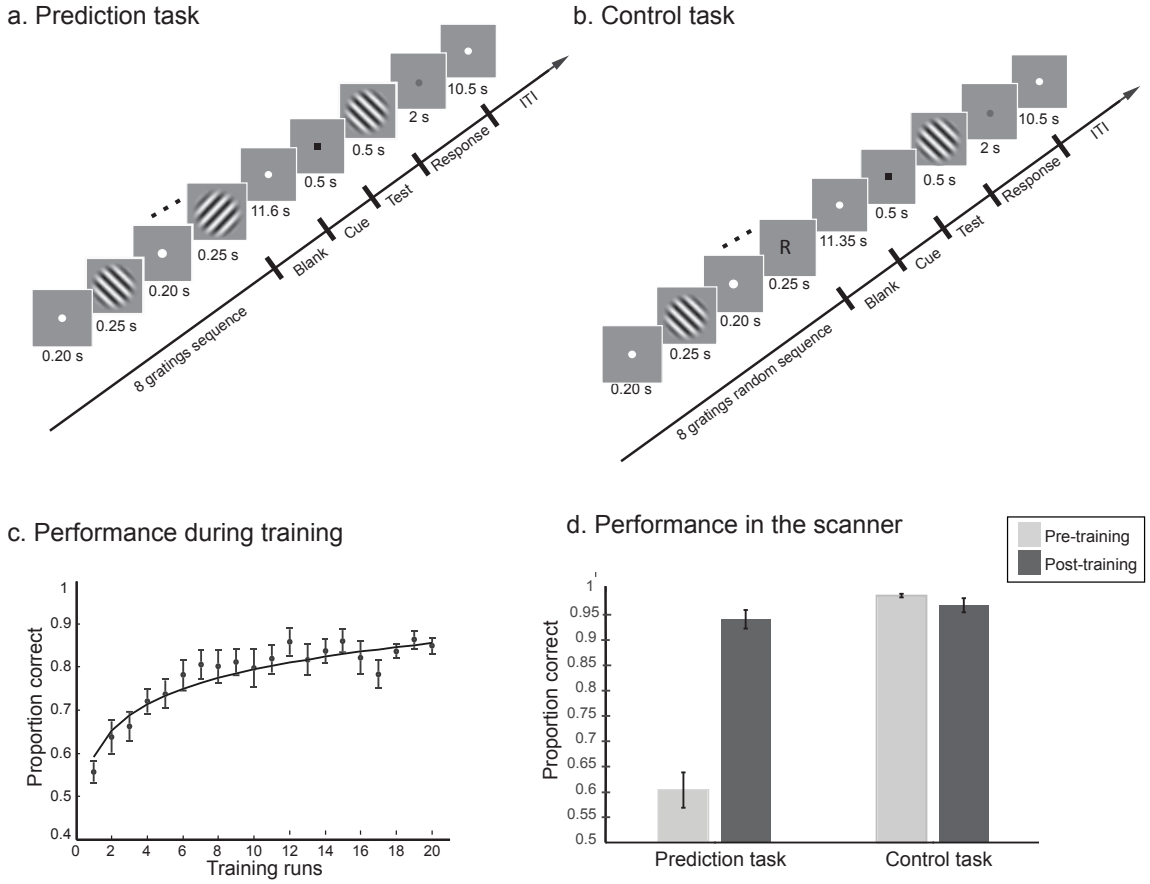
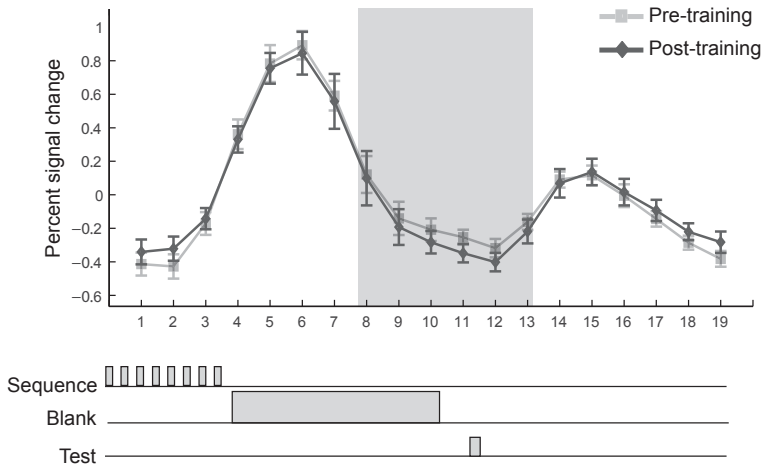

a. Prediction task
0.20 s
0.25 s
0.20 s
0.25 s
11.6 s
0.5 s
0.5 s
2 s
10.5 s
8 gratings sequence
Blank
Cue
Test
Response
ITI
b. Control task
0.20 s
0.25 s
R
0.25 s
11.35 s
0.5 s
0.5 s
2 s
10.5 s
8 gratings random sequence
Blank
Cue
Test
Response
ITI
c. Performance during training
Proportion correct
Training runs
d. Performance in the scanner
Pre-training
Post-training
Proportion correct
Prediction task
Control task

Figure 2

a. Percent signal change per volume




b. Classification accuracy per volume

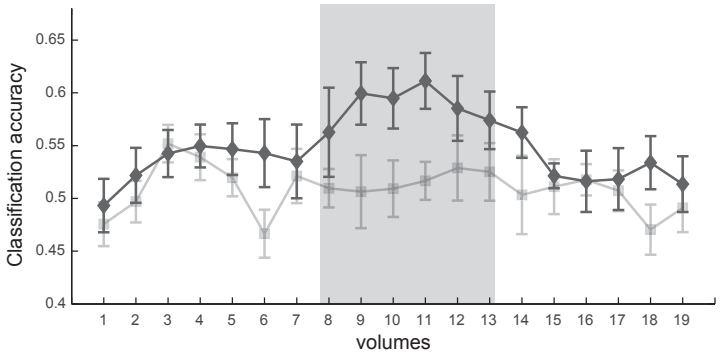
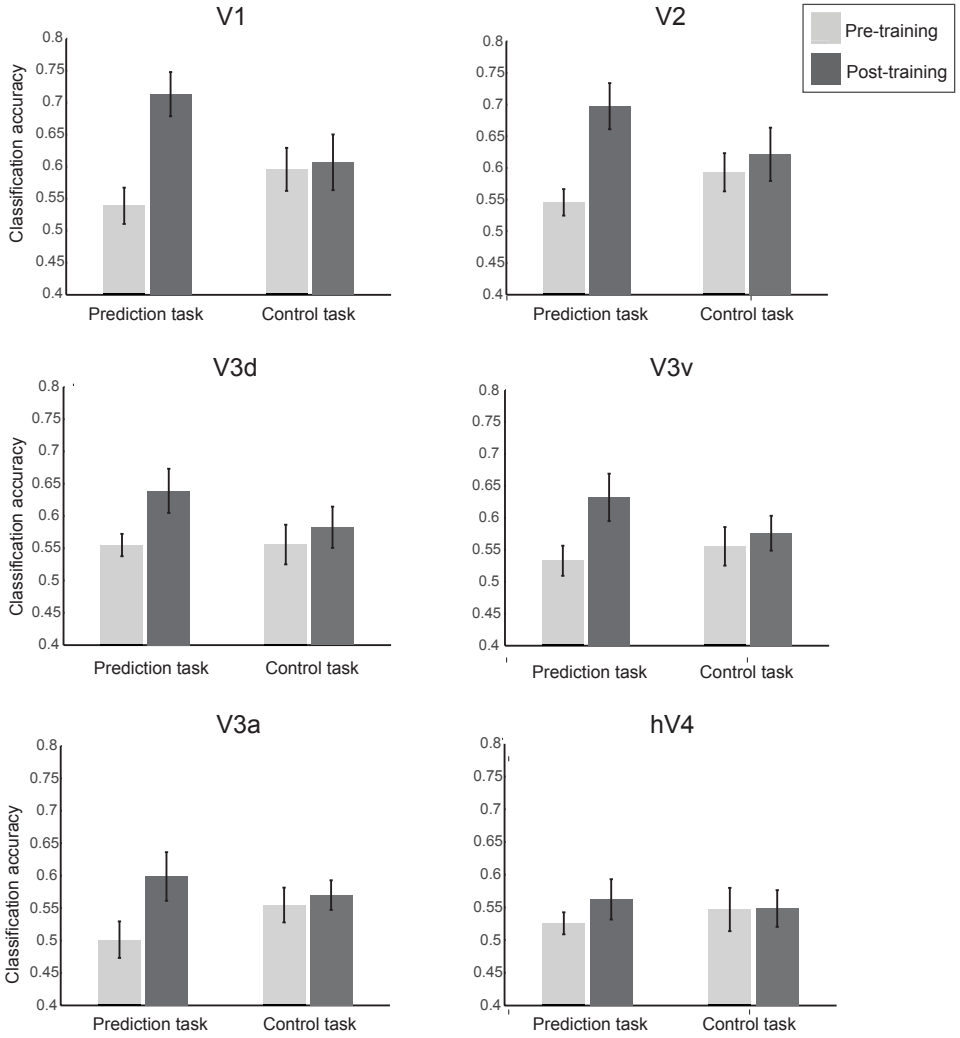
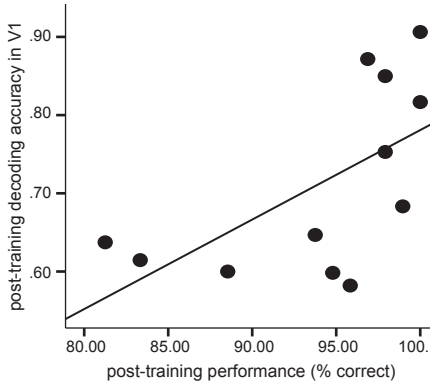


Figure 3


V1
V2
V3d
V3v
V3a
hV4
Pre-training
Post-training
Classification accuracy
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
Prediction task
Control task

igure 4

a.


post-training decoding accuracy in V1
.90
.80
.70
.60
80.00
85.00
90.00
95.00
100.
post-training performance (% correct)

b.

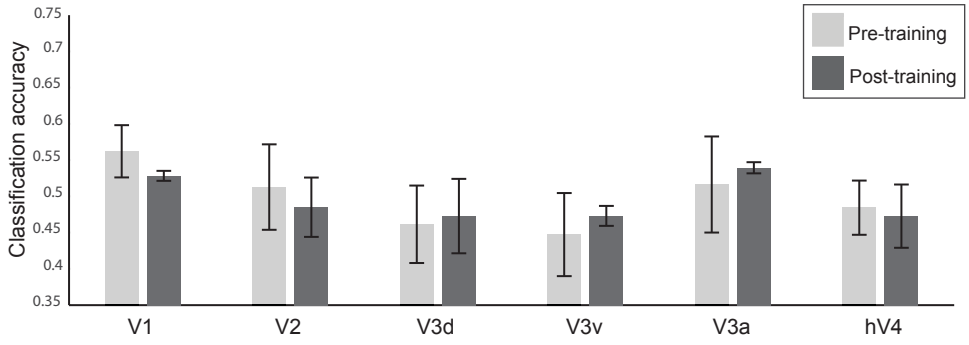
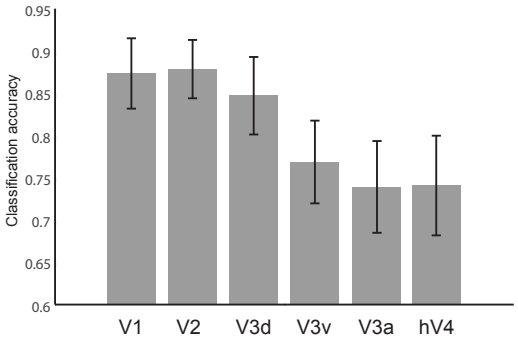
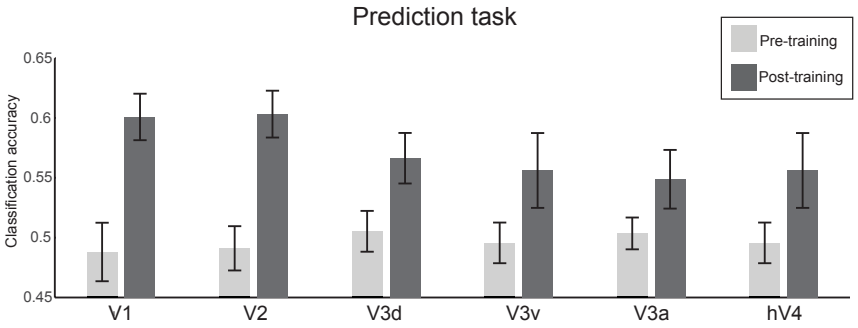

Classification accuracy
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
Pre-training
Post-training
V1
V2
V3d
V3v
V3a
hV4

Figure 5

a.


Classification accuracy
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
V1
V2
V3d
V3v
V3a
hV4

b.


Prediction task
Pre-training
Post-training
Classification accuracy
0.65
0.6
0.55
0.5
0.45
V1
V2
V3d
V3v
V3a
hV4

c.

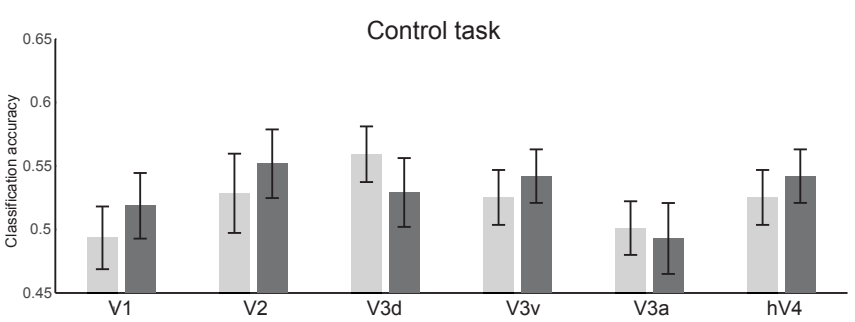
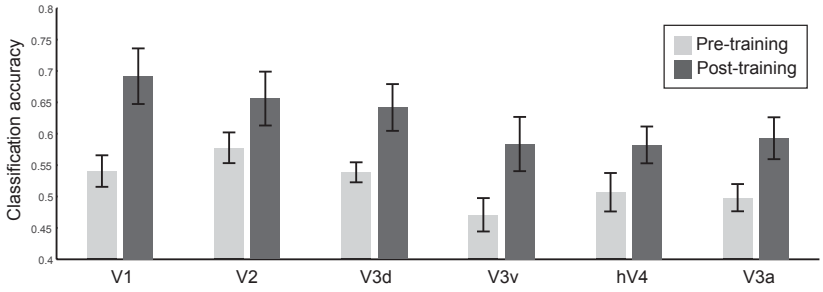

Control task
Classification accuracy
0.65
0.6
0.55
0.5
0.45
V1
V2
V3d
V3v
V3a
hV4

Figure 6


Classification accuracy
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
Pre-training
Post-training
V1
V2
V3d
V3v
hV4
V3a