Neuron
Article

Learning Shapes the Representation of Behavioral Choice in the Human Brain

Sheng Li,[1] Stephen D. Mayhew,[1] and Zoe Kourtzi[1],*
[1]School of Psychology, University of Birmingham, Birmingham B15 2TT, UK
*Correspondence: z.kourtzi@bham.ac.uk
DOI 10.1016/j.neuron.2009.03.016

SUMMARY

Making successful decisions under uncertainty due to noisy sensory signals is thought to benefit from previous experience. However, the human brain mechanisms that mediate flexible decisions through learning remain largely unknown. Comparing behavioral choices of human observers with those of a pattern classifier based on multivoxel single-trial fMRI signals, we show that category learning shapes processes related to decision variables in frontal and higher occipitotemporal regions rather than signal detection or response execution in primary visual or motor areas. In particular, fMRI signals in prefrontal regions reflect the observers' behavioral choice according to the learned decision criterion only in the context of the categorization task. In contrast, higher occipitotemporal areas show learning-dependent changes in the representation of perceived categories that are sustained after training independent of the task. These findings demonstrate that learning shapes selective representations of sensory readout signals in accordance with the decision criterion to support flexible decisions.

INTRODUCTION

Successful everyday actions rely on the brain making informed choices based on an assessment of the demands of the environment and the costs and benefits associated with different behaviors. In selecting the best choice of action, the brain has to deal with information that is inherently uncertain at different stages of the decision process: from extracting a robust estimate of the current state of the environment given noisy sensory signals to judging the rewards and risks associated with a particular decision (Platt and Huettel, 2008; Rushworth and Behrens, 2008; Schultz et al., 2008; Yu and Dayan, 2005). Recent theoretical work suggests ways that the primate brain meets this challenge by taking into account knowledge from previous experience (Daw and Doya, 2006; Sutton and Barto, 1998). However, a thorough understanding of the human brain plasticity mechanisms that mediate learning to support efficient and flexible decisions remains a considerable open challenge (Dayan and Niv, 2008; O'Doherty et al., 2007).

Computational models (Smith and Ratcliff, 2004) and experimental studies (for reviews, Glimcher, 2003; Gold and Shadlen, 2007; Heekeren et al., 2008; Schall, 2001) suggest that decision making comprises a set of interacting processes. First, sensory signals are detected and decision variables (i.e., quantities that relate to the likelihood of an event occurring) are computed by comparing and accumulating sensory evidence toward a threshold or criterion (i.e., the internal representation of a comparison quantity) for response. A second process monitors uncertainty in the sensory signals and the probable outcome of a choice as well as task demands. Further, performance during a task is monitored for errors and the need to adjust strategies for optimizing decisions. Recent neurophysiology and brain imaging studies have identified the neural circuits involved in these processes. However, the role of learning in shaping these processes and the human brain circuits involved in decision making remain largely unknown.

Here, we investigate how category learning shapes decision making processes in the human brain by combining psychophysical measurements and advanced fMRI analysis methods (multivoxel pattern analysis, MVPA) that are sensitive to neural information encoded at a finer-scale than the standard resolution of fMRI measurements (Cox and Savoy, 2003; Haynes and Rees, 2006; Norman et al., 2006). We employed a categorization task in which observers were presented with stimuli from a morphing space generated by varying the spiral angle between radial and concentric patterns (Figure 1A). Observers were asked to decide whether the viewed stimulus was radial or concentric. This task required observers to compare the sensory input (i.e., the stimulus on each trial) to an internal decision criterion that was defined as the categorical boundary in the stimulus space. Uncertainty in this task increased as stimuli approached the boundary between the stimulus categories (45° spiral angle based on physical stimulus properties). We investigated how learning shapes the observers' behavioral choice in two separate experiments. In the first experiment, observers were trained to use different decision criteria (i.e., category boundaries) when performing categorical decisions in each of two sessions. In the second experiment, we measured the observers' performance before and after training on a given decision criterion. In both experiments, we were able to shift the observers' categorical boundary after training with feedback, thereby dissociating the physical stimuli from their categorical interpretation.

Using fMRI, we asked which cortical regions carry the neural signature of this learning-dependent flexibility in categorical

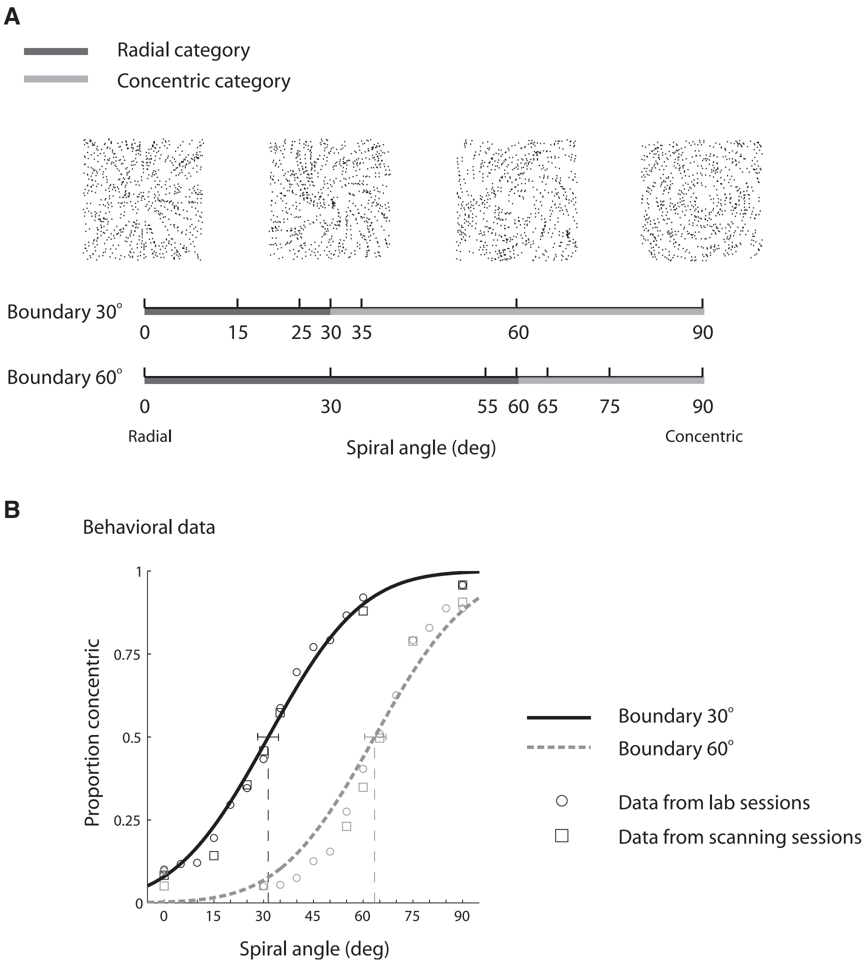

Figure 1. Stimulus and Behavioral Data

(A) Stimuli: four example Glass pattern stimuli (100% signal) at spiral angles of 0°, 30°, 60°, and 90°. The two boundaries and spiral angles tested are shown (black bar, stimuli that resemble radial; gray bar, stimuli that resemble concentric) that indicate the categorical membership of the stimuli for each boundary.

(B) Behavioral data from the lab (circles) and the scanner (squares) for each boundary. Lines indicate the cumulative Gaussian fits of the behavioral data from the lab. Error bars indicate the 95% confidence interval at 50% concentric threshold.

decisions. We reasoned that such regions would show trial-by-trial variations that correspond to the behavioral choice of the observers and change to reflect the shift in the decision criterion (i.e., category boundary). As the learning manipulation we employed changed the decision criterion and uncertainty during the categorization task, we predicted that frontal circuits previously implicated in these decision processes would show experience-dependent changes in neural processing with learning. Consistent with this prediction, we show that ventral prefrontal and premotor regions contain information that allows us to reliably decode (1) the observers' choice on single trials and (2) learning-dependent changes on the decision criterion used by the observers for classifying sensory input. Interestingly, these learning-dependent changes in decision processes were also evident in higher occipitotemporal regions that are known to be involved in the encoding of global form patterns. Importantly, the learned categories were represented in occipitotemporal regions independent of the task performed by observers during scanning suggesting learning-dependent changes in the representation of the perceived categories rather than simply task-related modulations. In contrast, no changes were observed in primary visual or motor areas indicating that learning did not change signal detection or response execution related processes. These findings provide evidence that category learning shapes neural representations to reflect the observers' behavioral choice during categorical decisions. In particular, in prefrontal circuits learning shapes the estimation of the decision criterion and task uncertainty, while in higher occipitotemporal regions the representations of perceived categories that may serve as selective readout signals for optimal decisions.

RESULTS

Behavioral Data: Learning-Dependent Changes on the Decision Criterion

We tested observers' ability to categorize global form patterns as radial or concentric (Figure 1A) and plotted their performance (proportion concentric) as a function of stimulus spiral angle (psychometric function). Before training (pretraining test), the mean categorization boundary (50% point on the psychometric function) was 48.96° (±3.57°) spiral angle, matching closely the mean of the physical stimulus space (45° spiral angle). We then trained the observers with feedback to assign stimuli into categories based on two different category boundaries: 30°, 60° spiral angle (see Experimental Procedures for details). Observers were first trained on one of the two boundaries and then retrained on the other. Testing the observers without feedback after training demonstrated that training had shifted the observers' criteria for categorization to 31.4° (±3.15°) for the 30° boundary and 63.5° (±3.24°) for the 60° boundary (Figure 1B). Fitting the behavioral data with a cumulative Gaussian (see Supplemental Data available online) showed a significant shift (bootstrapped 95% confidence intervals) in the threshold of the psychometric function when observers were trained with different categorization boundaries. Similar effects were observed during the scanning sessions (30° boundary, 34.4° ± 0.84°; 60° boundary, 64.1° ± 0.87°), indicating that the training procedure successfully modified the observers' decision criterion (i.e., categorization boundary).

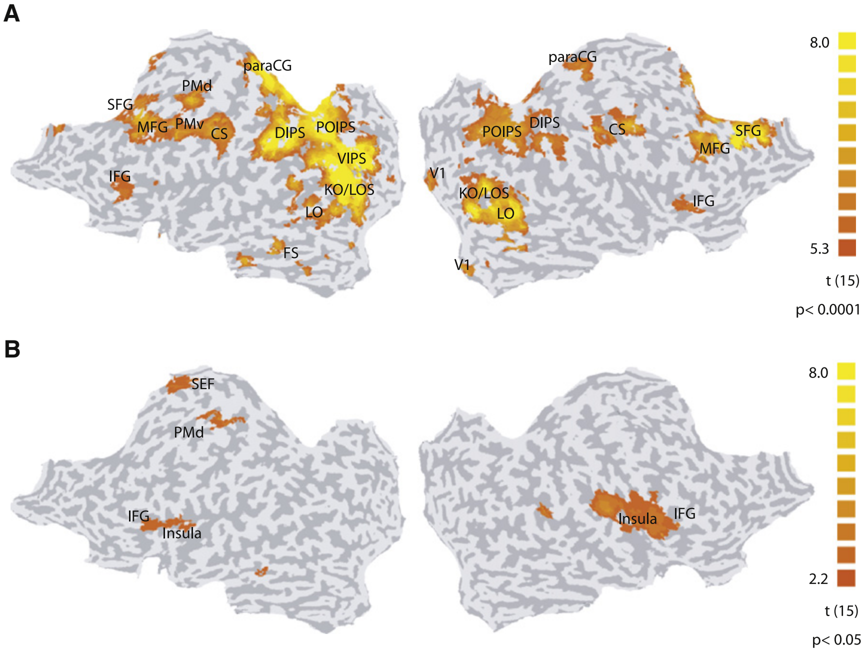

Figure 2. Activation Patterns Based on the Observers' Behavioral Choice

(A) Searchlight map (data across observers and task on the two boundaries) showing areas with significantly higher accuracy than chance (50% correct) ($p < 0.0001$, cluster threshold estimation 5 mm^2) for the classification of fMRI signals based on the observers' behavioral choice.

(B) Group map (data across observers and task on the two boundaries) based on covariance analysis. *t* value maps are superimposed on flattened cortical surfaces of both hemispheres (Table S1: Talairach coordinates).

Predicting Categorical Decisions from Activation Patterns across the Whole Brain

We first investigated which cortical areas contain information that is diagnostic of the observer's categorical decision (radial versus concentric) on each trial. Using a multivoxel searchlight approach (Kriegeskorte et al., 2006), we tested the accuracy of a linear support vector machine (SVM) in classifying fMRI signals based on each observer's behavioral choice (radial versus concentric) per trial. We performed this classification analysis on fMRI data when observers performed the categorization task based on the 30° or 60° boundary conditions. We defined an aperture (9 mm radius) that was moved serially throughout the entire cortex to assess the information content of voxel patterns in different brain regions. This analysis (Figures 2A and S1A) revealed voxel patterns with classification accuracy significantly higher than chance ($p < 0.0001$, between-observers, cluster threshold estimation 5 mm^2) in occipitotemporal areas (V1, KO/LOS, LO), parietal regions along the IPS (VIPS, POIPS, DIPS), and frontal regions in the motor cortex (CS, PMd, PMv), ventral (IFG/insula) and dorsal prefrontal (MFG, SFG) cortex. These findings suggest that activation patterns in frontoparietal and occipitotemporal regions afford the reliable prediction of observers' categorical decisions from single trial fMRI data.

In contrast, a standard GLM analysis (Figure 3A) on the univariate signals showed weak activations in a subset of these areas. Further, analysis of the average fMRI response across voxels per area did not show any significant differences between concentric and radial trials (Figure 3B). These results show that MVPA capitalizes on these small biases, to discern statistically reliable differences related to the behavioral choice. This is consistent with the higher sensitivity of multivariate methods in detecting neural preferences encoded at a finer spatial resolution than that of typical fMRI measurements.

To constrain our analysis and select regions of interest, we tested for interactions between areas in the identified cortical network based on the hypothesis that higher-level areas may pool the output of neural populations in sensory areas to form decisions (Heekeren et al., 2004; Kim and Shadlen, 1999; Newsome et al., 1989; Romo and Salinas, 2003; Shadlen and Newsome, 2001). In particular, we performed MVPA based on the observers' behavioral choice (radial versus concentric) on voxel patterns in KO/LOS and LO (pattern size of 200 voxels) that are known to be involved in the selective representation of visual forms (Ostwald et al., 2008). We then correlated the output of the MVPA classifier across trials with the time course from each voxel across the whole brain. Significant correlations were identified in areas IFG/Insula, PMd, and SEF in which the fMRI signal covaried with the difference in the response for radial versus concentric stimuli in sensory areas (Figure 2B; Supplemental Data).

Learning-Dependent Plasticity: Comparing fMR-Metric and Psychometric Functions

We tested how learning shapes the neural processing in frontal and occipitotemporal regions involved in the representation of categorical choices. We asked which of these cortical regions show changes in fMRI activation patterns that relate to the behavioral changes in the observers' decision criterion after training as shown by the shift in the observers' psychometric functions (Figure 1B). For each observer, we identified activation patterns based on the searchlight multivariate analysis (Figure 2). We focused on activation patterns in extrastriate visual areas (KO/LOS, LO) that showed strong preferential response for radial versus concentric stimulus choices, higher frontal areas (IFG, PMd, SEF) that were shown to covary with stimulus bias in sensory areas (KO/LOS, LO) and primary visual (V1) and motor (CS) cortex that are involved in the analysis of the physical input and the execution of the response, respectively. We trained a linear SVM to classify fMRI signals in these areas based on the observer's behavioral choice (radial versus concentric) on each trial and tested for accuracy in predicting the observers' choice for an independent dataset. For each observer, we calculated the mean performance of the classifier (proportion of trials

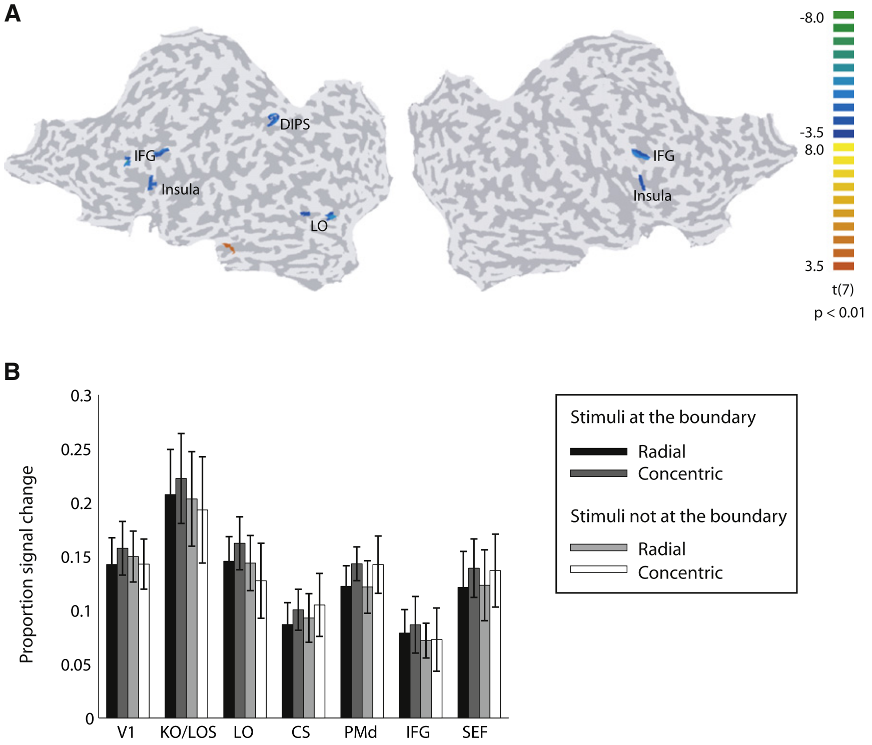

Figure 3. Univariate Analysis

(A) Random effect GLM analysis (across observers and task on the two boundaries) based on the observers' behavioral choices (radial versus concentric) ($p < 0.01$, cluster threshold estimation 28 mm^2).

(B) Analysis of fMRI time courses: for each ROI, we extracted fMRI responses to radial versus concentric trials based on the observers' responses. fMRI signals are shown for stimuli at the category boundary (30° and 60°) for each session and the same stimuli when they were not assigned as the category boundary. Proportion signal change from fixation baseline is plotted for radial versus concentric choices for each stimulus condition (at the boundary or not). Error bars indicate the standard error of the mean across observers.

classified as concentric for each stimulus condition) across cross-validations. We calculated the mean classifier performance across observers and plotted the data after linear scaling (i.e., for each fMR-metric function we scaled each prediction by subtracting the minimum value across conditions and dividing by the difference between maximum and minimum prediction values) to allow comparison between cortical areas. We fitted the data (Figure 4, scaled data; Figure S2, nonscaled data) using a cumulative Gaussian, just as we had done for the behavioral data. We refer to these functions as fMR-metric functions.

In line with changes in psychophysical performance based on training, fMR-metric functions in frontal (IFG, PMd) and extrastriate visual (KO/LOS, LO) areas showed significant shifts (nonoverlapping bootstrapped confidence intervals) that matched training-induced shifts in the category boundary (Figure 4A). We quantified this using the 50% threshold for fMR-metric functions obtained when observers performed the categorization task on the two different boundaries (30° versus 60°; Figure 4B). We observed significant changes in the 50% threshold of the psychometric ($F(1,7) = 70.82$, $p < 0.001$) and fMR-metric ($F(1,7) = 5.4$, $p < 0.05$) functions related to the categorization task on the 30° versus 60° boundary. However, the slopes of the psychometric ($F(1,7) < 1$, $p = 0.86$) and fMR-metric ($F(1,7) = 1.9$, $p = 0.21$) functions did not change significantly, suggesting behavioral and neural changes in the criterion for categorization rather than sensitivity to the stimuli. Importantly, these effects were not universal. In particular, the 50% thresholds for fMR-metric functions in primary visual (V1) and motor (CS) cortex were not shifted significantly between sessions, suggesting that training did not change processing in these regions. This result was confirmed by an additional analysis (Supplemental Data) based on the common stimulus conditions (0°, 30°, 60°, and 90° spiral angles) between the two sessions showing significant correlations between behavioral and classifier performance within the same session but not across sessions with different category boundaries for KO/LOS ($F(1,7) = 8.46$, $p < 0.05$), LO ($F(1,7) = 12.99$, $p < 0.01$), PMd ($F(1,7) = 7.39$, $p < 0.05$), IFG ($F(1,7) = 17.25$, $p < 0.01$), but not SEF (($F(1,7) < 1$, $p = 0.97$), V1 ($F(1,7) < 1$, $p = 0.79$) or CS ($F(1,7) < 1$, $p = 0.81$). Interestingly, for V1 the 50% thresholds of fMR-metric functions for both sessions were close to the mean (45°) of the physical stimulus space rather than the learned category boundary, suggesting a physical based representation of the stimuli that remained unaffected by learning induced changes in categorization. This result in V1 suggests that the shift in the fMR-metric functions in frontal and higher occipitotemporal areas could not be simply due to low-level differences between the stimulus categories presented in the two sessions. Further, it is possible that the results observed in the motor cortex reflect a small bias in the mapping between behavioral choices and button presses (Figure S3A). To decouple the observers' decision (radial versus concentric choice) from the motor response (button press) we introduced a cue that switched across trials indicating two different mappings between behavioral choice and button presses. The behavioral psychometric functions (Figure 1B) revealed that observers were very good at changing their button press behavior according to the cue. However, analysis of the distribution of button presses across conditions showed that observers had a small idiosyncratic preference for one of the mappings. This resulted in a small bias in the distribution of button presses across trials that may drive the BOLD signal in the primary motor cortex and be discriminated by the classifier. The idiosyncratic nature of the button press preferences means that there is no significant shift in the fMR-metric functions between the two sessions in CS. This result suggests that the shifts in the fMR-metric functions for prefrontal and higher occipitotemporal areas could not be explained simply on the basis of motor responses.

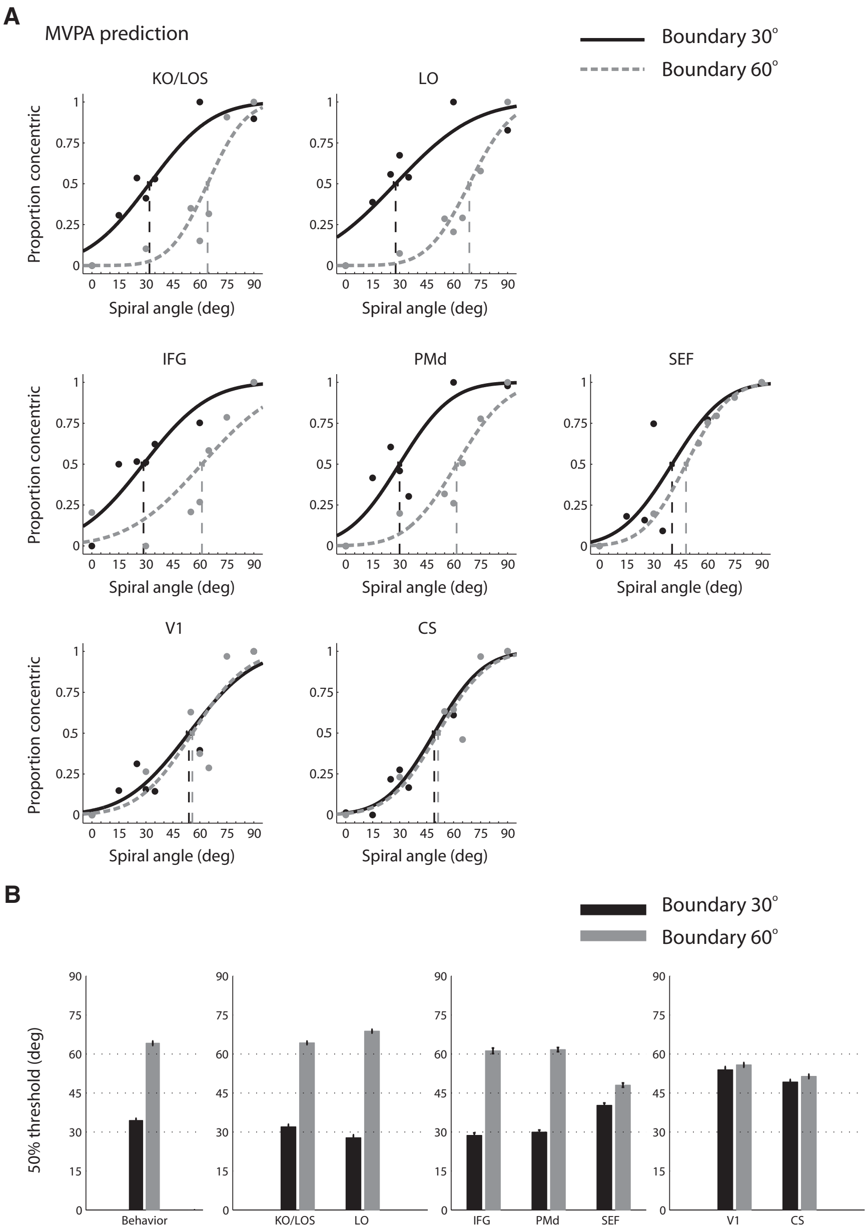

Figure 4. fMR-Metric Curves Based on the Observers' Behavioral Choice

(A) fMR-metric curves based on the observers' behavioral choice for each task (30° and 60° boundary) and ROI. Average classifier prediction data across observers are scaled from 0 to 1 and fitted with cumulative Gaussian functions. Error bars indicate the 95% confidence interval at 50% concentric threshold. Significant fits were obtained across ROIs and task on the two boundaries (lowest Pearson correlation coefficient R = 0.838, p = 0.018).

(B) Mean 50% concentric thresholds across observers are shown for the psychometric and fMR-metric data collected in the scanner for each ROI. Error bars indicate the 95% confidence interval at 50% concentric threshold.

As an extension of this analysis, we tested the performance of the SVM classifier when it was trained only on fMRI data obtained when observers viewed the radial (0°) and concentric (90°) stimuli. Having trained the classifier at the extremes of the stimulus classes, we then tested its performance for intermediate stimuli (i.e., stimulus spiral angles between 0° and 90°). We observed consistent shifts in the fMR-metric functions in prefrontal and higher occipitotemporal areas (Figure 5). Moreover, fitting the fMRI data using a scaled version of the psychometric function obtained during scanning showed that fMR-metric functions in frontal (IFG, PMd) and extrastriate visual (KO/LOS, LO) areas (but not SEF, V1 or CS) were shifted in correspondence with the learned categorization boundary (Figure 6A). To control for the possibility that these results are due to random correlations in the data, we correlated the psychometric data with an fMR-metric function computed from randomly permuted fMRI patterns (i.e., we randomized the correspondence between fMRI data and training labels and estimated the classifier prediction for each stimulus condition). The lack of significant correlations (Figure 6B) in this control analysis supports our interpretation for a link between task-relevant behavioral performance and neural preferences.

Further, we used a trial-by-trial analysis borrowing techniques from physiological studies (e.g., Britten et al., 1996; Uka and DeAngelis, 2003). We used choice probability analysis to quantify the relationship between the observers' choice and fMRI responses. We used voxel pattern responses (i.e., classification accuracy) evoked by stimuli near threshold performance (i.e., stimuli at the boundary and ±5° of spiral angle), as observers' performance included a useful number of errors for these stimuli. To avoid circularity in the analysis, we trained the classifier based on one dataset and estimated choice probabilities based on an independent data set. In particular, for each observer, we labeled all trials based on the observer's behavioral choice and tested the classifier's performance in predicting the observer's choice using a leave-one-run-out cross-validation procedure. We compared the output (i.e., discriminant function value) of the classifier for trials associated with a radial choice against trials associated with a concentric choice. We then plotted the probability with which the classifier predicted the behavioral choice per stimulus trial on an independent data set (ROC: receiver operator characteristic curve). The area under the ROC curve signified the choice probability for each observer and cortical region. We evaluated the significance of choice probabilities per observer using a bootstrap procedure (1000 random permutations of the trial labels). As

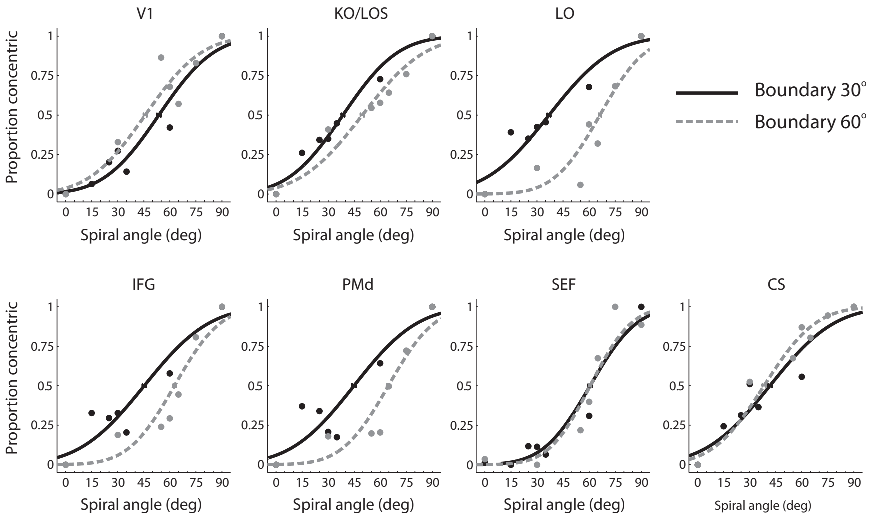

Figure 5. fMR-Metric Curves Based on the Classification of Stimuli 0° versus 90°

fMR-metric curves based on the classification of stimuli 0° versus 90° for each task (30° and 60° boundary) and ROI. Average classifier prediction data across observers are scaled from 0 to 1 and fitted with cumulative Gaussian functions. Error bars indicate the 95% confidence interval at 50% concentric threshold. Significant fits were obtained across ROIs and task on the two boundaries (lowest Pearson correlation coefficient $R = 0.912$, $p = 0.004$).

the stimuli at or nearby the boundary were ambiguous, we reasoned that significant ($p < 0.05$, based on permutation test) choice probabilities would indicate activity associated with the observers' perceptual interpretation of the stimuli. Significant choice probabilities (values outside the 95% confidence intervals of the random distributions) were observed in 7/8 observers in frontal areas (IFG, mean = 0.56; max = 0.6; PMd, mean = 0.55; max = 0.61), 5/8 observers in occipitotemporal areas (KO/LOS, mean = 0.55, max = 0.62; LO, mean = 0.56, max = 0.61) while only in 3/8 observers in V1, CS, and SEF. These choice probability values are comparable to those reported by previous fMRI (e.g., Hampton and O'Doherty, 2007; Pessoa and Padmala, 2007) and physiology (e.g., Britten et al., 1996; Uka and DeAngelis, 2003) studies using similar methods to decode the observers' choice from single trial signals. These rather low but significant values could be due to lower fMRI signals recorded when event-related designs are used for investigating trial-by-trial responses and the weak response biases at the resolution of fMRI voxels exploited by the MVPA. Finally, choice probabilities in frontal and occipitotemporal areas were significantly higher than in primary visual and motor areas ($F(1,80) = 4.9$, $p < 0.05$). Thus, this analysis is in good agreement with our preceding results in suggesting that pattern-based fMRI responses primarily in prefrontal and occipitotemporal areas account for the behavioral choice of the observers.

The design of our study allowed us to rule out a number of less likely interpretations of our results. First, our design ensured that the observers were not biased in their responses by equating the number of conditions and stimuli across categories. As a result, the stimulus set tested in the two sessions could not remain identical when the category boundary changed across sessions. However, our design allowed us to directly compare between critical stimulus conditions (0°, 90°, 30°, 60°) that were common in the two sessions. Analysis of the univariate fMRI signals (Figure 3B) showed similar fMRI responses across stimulus conditions (i.e., when the stimuli were interpreted as the category boundary or not) suggesting that differences in the MVPA performance reflect the observers' behavioral choice rather than differences in the stimulus statistics across conditions. Second, the learning-dependent changes we observed in the representation of categorical decisions could not be due to differences in task difficulty across conditions, as the classification analysis compared trials associated with different behavioral responses (radial versus concentric) for each stimulus condition. Third, the cued-delay paradigm we used controlled for differences in the observers' response time. That is, observers made their decision during the delay after stimulus offset and waited for the cue before they could select the correct motor response, resulting in similar response times across stimulus conditions. As the stimulus-response association was randomized across trials, the motor response could not be anticipated on a given trial. As an additional control, we used the searchlight approach to search for brain patterns that contained reliable information with which to classify the finger (i.e., button press) used by the observers to indicate their behavioral choice. No significant accuracies for this classification were observed in occipitotemporal (KO/LOS, LO) or the IFG/insula regions (Figures S3B and S3C), suggesting that results in these areas can not be simply explained on the basis of motor responses. Finally, eye movement recordings during scanning showed that there were no significant differences in the eye position, number and amplitude of saccades across stimulus conditions and categorization on the two boundaries (Figure S4). This analysis suggests that it is unlikely that the learning-dependent changes we observed were significantly confounded by eye movements.

Control Experiment: Task-Related Learning Changes

Our results show that frontal and higher occipitotemporal regions contain information about the behavioral choice of the observers and change their processing with learning to reflect the observers' criterion (i.e., categorical boundary) in making categorical judgments. Next, we investigated whether these learning-related changes depend on the task performed by the observers during scanning.

We tested two groups of observers for two sessions. In the first session, observers were instructed to categorize the stimuli as radial versus concentric (same task as in main experiment) without feedback, allowing us to determine the categorical boundary per observer before training. In the second session, observers were trained with feedback to categorize the stimuli based on a predefined boundary (30° or 60° spiral angle).

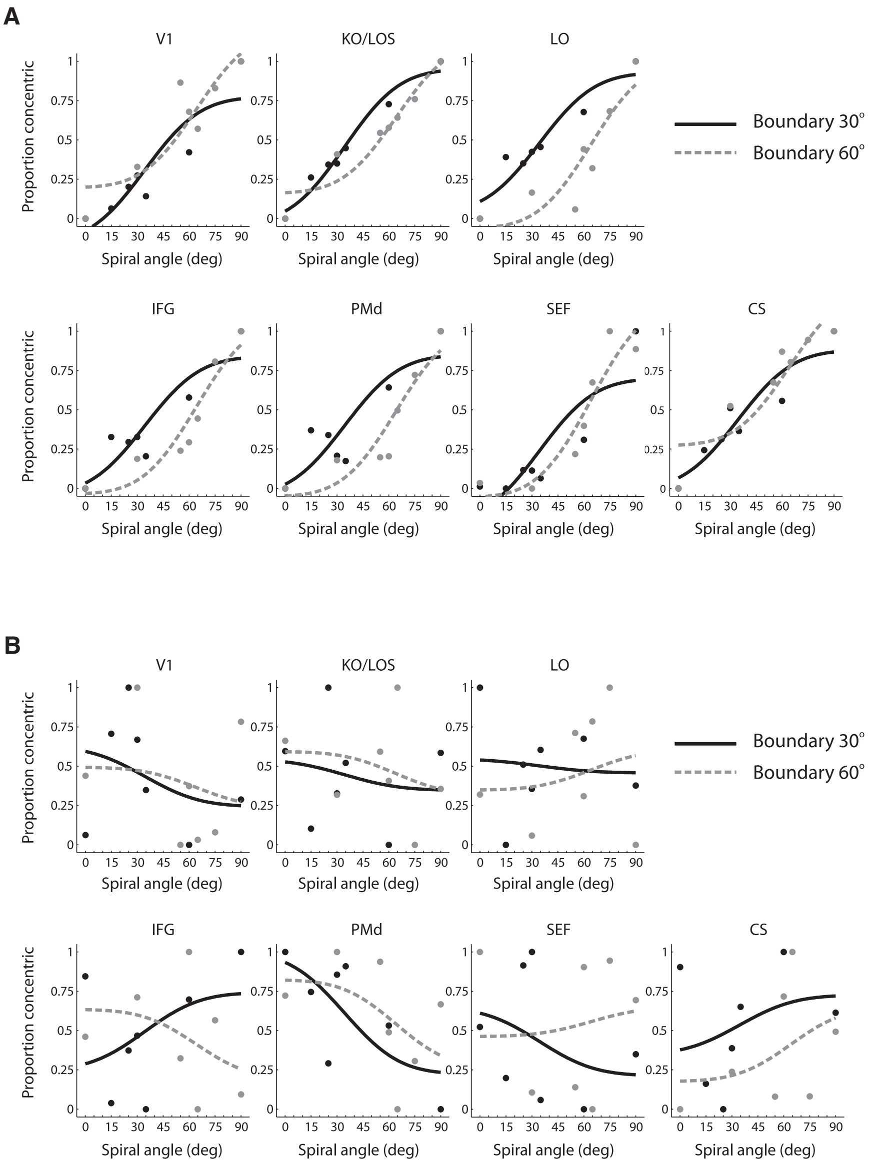

Figure 6. Correlating Psychometric and fMR-Metric Functions

(A) Fitting of fMR-metric curves based on the classification of stimuli 0° versus 90° with the scaled psychometric function obtained from the behavioral data (Supplemental Data). Significant fits were obtained across ROIs and task on the two boundaries (lowest Pearson correlation coefficient $R = 0.841$, $p = 0.018$).

(B) Fitting of fMR-metric curves based on the classification of stimuli 0° versus 90° when the data labels were permuted with the scaled psychometric function. No significant fits were observed.

However, during scanning observers performed a target detection task (i.e., pressed a button to indicate the presence of a square stimulus pattern rendered by dot pairs similar to the radial and concentric stimuli). The observers' performance ranged from 56.8% to 83.2% correct for response times between 839 ms (mean response time) and 1000 ms from stimulus onset, ensuring that the observers engaged fully with the task. Training shifted the observers' criterion from 43.74° (±1.55°) for the 45° boundary to 31.4° (±1.66°) for the 30° boundary, and to 57.2° (±2.36°) for the 60° boundary (Figure 7A). Fitting the behavioral data with a cumulative Gaussian (Supplemental Data) showed a significant shift in the observers' criterion (i.e., 45° versus 30°, $F(1,6) = 16.1$, $p < 0.01$; 45° versus 60°, $F(1,4) = 20.44$, $p < 0.05$).

A searchlight analysis showed significant accuracies for the classification of radial versus concentric stimuli primarily in KO/LOS but not frontal areas (Figure S1B). We then trained an SVM to discriminate between the two extreme stimulus conditions (radial versus concentric) stimuli, tested the classifier's prediction (radial versus concentric) for each stimulus condition and generated fMR-metric functions as before. This analysis showed that fMR-metric functions in KO/LOS for the trained categories (30°, 60° boundary) were shifted significantly ($F(1,10) = 25.89$, $p < 0.001$) against each other (Figure 7B). In contrast, fMR-metric functions were not significantly fitted in the IFG/insula region. These results suggest that higher occipitotemporal regions contain information related to the learned stimulus categories even when observers are not engaged in a categorization task. However, frontal circuits show learning-dependent changes that reflect the observers' behavioral choice when observers are engaged in a categorization task. Thus, learning in higher occipitotemporal regions shapes the selective representation of perceived categories that are sustained independent of the task, whereas task-related changes in frontal areas reflect changes in the decision criterion that the observers use when comparing sensory evidence in the context of the categorization task.

DISCUSSION

Our findings provide insights into the learning brain processes that are important for making decisions under uncertainty. First, using advanced mutlivoxel pattern classification methods, we identify the neural decision processes that are shaped by category learning. We show that learning shapes neural processing related to selecting the appropriate criterion (i.e., categorical boundary) in frontal and higher occipitotemporal regions rather than signal detection or response execution in primary visual or motor areas. Second, we discern differential mechanisms of learning-dependent plasticity in frontal versus occipitotemporal areas. In particular, learning shapes selective readout signals in higher occipitotemporal regions that reflect the perceived

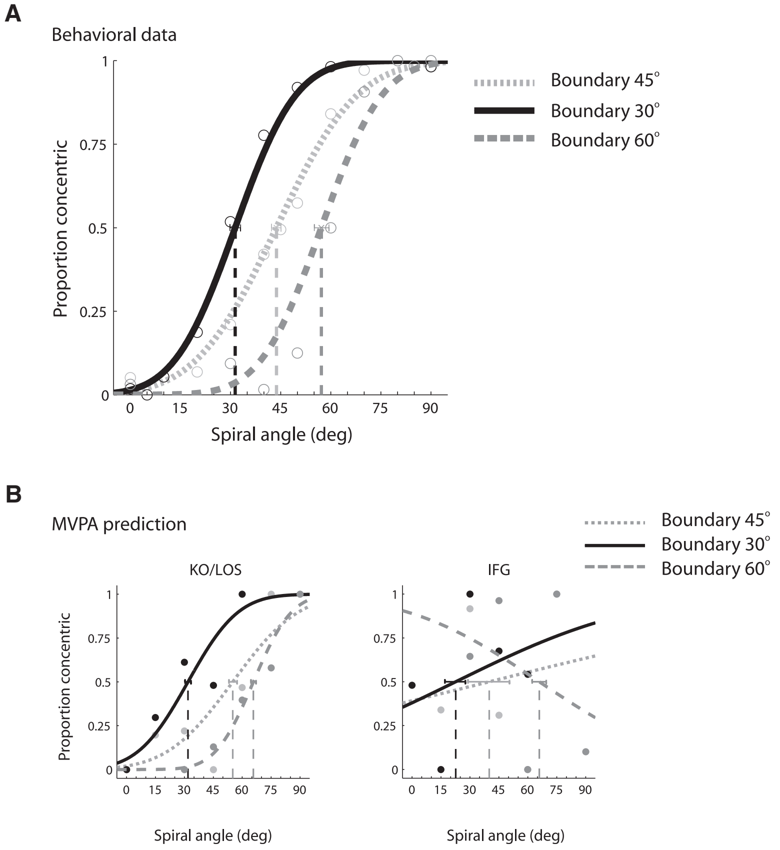

Figure 7. Control Experiment

(A) Average behavioral data across observers from the lab (circles) are shown for each boundary. Lines indicate the cumulative Gaussian fits. Error bars indicate the 95% confidence interval at 50% concentric threshold.

(B) fMR-metric curves for each task (45°, 30°, and 60° boundary) and ROI. Average classifier prediction data across observers are scaled from 0 to 1 and fitted with cumulative Gaussian functions. Error bars indicate the 95% confidence interval at 50% concentric threshold.

stimulus category independent of task context and may contribute to the estimation of the decision criterion specific to the task in prefrontal circuits. Thus, our findings provide novel evidence for distributed experience-dependent plasticity mechanisms that shape the estimation of decision variables in the human brain.

Our work advances our understanding of adaptive decision making mechanisms beyond previous physiology and imaging studies in three main respects. First, previous studies have provided a detailed roadmap of the cortical circuits involved in decision making. In particular, previous studies have investigated the neural circuits involved in categorical decisions under uncertainty related to stimulus detection (i.e., due to external noise applied to the stimulus) or criterion selection (i.e., distance from the categorical boundary). Dorsolateral prefrontal and intraparietal regions have been suggested to accumulate information about noisy signals in order to make a decision (Grinband et al., 2006; Heekeren et al., 2004, 2006; Kim and Shadlen, 1999; Newsome et al., 1989; Romo and Salinas, 2003; Shadlen and Newsome, 2001). Ventrolateral prefrontal regions including insular cortex have been suggested to monitor uncertainty in stimulus detection, discriminability, and probability of reward. In the context of categorical decisions (for reviews, Ashby and Maddox, 2005; Keri, 2003), these areas are suggested to maintain information in short term memory for comparative stimulus judgments (Philiastides and Sajda, 2007) relative to the category boundary (Grinband et al., 2006). Further, motor areas are thought to be involved in both computing the decision variables as well as the planning and execution of the response (Gold and Shadlen, 2003), while supplementary eye field (Corbetta and Shulman, 2002) and cingulate (Rushworth and Behrens, 2008) regions are suggested to engage in error-monitoring and performance adjustment through the allocation of attentional resources. Extending beyond this previous work, our study demonstrates that learning shapes neural processing in these circuits in accordance with changes in behavioral decisions after training.

Second, we investigate how learning shapes the link between behavioral and neural choices and supports adaptive decision making. Previous physiology studies have shown that sensory (Britten et al., 1996; Uka and DeAngelis, 2003) and prefrontal regions (e.g., Kim and Shadlen, 1999) reflect the animals' behavioral choice in perceptual decision tasks (e.g., motion direction discrimination). Previous imaging studies have successfully decoded decisions and intentions in the human brain (Hampton and O'Doherty, 2007; Haynes et al., 2007; Pessoa and Padmala, 2005, 2007; Philiastides et al., 2006; Philiastides and Sajda, 2006; Serences and Boynton, 2007; Williams et al., 2007). Here, we show that distributed signals across voxel patterns in ventral prefrontal (IFG/insula), and premotor (PMd) cortex contain information that allows us to decode the observers' behavioral choice as shaped by previous experience. We develop a methodology adapting established psychophysical and physiological procedures to fMRI data collection and multivariate analysis. In particular, we collect single-trial fMRI data for stimuli that vary parametrically in their physical similarity. We then compute psychometric functions based on the observers' behavioral choice on each stimulus trial and fMR-metric functions based on the classifier's choice on single-trial fMRI signals. Comparing the classifier's choices with the observer's choices shows that fMR-metric functions resemble psychometric functions and have similar thresholds, suggesting a link between behavioral and neural responses. More importantly, we show that learning-dependent changes in the behavioral decision criterion (i.e., categorical boundary) are reflected by changes in the threshold (50% point) of the fMR-metric functions in ventrolateral prefrontal (IFG/insula) and premotor regions. An analysis of the slopes of the psychometric and fMR-metric functions suggest that learning changes processes related to decision variables (criterion for stimulus comparison, stimulus uncertainty relative to criterion) rather than signal detection (i.e., sensitivity in detecting stimuli from noise). This is supported by the lack of a significant shift in the fMR-metric functions in V1, CS, or SEF suggesting that learning a new criterion for categorization does not affect the physical stimulus- or response-related representations in these regions that may support signal detection and motor execution respectively.

Third, in contrast to the prediction that learning-dependent changes in decision variables should only engage higher frontal circuits involved in reading out and comparing sensory evidence, we observed shifts in the fMR-metric functions in higher occipitotemporal regions (KO/LOS, LO) that are known to be involved in the representation of visual forms (Ostwald et al., 2008). Recent work has suggested that these areas accumulate information to the time of recognition (Ploran et al., 2007), support the persistence of a percept (Philiastides and Sajda, 2007), and therefore may contribute to the comparison of sensory evidence during decision making. Our findings demonstrate the learning shapes this comparison process and the neural representation of visual categories in occipitotemporal areas. Importantly, these neural changes were evident without observers performing the categorization task, suggesting learning-dependent changes in the representation of the perceived categories rather than simply task-related modulations. The lack of learning-dependent changes in prefrontal regions when observers did not perform the categorization task is in agreement with a recent imaging study (Jiang et al., 2007) and the proposed role of these areas in adaptive coding for complex cognitive tasks (Duncan, 2001; Koechlin and Summerfield, 2007; Miller, 2000). It is unlikely that the learning-induced changes we observed in occipitotemporal areas resulted from learning specific category exemplars or stimulus-response associations. In particular, the stimuli tested during scanning differed in their visual properties (i.e., signal level and spiral angle) from the stimuli presented during training, suggesting that observers performed the categorization task based on the learned boundary rather than specific exemplars. Finally, we controlled for the possibility that the results could be due to memorized stimulus-response associations by randomizing the motor responses based on the cue in the main experiment, and introducing a task requiring a motor response orthogonal to the stimulus categories in the control experiment.

These findings provide insights in the contested role of temporal areas in visual categorization. Previous studies have proposed that the temporal cortex represents primarily the visual similarity between stimuli and their identity (Freedman et al., 2003; Jiang et al., 2007; Op de Beeck et al., 2001) while other studies show that it represents the critical stimulus dimensions for categorization (Li et al., 2007; Mirabella et al., 2007; Sigala and Logothetis, 2002) and is modulated by task demands (Koida and Komatsu, 2007) as well as experience (e.g., Gauthier et al., 1997; Op de Beeck et al., 2006; Op de Beeck et al., 2008). Our findings provide evidence for flexible neural representations of visual stimuli that reflect the learned categorical similarity rather than the physical similarity between stimuli. It is possible that these flexible representations are formed based on recurrent processes that integrate bottom-up selectivity based on feature similarity in occipitotemporal areas and top-down influences based on the perceptual interpretation of the stimuli in prefrontal circuits. However, our findings suggest that such top-down influences during learning shape the neural representations in occipitotemporal areas to reflect the perceived categories even when observers are not engaged in a visual categorization task.

The limited resolution of fMRI does not allow us to discern whether these learning-dependent changes in the neural representations reflect changes in the selectivity of single neurons, correlations across local neural populations, or connectivity across frontoparietal areas engaged in computing decision variables and sensory-driven occipitotemporal regions. Recent neurophysiological studies show that small neural populations in the temporal cortex contain information about abstract categories that are established through training (Meyers et al., 2008). Further, learning is shown to establish and refine sensory-motor associations (Boettiger and D'Esposito, 2005; Toni et al., 2001) that provide a more selective readout of highly sensitive signals in sensory areas (Law and Gold, 2008). Our findings provide evidence that once such sensory-motor associations have been established through training on a visual categorization task, neural representations in occipitotemporal areas remain sensitive to the learned stimulus categories even when the observers are not engaged in a complex cognitive task that entails selective information readout by higher frontoparietal circuits. Such representations may further support identification of individual category members, generalization to new stimuli similar to the category members, and expertise in familiar abstract categories.

In summary, our findings provide evidence that category learning shapes decision processes in the human brain related to the choice of the behaviorally relevant criterion for assigning sensory input into meaningful categories. Using multivoxel pattern analysis on single-trial fMRI data, we compare fMR-metric functions that reflect the choices of an MVPA classifier to psychometric functions that reflect the observers' choices. MVPA allows us to evaluate whether small biases across voxels related to the stimulus preference of the underlying neural populations are statistically reliable. However, cautious interpretation of the results is necessary due to the complex nature of the BOLD signal. MVPA on fMRI signals allows us to reliably extract information about the sensitivity of neural populations at a finer spatial resolution than that of typical fMRI measurements by pooling small biases across voxels, but it does not enable us to discern the nature of the signals that determine this sensitivity. In the context of our study, learning-dependent changes in the classifier's choice may reflect changes in the selectivity of single neurons, correlations across local neural populations, or input from local or distant neural circuits. Our findings suggest that learning may shape selective readout signals from neurons in higher occipitotemporal areas that support adaptive coding in frontal neural populations for complex cognitive tasks. Future work employing the methodology that we employed here for the analysis of both fMRI and electrophysiology signals will provide further insights into the neural mechanisms that mediate learning-dependent changes for adaptive decision making in humans and monkeys.

EXPERIMENTAL PROCEDURES

Observers

Eight observers (four male, four female) participated in the main experiment, and fourteen observers (seven male, seven female) participated in the control experiment. Two observers were excluded from the data analysis in the control experiment due to excessive head movement during scanning. All observers were from the University of Birmingham, had normal or corrected to normal vision, and gave written informed consent. The study was approved by the local ethics committee.

Stimuli

We used Glass pattern stimuli (Glass, 1969) defined by white dot pairs (dipoles) displayed within a square aperture (7.7° × 7.7°) on a black background (100% contrast). The dot density was 3% and the Glass shift (i.e., the distance between two dots in a dipole) was 16.2 arc min. The size of each dot was 2.3 × 2.3 arc min^2. These parameters were chosen based on pilot psychophysical studies and in accordance with previous studies (e.g., Wilson and Wilkinson, 1998) showing that coherent form patterns are reliably perceived for these parameters. We generated concentric and radial Glass patterns by placing dipoles tangentially (concentric stimuli) or orthogonally (radial stimuli) to the circumference of a circle centered on the fixation dot. Further, we generated intermediate patterns between these two Glass pattern types by parametrically varying the spiral angle of the pattern from 0° (radial pattern) to 90° (concentric pattern) (Figure 1A). For each dot dipole, the spiral angle was defined as the angle between the dot dipole orientation and the radius from the center of the dipole to the center of the stimulus aperture. Each stimulus comprised dot dipoles that were aligned according to the specified spiral angle (signal dipoles) for a given stimulus, and noise dipoles for which the spiral angle was randomly selected. Half of the observers were presented with clockwise spiral patterns (0° to 90° spiral angle) and half with counterclockwise spiral patterns (0° to −90° spiral angle). A new pattern was generated for each stimulus presented in a trial, resulting in stimuli that were locally jittered in their position.

To control for stimulus-specific training effects and ensure generalization of learning, we trained the observers with 60% signal stimuli but tested (pre- and posttraining test) and scanned on 40% signal stimuli. These values were selected based on behavioral pilot experiments showing that degrading the Glass patterns with noise resulted in gradual changes in the observers' categorization performance across conditions. This procedure ensured that learning could not be due to similar local cues between the stimuli used for training, tests and scanning, but rather global features (i.e., spiral angle) used by the observers for stimulus categorization.

Design

For the main experiment (see Supplemental Data for details on control experiment), all observers participated in two fMRI sessions. Each session was preceded by psychophysical training outside the scanner, and the observers' behavioral performance was matched before the two fMRI sessions (85% correct performance).

Psychophysical Training

First, observers were familiarized with the task and stimuli in a short practice session. Observers were shown the 100% signal Glass patterns and were instructed to categorize each stimulus into one of two categories: similar to a radial Glass pattern (0° spiral angle) versus similar to a concentric Glass pattern (90° spiral angle). Then, during the pretraining test observers were presented with 40% signal Glass patterns and were instructed to perform the same categorization task. This pretraining test allowed us to identify the categorical boundary between radial and concentric for each observer before training.

Following the pretraining test, observers were presented with 60% signal stimuli and were trained (self-paced procedure with audio error feedback) to shift this boundary to either 30° (training runs, 4.1 ± 2.36) or 60° (training runs, 3.8 ± 1.67) of spiral angle. In the first training session, half of the observers (group 1) were trained to categorize the stimuli based on a boundary at 30° spiral angle, whereas the other half (group 2) were trained to categorize the stimuli based on 60° boundary. In the second training session, observers from group 1 were trained on the 60° boundary, while observers from group 2 were trained on the 30° boundary. For the 30° boundary session observers were trained at steps: 5°, 10°, 20°, 25°, 35°, 40°, 50°, and 55° of spiral angle while for the 60° boundary at the following steps: 35°, 40°, 50°, 55°, 65°, 70°, 80°, and 85° of spiral angle. Each training session comprised multiple runs (ranging from 2 to 9 runs) with 144 trials per run. For each trial during training, the stimulus was presented for 300 ms. A white fixation square (7.7 × 7.7 arc min^2) was presented at the center of each stimulus. Observers were instructed to indicate which category the stimulus belonged to by pressing one of two keys. Observers were trained until their performance reached a stable criterion level (85% correct) twice. This training procedure ensured that the performance of the observers was similar for both boundaries before scanning.

After training, observers were tested in a posttraining test (420 trials) during which stimuli (40% signal level) were presented for 200 ms. For the 30° boundary session observers were tested at steps 0°, 5°, 10°, 15°, 20°, 25°, 30°, 35°, 40°, 45°, 50°, 55°, 60°, and 90° of spiral angle, while for the 60° boundary at steps 0°, 30°, 35°, 40°, 45°, 50°, 55°, 60°, 65°, 70°, 75°, 80°, 85°, and 90° of spiral angle. To assess the result of training, no feedback was given during this posttraining test.

fMRI Measurements

For the main experiment (see Supplemental Data for control experiment), all observers participated in two scanning sessions during which they performed the categorization task on the Glass pattern stimuli after training on each of the two boundaries (30° and 60° spiral angles).

For each observer, we collected data from seven to eight event-related runs in each session. The order of trials was matched for history (one trial back) such that each trial was equally likely to be preceded by any of the conditions. The order of the trials differed across runs and observers. Eight conditions (seven stimulus conditions and one fixation condition during which only the fixation point was displayed at the center of the screen) with 16 trials per condition were presented in each run. Each run comprised 129 trials (128 trials across conditions and one initial trial for balancing the history of the second trial) and two 9 s fixation periods (one in the beginning and one at the end of the run). For the 30° boundary, the stimulus conditions comprised Glass patterns of 0°, 15°, 25°, 30°, 35°, 60°, and 90° spiral angles. For the 60° boundary, the stimulus conditions comprised Glass patterns of 0°, 30°, 55°, 60°, 65°, 75°, and 90° spiral angles. Four stimulus conditions (0°, 30°, 60°, and 90° spiral angle) were common between sessions. The choice of the rest of the stimuli was constrained by two factors. (1) We equated the number of conditions and stimuli across categories while avoiding stimulus repetition to ensure that observers were not biased in their responses due to uneven number of conditions (stimuli) in one of the two categories. (2) We aimed to sample representative points on the psychometric function while selecting a limited but adequate number of conditions to ensure that enough trials were recorded per condition, and high-quality signals were measured within the time constraints of fMRI scanning.

For fixation trials, the fixation square was displayed for 3 s. For experimental trials (3 s long), each trial started with 200 ms stimulus presentation followed by 1300 ms delay during which a white fixation square was displayed at the center of the screen. After this fixed delay, the fixation dot changed color to either green or red. This change in fixation color served as a cue for the motor response using one of two buttons. If the color cue was green observers used the same finger-key matching as during training (i.e., index finger for radial), while if the color cue was red, observers switched finger-key matching (i.e., index finger for concentric). The fixation color was changed back to white 300 ms before the next trial onset. This procedure aimed to dissociate the motor response (button press) from the learned stimulus categories. Observers were familiarized with this procedure before scanning.

fMRI Data Acquisition

The experiments were conducted at the Birmingham University Imaging Center (3T Achieva scanner; Philips, Eindhoven, The Netherlands). EPI and T1-weighted anatomical (1 × 1 × 1 mm) data was collected with an eight channel SENSE head coil. For the main experiment, EPI data (Gradient echo-pulse sequences) were acquired from 24 slices (whole-brain coverage, TR, 1500 ms; TE, 35 ms;

flip-angle, 73 degrees; 2.5 × 2.5 × 4 mm resolution). For the control experiment EPI data were acquired from 32 slices (whole-brain coverage, TR, 2000 ms; TE, 35 ms; flip-angle, 80 degrees; 2.5 × 2.5 × 3 mm resolution).

fMRI Data Analysis

MRI data was processed using Brain Voyager QX (Brain Innovations, Maastricht, The Netherlands). Anatomical data was used for 3D cortex reconstruction, inflation and flattening. Preprocessing of functional data included slice-scan time correction, head movement correction, temporal high-pass filtering (3 cycles) and removal of linear trends. Trials with head motion larger than 1 mm of translation, or 1° of rotation were excluded from the analysis. Spatial smoothing (Gaussian filter; full-width at half maximum, 6 mm) was performed only for group random effect analysis (Figure 3A) but not for data used for the multivoxel pattern classification 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 the two sessions were coaligned registering all volumes of each observer to the first functional volume of the first run and session. This procedure ensured a cautious registration across sessions. To avoid confounds from any remaining registration errors we compared fMRI signals between stimulus conditions within each session rather than across sessions. A gray-matter mask was generated for each observer in Talairach space from the anatomical data for selecting only gray-matter voxels for further analyses.

Multivariate Mapping of Regions Based on the Observer's Response

For each observer, we identified cortical regions whose activations correlated with the observers' response in the categorization task by performing a multivoxel searchlight analysis (Kriegeskorte et al., 2006) on the data for each categorization boundary. In particular, we defined a spherical aperture with radius of 9 mm and moved this aperture voxel by voxel across the whole brain (only gray-matter voxels were included). For voxels within the aperture (98 voxels per aperture on average), we used a linear support vector machine pattern classifier to classify fMRI signals across all stimulus conditions based on the observers' behavioral choice (radial versus concentric). That is, we trained the classifier to associate the fMRI signal from each trial with a label (radial versus concentric) that was determined by the observer's interpretation of the stimulus in the context of the categorization task rather than by the stimulus condition. To control for the unequal numbers of trials categorized by the observers as radial versus concentric in each stimulus condition, we weighed the classification by the ratio of concentric over radial fMRI patterns (Supplemental Data).

We performed this pattern classification on individual trials by averaging the two volumes from each trial (trial duration = 3 s, TR = 1.5 s) to generate one training pattern. Modeling of the hemodynamic response function (HRF) with a double Gaussian function for all data across conditions and observers showed that the average peak response across areas was at 4.71 s (±0.37 s) after stimulus onset. To account for this hemodynamic delay, we shifted the fMRI time series by 3 volumes (4.5 s). To ensure generalization of the classification, we used a leave-one-run-out cross-validation procedure. For each cross-validation, one run was left out as an independent test dataset and the data from the rest of the runs was used as the training set. The classification accuracy for each aperture was obtained by averaging the prediction accuracy across cross-validations. The accuracy value for each voxel was obtained by averaging the accuracy values from all apertures in which this voxel was included. To identify voxels with accuracy significantly higher than chance across observers we conducted a second level statistical analysis (t test, $p < 0.0001$, cluster threshold estimation 5 mm^2).

SUPPLEMENTAL DATA

The Supplemental Data include four figures, one table, and Supplemental Experimental Procedures and can be found with this article online at http://www.neuron.org/supplemental/S0896-6273(09)00239-6.

ACKNOWLEDGMENTS

We thank J. Storrar for help with data collection. We are grateful to N. Kriegeskorte for kindly providing code for data analysis. We would also like to thank B. Tjan, M. Eckstein, and A.E. Welchman for helpful comments and discussions. This work was supported by grants from the Biotechnology and Biological Sciences Research Council to Z.K. (D52199X, E027436).

Accepted: March 19, 2009
Published: May 13, 2009

REFERENCES

Ashby, F.G., and Maddox, W.T. (2005). Human category learning. Annu. Rev. Psychol. *56*, 149–178.

Boettiger, C.A., and D'Esposito, M. (2005). Frontal networks for learning and executing arbitrary stimulus-response associations. J. Neurosci. *25*, 2723–2732.

Britten, K.H., Newsome, W.T., Shadlen, M.N., Celebrini, S., and Movshon, J.A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis. Neurosci. *13*, 87–100.

Corbetta, M., and Shulman, G.L. (2002). Control of goal-directed and stimulus-driven attention in the brain. Nature Rev. *3*, 201–215.

Cox, D.D., and Savoy, R.L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage *19*, 261–270.

Daw, N.D., and Doya, K. (2006). The computational neurobiology of learning and reward. Curr. Opin. Neurobiol. *16*, 199–204.

Dayan, P., and Niv, Y. (2008). Reinforcement learning: the good, the bad and the ugly. Curr. Opin. Neurobiol. *18*, 185–196.

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. Nat. Rev. Neurosci. *2*, 820–829.

Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J. Neurosci. *23*, 5235–5246.

Gauthier, I., Anderson, A.W., Tarr, M.J., Skudlarski, P., and Gore, J.C. (1997). Levels of categorization in visual recognition studied using functional magnetic resonance imaging. Curr. Biol. *7*, 645–651.

Glass, L. (1969). Moire effect from random dots. Nature *223*, 578–580.

Glimcher, P.W. (2003). The neurobiology of visual-saccadic decision making. Annu. Rev. Neurosci. *26*, 133–179.

Gold, J.I., and Shadlen, M.N. (2003). The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. J. Neurosci. *23*, 632–651.

Gold, J.I., and Shadlen, M.N. (2007). The neural basis of decision making. Annu. Rev. Neurosci. *30*, 535–574.

Grinband, J., Hirsch, J., and Ferrera, V.P. (2006). A neural representation of categorization uncertainty in the human brain. Neuron *49*, 757–763.

Hampton, A.N., and O'Doherty, J.P. (2007). Decoding the neural substrates of reward-related decision making with functional MRI. Proc. Natl. Acad. Sci. USA *104*, 1377–1382.

Haynes, J.D., and Rees, G. (2006). Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. *7*, 523–534.

Haynes, J.D., Sakai, K., Rees, G., Gilbert, S., Frith, C., and Passingham, R.E. (2007). Reading hidden intentions in the human brain. Curr. Biol. *17*, 323–328.

Heekeren, H.R., Marrett, S., Bandettini, P.A., and Ungerleider, L.G. (2004). A general mechanism for perceptual decision-making in the human brain. Nature *431*, 859–862.

Heekeren, H.R., Marrett, S., Ruff, D.A., Bandettini, P.A., and Ungerleider, L.G. (2006). Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality. Proc. Natl. Acad. Sci. USA *103*, 10023–10028.

Heekeren, H.R., Marrett, S., and Ungerleider, L.G. (2008). The neural systems that mediate human perceptual decision making. Nat. Rev. Neurosci. *9*, 467–479.

Jiang, X., Bradley, E., Rini, R.A., Zeffiro, T., Vanmeter, J., and Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. Neuron *53*, 891–903.

Keri, S. (2003). The cognitive neuroscience of category learning. Brain Res. Brain Res. Rev. *43*, 85–109.

Kim, J.N., and Shadlen, M.N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat. Neurosci. *2*, 176–185.

Koechlin, E., and Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. Trends Cogn. Sci. *11*, 229–235.

Koida, K., and Komatsu, H. (2007). Effects of task demands on the responses of color-selective neurons in the inferior temporal cortex. Nat. Neurosci. *10*, 108–116.

Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. Proc. Natl. Acad. Sci. USA *103*, 3863–3868.

Law, C.T., and Gold, J.I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. Nat. Neurosci. *11*, 505–513.

Li, S., Ostwald, D., Giese, M., and Kourtzi, Z. (2007). Flexible coding for categorical decisions in the human brain. J. Neurosci. *27*, 12321–12330.

Meyers, E.M., Freedman, D.J., Kreiman, G., Miller, E.K., and Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. J. Neurophysiol. *100*, 1407–1419.

Miller, E.K. (2000). The prefrontal cortex and cognitive control. Nat. Rev. Neurosci. *1*, 59–65.

Mirabella, G., Bertini, G., Samengo, I., Kilavik, B.E., Frilli, D., Della Libera, C., and Chelazzi, L. (2007). Neurons in area V4 of the macaque translate attended visual features into behaviorally relevant categories. Neuron *54*, 303–318.

Newsome, W.T., Britten, K.H., and Movshon, J.A. (1989). Neuronal correlates of a perceptual decision. Nature *341*, 52–54.

Norman, K.A., Polyn, S.M., Detre, G.J., and Haxby, J.V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. *10*, 424–430.

O'Doherty, J.P., Hampton, A., and Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. Ann. N Y Acad. Sci. *1104*, 35–53.

Op de Beeck, H., Wagemans, J., and Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. Nat. Neurosci. *4*, 1244–1252.

Op de Beeck, H.P., Baker, C.I., DiCarlo, J.J., and Kanwisher, N.G. (2006). Discrimination training alters object representations in human extrastriate cortex. J. Neurosci. *26*, 13025–13036.

Op de Beeck, H.P., Torfs, K., and Wagemans, J. (2008). Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. J. Neurosci. *28*, 10111–10123.

Ostwald, D., Lam, J.M., Li, S., and Kourtzi, Z. (2008). Neural coding of global form in the human visual cortex. J. Neurophysiol. *99*, 2456–2469.

Pessoa, L., and Padmala, S. (2005). Quantitative prediction of perceptual decisions during near-threshold fear detection. Proc. Natl. Acad. Sci. USA *102*, 5612–5617.

Pessoa, L., and Padmala, S. (2007). Decoding near-threshold perception of fear from distributed single-trial brain activation. Cereb. Cortex *17*, 691–701.

Philiastides, M.G., and Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb. Cortex *16*, 509–518.

Philiastides, M.G., and Sajda, P. (2007). EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. J. Neurosci. *27*, 13082–13091.

Philiastides, M.G., Ratcliff, R., and Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. J. Neurosci. *26*, 8965–8975.

Platt, M.L., and Huettel, S.A. (2008). Risky business: the neuroeconomics of decision making under uncertainty. Nat. Neurosci. *11*, 398–403.

Ploran, E.J., Nelson, S.M., Velanova, K., Donaldson, D.I., Petersen, S.E., and Wheeler, M.E. (2007). Evidence accumulation and the moment of recognition: dissociating perceptual recognition processes using fMRI. J. Neurosci. *27*, 11912–11924.

Romo, R., and Salinas, E. (2003). Flutter discrimination: neural codes, perception, memory and decision making. Nat. Rev. Neurosci. *4*, 203–218.

Rushworth, M.F., and Behrens, T.E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. Nat. Neurosci. *11*, 389–397.

Schall, J.D. (2001). Neural basis of deciding, choosing and acting. Nat. Rev. Neurosci. *2*, 33–42.

Schultz, W., Preuschoff, K., Camerer, C., Hsu, M., Fiorillo, C.D., Tobler, P.N., and Bossaerts, P. (2008). Explicit neural signals reflecting reward uncertainty. Philos. Trans. R. Soc. Lond. B Biol. Sci. *363*, 3801–3811.

Serences, J.T., and Boynton, G.M. (2007). The representation of behavioral choice for motion in human visual cortex. J. Neurosci. *27*, 12893–12899.

Shadlen, M.N., and Newsome, W.T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J. Neurophysiol. *86*, 1916–1936.

Sigala, N., and Logothetis, N.K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. Nature *415*, 318–320.

Smith, P.L., and Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. Trends Neurosci. *27*, 161–168.

Sutton, R., and Barto, A.G. (1998). Reinforcement Learning (Cambridge, MA: MIT Press).

Toni, I., Rushworth, M.F., and Passingham, R.E. (2001). Neural correlates of visuomotor associations. Spatial rules compared with arbitrary rules. Exp. Brain Res. *141*, 359–369.

Uka, T., and DeAngelis, G.C. (2003). Contribution of middle temporal area to coarse depth discrimination: comparison of neuronal and psychophysical sensitivity. J. Neurosci. *23*, 3515–3530.

Williams, M.A., Dang, S., and Kanwisher, N.G. (2007). Only some spatial patterns of fMRI response are read out in task performance. Nat. Neurosci. *10*, 685–686.

Wilson, H.R., and Wilkinson, F. (1998). Detection of global structure in Glass patterns: implications for form vision. Vision Res. *38*, 2933–2947.

Yu, A.J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. Neuron *46*, 681–692.