Current Biology 23, 1799–1804, September 23, 2013 ©2013 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2013.07.052

Report

Learning Optimizes Decision Templates in the Human Visual Cortex

Shu-Guang Kuai,[1] Dennis Levi,[2] and Zoe Kourtzi[1,3,*]
[1]School of Psychology, University of Birmingham, Birmingham B15 2TT, UK
[2]School of Optometry and Helen Wills Neuroscience Institute, University of California, Berkeley, Berkley, CA 94720, USA
[3]Laboratory for Neuro- and Psychophysiology, KU Leuven, 3000 Leuven, Belgium

Summary

Translating sensory information into perceptual decisions is a core challenge faced by the brain. This ability is understood to rely on weighting sensory evidence in order to form mental templates of the critical differences between objects. Learning is shown to optimize these templates for efficient task performance [1–4], but the neural mechanisms underlying this improvement remain unknown. Here, we identify the mechanisms that the brain uses to implement templates for perceptual decisions through experience. We trained observers to discriminate visual forms that were randomly perturbed by noise. To characterize the internal stimulus template that observers learn when performing this task, we adopted a classification image approach (e.g., [5–7]) for the analysis of both behavioral and fMRI data. By reverse correlating behavioral and multivoxel pattern responses with noisy stimulus trials, we identified the critical image parts that determine the observers' choice. Observers learned to integrate information across locations and weight the discriminative image parts. Training enhanced shape processing in the lateral occipital area, which was shown to reflect size-invariant representations of informative image parts. Our findings demonstrate that learning optimizes mental templates for perceptual decisions by tuning the representation of informative image parts in higher ventral cortex.

Introduction

Extracting Classification Images from Behavioral and fMRI Data

We tested the ability of observers (n = 9) to discriminate between two classes of polygons (Figure 1A, class I or class II). Although these stimuli are simpler than familiar objects, they are advantageous in several respects. First, to investigate learning, we chose a novel, rather than a familiar, stimulus space. Second, to ensure that observers discriminated global shapes rather than local differences between stimuli, we parametrically manipulated the stimuli with linear morphing and rotated them in the image plane across trials. Third, we used positional noise, which has been shown to support the efficient extraction of classification images from smaller samples than needed when luminance noise is used [1]; therefore, this is an ideal method for extracting classification images from a limited number of fMRI trials.

To identify the specific stimulus components that determine the observer's choice (i.e., the discriminative features), we reverse correlated behavioral choices and fMRI signals with noisy stimulus trials. This approach has been used widely in psychophysics (for reviews, see [5, 6]); however, its application to neuroimaging has been limited by noisy single-trial fMRI signals and the small number of samples that can be acquired during fMRI scans [8, 9]. To overcome these limitations, we developed a new method that uses reverse correlation in conjunction with multivoxel pattern analysis. We calculated decision templates on the basis of the choices made by a linear support vector machine (SVM) classifier that decodes the stimulus class from the fMRI data measured on individual stimulus trials. Thus, we combined the power of SVM stimulus decoding to uncover neuronal preferences [10] with reverse correlation classification images in order to reveal discriminative image features that are enhanced through learning.

To directly test the link between human behavior and fMRI data, we compared human performance and behaviorally relevant fMRI responses to an ideal observer. To extract behaviorally relevant fMRI signals, we trained a linear classifier to predict the observers' choice. After this, we regressed both behavioral responses and behaviorally relevant fMRI signals to the input stimuli and computed the classification images. To avoid circularity and evaluate whether behavioral performance and neural representations become more efficient with learning, we correlated behavioral and fMRI classification images to an ideal observer rather than to each other.

Behavioral and fMRI Classification Images

To ensure that observers learned to classify the two polygon classes, we trained them with auditory feedback (minimum three sessions, 900–1,100 trials per session), resulting in improved performance, as quantified by a 32.2% reduction in class discrimination thresholds ($F_{(1,8)} = 58.44$, $p < 0.01$). Importantly, this improvement was reflected in the participants' use of particular portions of the image when making their decisions. Classification images based on the observers' performance after training showed marked differences between image parts associated with the two stimulus classes (Figure 2A). In contrast, we did not observe any consistent image parts associated with the two stimulus classes before training, ensuring that the classification images reflected the perceived differences between classes rather than local image differences between stimuli.

Having characterized the behavioral decision template, we used fMRI to determine where in the visual cortex this template is implemented. Given the known role of the ventral visual pathway in shape processing, we chose to study this pathway in detail with the use of high-resolution fMRI recordings. Our results show that classification images in the lateral occipital area (LO), but not in early visual areas, revealed image parts that were perceptually distinct between the two stimulus classes (Figure 2B, Figure S1 available online). Importantly, there was little information about this perceptual template before training, ensuring that fMRI classification templates reflect the perceived classes rather than stimulus examples. Comparing classification images derived from behavioral and fMRI data showed that similar image parts became more discriminable between the two stimulus classes after training, suggesting a correspondence between behavioral and fMRI templates.

*Correspondence: z.kourtzi@bham.ac.uk

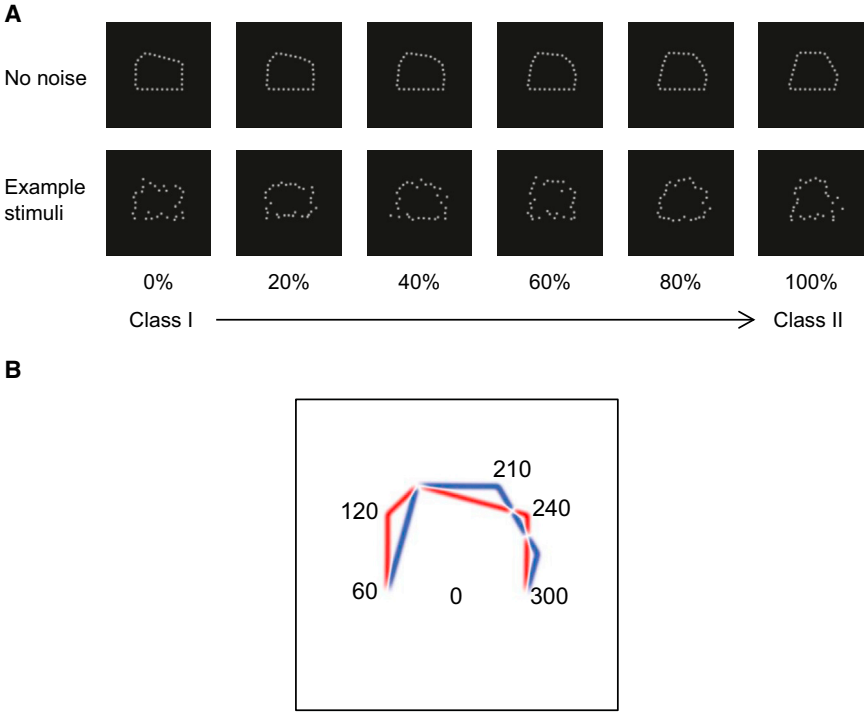

Figure 1. Stimuli and Image-Based Description

(A) Sample pentagon-like stimuli comprising 30 equally spaced Gaussian dots with SD = 0.1°. Two classes of shapes were generated by varying the location of the pentagon lines that differed in their length. The top panel shows the stimulus space generated by linear morphing between class I and II polygons (stimuli are shown as a function of the percent of class II). The bottom panel shows example stimuli with position noise, as presented in the experiment.

(B) Radial image decomposition. The class distance for each of 30 image parts (i.e., 12° large image regions centered on the position of each of the 30 Gaussian dots defining the shape contour) was computed by subtracting the mean distance of class I (red lines) from the mean distance of class II stimuli (blue lines). Positive values indicate larger distances for class I stimuli, whereas negative values indicate larger distances for class II stimuli. We identified three most informative image patches (120°, 210°, and 240°) with the highest class distances. White stimulus regions signify no differences between stimulus classes.

Quantifying Behavioral and fMRI Decision Template

To quantify behavioral and fMRI decision templates, we compared human and SVM classifier performance to an ideal observer model (i.e., the maximum performance that was possible for this stimulus discrimination task). We used a radial decomposition of the shapes (Figure 1B) and computed the distance between classes for each image location on the basis of the behavioral and fMRI data.

To assess whether decision templates became closer to the ideal after training, we correlated behavioral and fMRI class distances to ideal class distances for each participant before and after training. This analysis showed that human performance (Figure 3A) and fMRI activation patterns (Figure 3B) in higher ventral areas became closer to ideal performance after the observers had learned the stimulus classes. Importantly, we did not observe significant correlations before training, ensuring that our fMRI activation patterns reflect perceptual templates (i.e., representations of discriminative image parts between classes) rather than differential neural selectivity to local image features before observers became familiar with the stimulus classes. That is, linear regression analysis on the behavioral and ideal class distances showed that both the correlation ($t_{(8)} = 8.39$, $p < 0.01$) and the slope ($t_{(8)} = 8.86$, $p < 0.01$) values were enhanced significantly after training (Figure 3A).

Furthermore, correlation ($t_{(8)} = 3.59$, $p < 0.001$) and slope ($t_{(8)} = 3.68$, $p < 0.001$) values between fMRI and ideal class distances increased significantly after training in LO, suggesting that learning enhanced the representation of the perceived differences between classes (Figure 3B). In contrast, we did not observe any significant changes after training in early (correlation: $F_{(1,8)} = 3.43$, $p = 0.1$; slope: $F_{(1,8)} = 4.01$, $p = 0.08$) or ventral (correlation: $F_{(1,8)} = 1.49$, $p = 0.26$; slope: $F_{(1,8)} = 1.69$, $p = 0.23$) visual areas. This result in early visual areas was expected, given that the stimulus manipulations we employed (i.e., stimulus rotation across trials) prevented the learning of local image positions.

Is it possible that the learning-dependent improvement we observed in fMRI classification images in accordance with behavior was due to the fact that the classifier was trained on behaviorally relevant fMRI signals? To control for this, we trained the classifier on the choices of the ideal observer that contained all information about the stimulus space. This analysis (Figure S2A) resulted in similar correlation patterns, as shown in Figure 3B, suggesting that our results could not be confounded by classifier choice. This link between behavioral responses and fMRI activation patterns in higher ventral areas was further supported by enhanced trial-by-trial correlation after training between observer choices and classifier predictions (Figure S2B). Additional control analyses (for details, see the Supplemental Information) showed that our results could not be confounded by univariate signal differences across brain areas, motor responses, or eye movements.

Altogether, these results suggest that perceptual templates of shapes are implemented in higher ventral cortex. As well as being robust to local image rotations across trials, we found that these templates were also tolerant to stimulus size changes. In particular, after training, we tested observers' performance for stimuli that were presented at different sizes (1.5× or 2× larger) from the trained stimuli. We showed that behavioral and fMRI classification images for stimuli of trained and untrained size were highly similar. Specifically, there were no significant differences (R: $t_{(6)} = -0.07$, $p = 0.95$; slope: $t_{(6)} = -0.42$, $p = 0.69$) between the correlations of behavioral and ideal class distances for stimuli of trained and untrained size. This was also true for the correlation ($F_{(1,6)} < 1$, $p = 0.59$) and slope ($F_{(1,6)} < 1$, $p = 0.48$) values of fMRI and ideal class distances between stimuli of trained and untrained size in ventral visual areas. Although transfer of learning across image changes is highly debated [11, 12], our findings provide evidence that learning tunes representations of discriminative image parts in higher ventral cortex that are tolerant to image changes rather than specific local image positions.

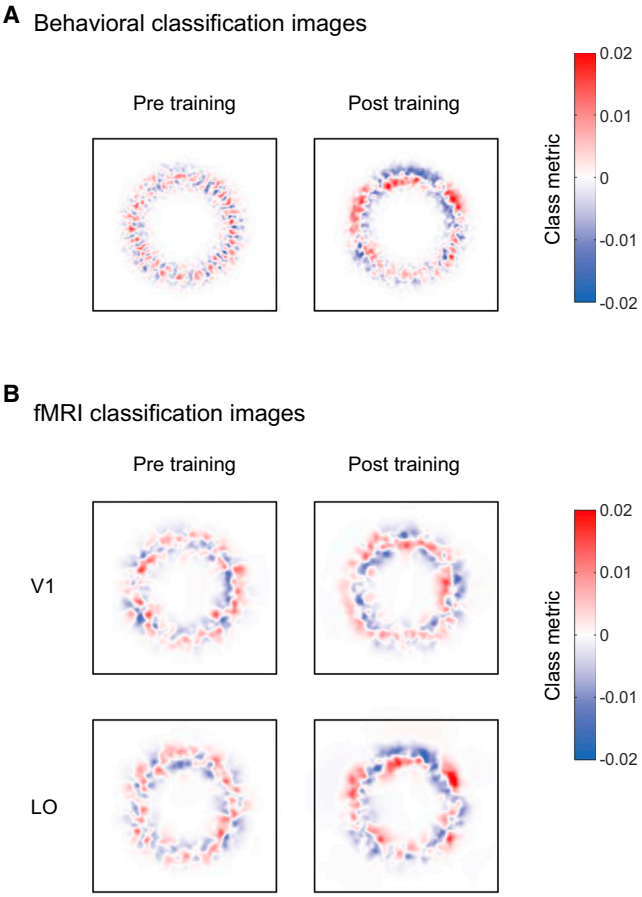

Figure 2. Classification Images

(A) Behavioral classification images before and after training averaged across participants.

(B) fMRI classification images for V1 and LO before and after training averaged across participants (see Figure S1 for all areas). Red indicates image locations associated with a class I decision, whereas blue indicates image locations associated with a class II decision.

Identifying Discriminative Image Parts

The optimal strategy for discriminating between the two stimulus classes is to take into account all image locations. Our results so far show that, after training, observers adopted this strategy using all informative image locations. To quantify this effect, we selected the three most informative image parts that corresponded to shape corners differing between the two stimulus classes (Figure 1B); that is, local maxima and minima from the stimulus distance metric description with the highest distance between classes.

Our results showed increased behavioral class distances after training ($F_{(1,8)}$ = 151.43, p = 0.001) for these informative image parts (Figure 4A), suggesting that the training enhanced the observers' ability to integrate information across discriminative parts. It is unlikely that spatial correlations across image locations could drive observer performance, given that noise samples were independently assigned to each stimulus location. However, to control for this possibility, we tested the performance of the ideal observer when information was provided only for each of the three informative image parts separately (i.e., the stimuli remained the same but different weights were applied across image locations). We found that no other image parts could be recovered reliably in the classification images when information was provided about one image part only (Figure S3). This analysis provides evidence that image locations are independent from each other, suggesting that the optimal strategy is to integrate information across them.

Analysis of the fMRI data (Figure 4B) showed that representations of these informative image parts are enhanced after training in LO rather than earlier visual areas. In particular, class distances increased significantly for informative image parts after training in LO ($F_{(1,8)}$ = 9.67, p < 0.01) but not earlier visual areas (V1: $F_{(1,8)}$ = 1.83, p = 0.21; V2: $F_{(1,8)}$ = 0.22, p = 0.65; V3v: $F_{(1,8)}$ = 3.3, p = 0.11; hV4: $F_{(1,8)}$ = 0.3, p = 0.6). Interestingly, the results based on behavioral and LO class distances indicate that observers may weight part three more than parts one and two, suggesting a stronger effect of training for weaker discriminative signals between stimulus classes.

Is it possible that the learning-dependent changes we observed in LO were simply due to higher classification accuracies in ventral areas after, rather than before, training? To control for this possibility, we conducted two additional analyses. First, we randomly selected 60% of the fMRI trials (given that mean classifier performance was 61% and ranged from 51% to 71% in LO), assigned correct labels to only these trials, and trained the classifier (using all correct and incorrect trials) to predict the stimulus class using an independent data set. This analysis showed no changes in the classification images in LO with training. Second, performing the class distance analysis on data from participants with classification accuracy higher than chance did not show any significant differences before and after training in early and ventral visual areas (see the Supplemental Information). Altogether, these analyses suggest that our findings could not be simply accounted for by differences in overall classification accuracies before and after training.

Discussion

Combining classification image approaches with multivariate fMRI analysis, we provide evidence for the mechanisms that the brain uses to optimize mental templates for perceptual decisions through experience. We demonstrate that higher ventral areas (LO) implement decision templates by integrating information across image locations and representing informative image parts in a size-invariant manner.

These findings advance our understanding of the brain mechanisms that optimize the neural code for efficient perceptual decisions in three main respects. First, previous behavioral studies have proposed that learning enhances perceptual efficiency by retuning the decision template [1–4]. Although fMRI analysis methods have successfully demonstrated changes in the overall activation magnitude (i.e., increased or decreased activations for trained stimuli) with learning (e.g., [13–17]), they have been less sensitive in distinguishing preferences of neural populations for distinct visual shapes. As a result, the link between enhanced perceptual efficiency due to learning and selectivity changes in brain patterns that support perceptual decisions remains unexplored.

Previous imaging studies (for review, see [18]) have speculated that decreased fMRI activations following training may be due to the enhanced tuning of small neural populations that encode behaviorally relevant information, resulting in enhanced performance. Only recently, with the use of multivoxel pattern classification methods, have we been able to

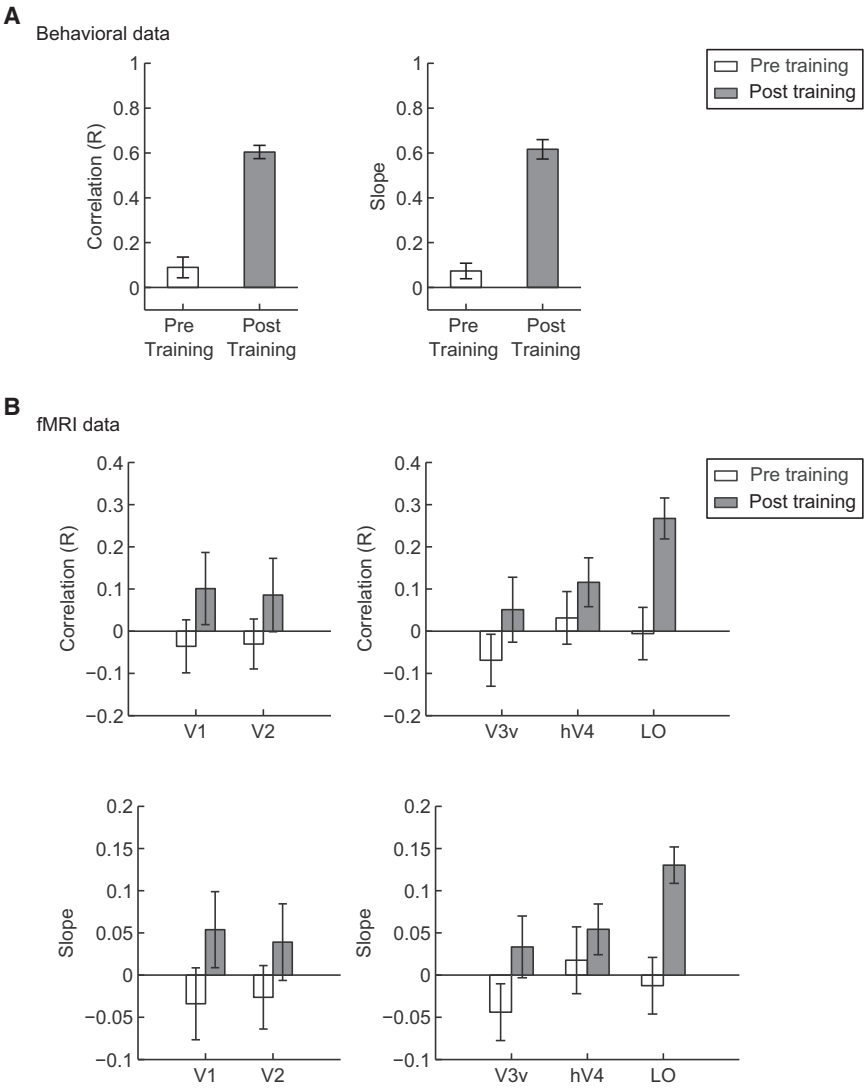

Figure 3. Comparison of Behavioral, fMRI, and Ideal Class Distances

(A and B) Correlations of behavioral (A) and fMRI (B) class distances with the ideal class distance for each of the 21 informative image parts. fMRI signals were derived from training the SVM classifier on the basis of the choices of the human observers (Figures 3 and S2B) or the ideal observer (Figure S2A) with the classifier's performance. Correlations were performed for each participant before and after training. Mean R coefficient and slope values across participants are plotted. Error bars indicate the SEM. Because of noisy BOLD signals, these values are lower for fMRI than behavioral data.

link behavioral improvement after training to enhanced fMRI selectivity [19]. However, this multivariate analysis alone does not allow us to identify the informative image regions that support perceptual decisions. Our approach—comparing human and fMRI classifier performance to an ideal observer—provides the first neuroimaging evidence that learning does not simply modulate overall activity magnitude but tunes the representation of informative image parts in the higher ventral cortex in order to support efficient shape discrimination.

Second, previous imaging studies using pattern classification approaches for the analysis of electroencephalography and fMRI data [8, 9] have shown that behaviorally relevant features (e.g., face features) are represented in visual areas selective for their processing. However, these studies did not test the role of learning in shaping decision templates given that familiar stimuli were used. We purposefully chose a novel stimulus space for investigating the role of learning in optimizing decision templates and simple, but carefully controlled, stimuli that are suitable for studying processing along the ventral visual stream. Our findings demonstrate that learning tunes the representation of image parts in higher ventral areas. Our approach combining reverse correlation with pattern classifiers could be extended further to extract classification images from fMRI signals in higher temporal areas related to more complex naturalistic stimuli. Given that these more anterior portions of the ventral hierarchy are understood to take their inputs from more posterior regions, we would expect similar effects in these regions for more complex objects, as we demonstrate here for posterior occipitotemporal regions with simple shapes.

Finally, the high-resolution imaging adopted in our study afforded us the signal quality necessary to reveal multivoxel patterns that represent fine image parts, but it restricted brain coverage to the posterior occipitotemporal cortex (i.e., no significant activations were observed for our stimuli anterior to LO). Previous work has also implicated frontoparietal circuits in flexible perceptual decisions [20, 21]. Given the complex nature of the BOLD signal, it is possible that the fMRI selectivity that we observed for informative image parts in higher ventral areas is enhanced by feedback from frontoparietal circuits that may reweight sensory signals in visual areas [22]. It is also important to note that—despite the enhanced sensitivity of our methodology—multivoxel pattern classification approaches reveal neural preferences at the scale of large neural populations rather than the tuning of individual neurons. Therefore, understanding the cortical circuits that support adaptive brain processes for perceptual decisions requires further whole-brain connectivity studies combining advanced imaging and neurophysiological techniques.

Experimental Procedures

Stimuli

Two classes of shapes (pentagons) were generated by manipulating the location of the pentagon lines that differed in their lengths. We added positional noise to all stimuli that were produced by radially shifting the position of each polygon dot on the basis of a Gaussian distribution (mean = 0, SD = 0.4°). Previous work has shown that the human classification images derived with positional noise are independent of the signal strength [23, 24]. To assess this, we tested the observers' ability to classify zero signal stimuli

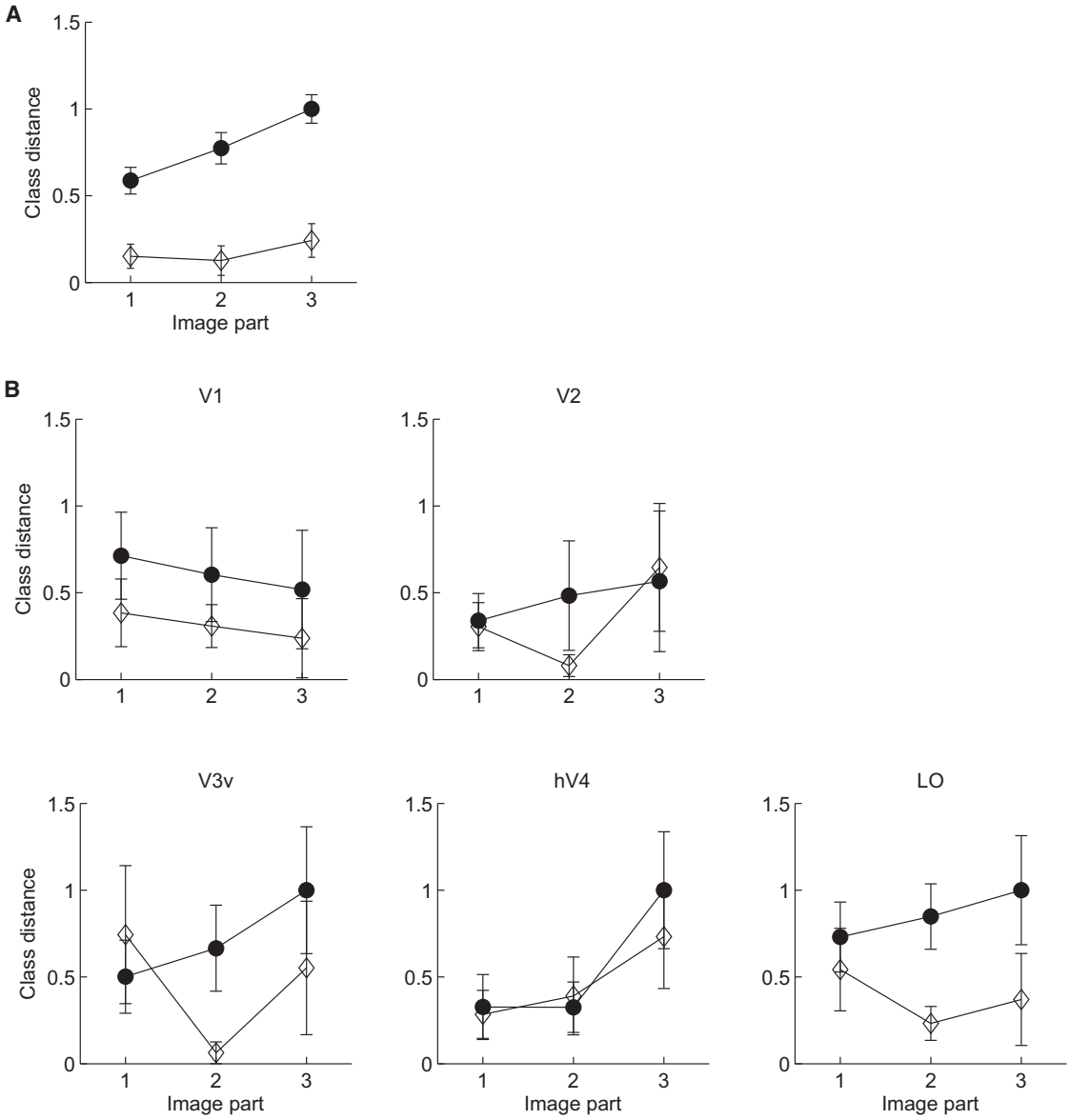

Figure 4. Class Distances before and after Training

(A and B) Behavioral (A) and fMRI (B) class distances are shown at three image locations with maximum and minimum class distance (see Figure 1B) before (open diamonds) and after (filled circles) training. Class distances were normalized by setting minimum and maximum values to 0 and 1, respectively. Error bars indicate SEM across participants. A control analysis (Figure S3) provides evidence that spatial correlations across image locations could not drive observer performance.

(i.e., random dot stimuli). After training, classification images for these stimuli were very similar to those obtained in our main experiment, suggesting that the classification images we observed are independent of signal strength.

Design

Observers participated in a pretraining fMRI session, three to five behavioral training sessions (900–1,100 trials per session, depending on participant availability), and two posttraining fMRI sessions (see the Supplemental Information). The study was approved by the University of Birmingham Ethics Committee.

Classification Image Analysis

We calculated behavioral and fMRI classification images after rotating and resizing each stimulus image to a standard orientation and size. We used the noise fields (i.e., noise perturbations across trials) to compute classification images. We subtracted the average of noise fields over all trials for which the observers responded class I from the average of noise fields over all trials for which the observers responded class II (see the Supplemental Information).

Supplemental Information

Supplemental Information contains Supplemental Experimental Procedures and three figures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2013.07.052.

Acknowledgments

We are grateful to M. Eckstein, D. Kersten, S. Klein, and B. Tjan for helpful discussions. This work was supported by grants from the Biotechnology and Biological Sciences Research Council (D52199X and E027436) and the European Community's Seventh Framework Programme

(FP7/2007-2013) under agreements 255577 and 214728 and Research Executive Agency grant agreement PITN-GA-2011-290011 to Z.K. and by grant RO1EY01728 from the National Eye Institute (NIH) to D.L.

Received: January 11, 2013
Revised: June 10, 2013
Accepted: July 16, 2013
Published: September 5, 2013

References

1. Li, R.W., Levi, D.M., and Klein, S.A. (2004). Perceptual learning improves efficiency by re-tuning the decision 'template' for position discrimination. Nat. Neurosci. *7*, 178–183.
2. Dobres, J., and Seitz, A.R. (2010). Perceptual learning of oriented gratings as revealed by classification images. J. Vis. *10*, 8.
3. Dupuis-Roy, N., and Gosselin, F. (2007). Perceptual learning without signal. Vision Res. *47*, 349–356.
4. Gold, J., Bennett, P.J., and Sekuler, A.B. (1999). Signal but not noise changes with perceptual learning. Nature *402*, 176–178.
5. Abbey, C.K., and Eckstein, M.P. (2002). Classification image analysis: estimation and statistical inference for two-alternative forced-choice experiments. J. Vis. *2*, 66–78.
6. Eckstein, M.P., and Ahumada, A.J., Jr. (2002). Classification images: a tool to analyze visual strategies. J. Vis. *2*, 1x.
7. Murray, R.F. (2011). Classification images: A review. J. Vis. *11*, 11.
8. Smith, F.W., Muckli, L., Brennan, D., Pernet, C., Smith, M.L., Belin, P., Gosselin, F., Hadley, D.M., Cavanagh, J., and Schyns, P.G. (2008). Classification images reveal the information sensitivity of brain voxels in fMRI. Neuroimage *40*, 1643–1654.
9. Schyns, P.G., Gosselin, F., and Smith, M.L. (2009). Information processing algorithms in the brain. Trends Cogn. Sci. *13*, 20–26.
10. Tong, F., and Pratte, M.S. (2012). Decoding patterns of human brain activity. Annu. Rev. Psychol. *63*, 483–509.
11. Furmanski, C.S., and Engel, S.A. (2000). Perceptual learning in object recognition: object specificity and size invariance. Vision Res. *40*, 473–484.
12. Rubin, N., Nakayama, K., and Shapley, R. (1997). Abrupt learning and retinal size specificity in illusory-contour perception. Curr. Biol. *7*, 461–467.
13. Sigman, M., Pan, H., Yang, Y., Stern, E., Silbersweig, D., and Gilbert, C.D. (2005). Top-down reorganization of activity in the visual pathway after learning a shape identification task. Neuron *46*, 823–835.
14. Yotsumoto, Y., Watanabe, T., and Sasaki, Y. (2008). Different dynamics of performance and brain activation in the time course of perceptual learning. Neuron *57*, 827–833.
15. Mukai, I., Kim, D., Fukunaga, M., Japee, S., Marrett, S., and Ungerleider, L.G. (2007). Activations in visual and attention-related areas predict and correlate with the degree of perceptual learning. J. Neurosci. *27*, 11401–11411.
16. Kourtzi, Z., Betts, L.R., Sarkheil, P., and Welchman, A.E. (2005). Distributed neural plasticity for shape learning in the human visual cortex. PLoS Biol. *3*, e204.
17. 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.
18. Kourtzi, Z. (2010). Visual learning for perceptual and categorical decisions in the human brain. Vision Res. *50*, 433–440.
19. Zhang, J., Meeson, A., Welchman, A.E., and Kourtzi, Z. (2010). Learning alters the tuning of functional magnetic resonance imaging patterns for visual forms. J. Neurosci. *30*, 14127–14133.
20. Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. Nat. Rev. Neurosci. *2*, 820–829.
21. Li, S., Ostwald, D., Giese, M., and Kourtzi, Z. (2007). Flexible coding for categorical decisions in the human brain. J. Neurosci. *27*, 12321–12330.
22. Dosher, B.A., and Lu, Z.L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. Proc. Natl. Acad. Sci. USA *95*, 13988–13993.
23. Li, R.W., Klein, S.A., and Levi, D.M. (2006). The receptive field and internal noise for position acuity change with feature separation. J. Vis. *6*, 311–321.
24. Li, R.W., Klein, S.A., and Levi, D.M. (2008). Prolonged perceptual learning of positional acuity in adult amblyopia: perceptual template retuning dynamics. J. Neurosci. *28*, 14223–14229.