Fingerprints of Learned Object Recognition Seen in the fMRI Activation Patterns of Lateral Occipital Complex

Zvi N. Roth[1,2,3] and Ehud Zohary[1,2,3]

[1]Department of Neurobiology, [2]Interdisciplinary Center for Neural Computation and [3]Edmond & Lily Safra Center for Brain Sciences, Hebrew University, Jerusalem 91904, Israel

Address correspondence to Zvi N. Roth, Department of Neurobiology, Life Sciences Institute and the Interdisciplinary Center for Neural Computation, Hebrew University, Jerusalem 91904, Israel. Email: zvi.roth@mail.huji.ac.il; Ehud Zohary, Department of Neurobiology, Life Sciences Institute and the Interdisciplinary Center for Neural Computation, Hebrew University, Jerusalem 91904, Israel. Email: udiz@cc.huji.ac.il.

One feature of visual processing in the ventral stream is that cortical responses gradually depart from the physical aspects of the visual stimulus and become correlated with perceptual experience. Thus, unlike early retinotopic areas, the responses in the object-related lateral occipital complex (LOC) are typically immune to parameter changes (e.g., contrast, location, etc.) when these do not affect recognition. Here, we use a complementary approach to highlight changes in brain activity following a shift in the perceptual state (in the absence of any alteration in the physical image). Specifically, we focus on LOC and early visual cortex (EVC) and compare their functional magnetic resonance imaging (fMRI) responses to degraded object images, before and after fast perceptual learning that renders initially unrecognized objects identifiable. Using 3 complementary analyses, we find that, in LOC, unlike EVC, learned recognition is associated with a change in the multivoxel response pattern to degraded object images, such that the response becomes significantly more correlated with that evoked by the intact version of the same image. This provides further evidence that the coding in LOC reflects the recognition of visual objects.

Keywords: fMRI, LOC, MVPA, object perception, perceptual learning

Introduction

The hierarchy of visual processing in the visual ventral stream is evident in the increasing degree of invariance to alterations of certain features in the image (as long as the object identity is clearly distinguishable; Grill-Spector et al. 2001; Grill-Spector and Malach 2004). Thus, while primary visual cortex (V1) is typically highly sensitive to changes in the location, contrast, or retinal size of visual stimuli, regions within the lateral occipital complex (LOC) are largely unaffected by such manipulations (Grill-Spector et al. 2001; Avidan et al. 2002; Sawamura et al. 2005; Konkle and Oliva 2012, respectively).

Furthermore, the response in LOC is often quite specific to an object category, such that various stimuli from a given category (e.g., teapots) elicit quite similar patterns of response across voxels, which are distinctly different from those elicited by images from another category (e.g., chairs; Eger et al. 2008). Importantly, this result is not merely a consequence of a possible difference in the low-level features (e.g., size) that may characterize the different object categories. These findings suggest that LOC activation patterns reflect a perceptual, high-level representation of visual objects.

Here, we attempt to corroborate these findings by manipulating the observers' *percept* while presenting the same physical stimulus. A similar approach has been utilized before

using various techniques. One classical method has been the induction of rival percepts, either binocular or monocular, such that the percept switches between the 2 images seen by the left and right eye (in the binocular case; Tong et al. 1998; Haynes and Rees 2005), between 2 interpretations of an ambiguous image such as Rubin's vase–face picture (Hasson et al. 2001; Andrews et al. 2002; Hesselmann et al. 2008) or between visibility and invisibility of part of a stimulus as a result of surrounding motion (Bonneh et al. 2001). Such studies utilizing bistable perception linked changes in perception (e.g., the percept switching between the binocular rivals) to fluctuations in brain activity in both low- and high-level visual areas (Tong et al. 2006; Donner et al. 2008; Sterzer et al. 2009). Another approach was to focus on permanent, long-term perceptual transitions by means of priming experiments, in which degraded or black-and-white Mooney images which were unrecognizable at first were easily identified after exposure to the original, intact images (Dolan et al. 1997).

These previous studies showed that the average *magnitude* of the response in the occipito-temporal cortex is modulated by the percept. For example, face-selective regions were more active when the ambiguous image was interpreted as a face than when the same image was perceived as another object (Andrews and Schluppeck 2004). More recent studies have focused on the patterns of functional magnetic resonance imaging (fMRI) activation, showing a correspondence between the evoked pattern and the specific percept [e.g., the object's perceived shape (Williams et al. 2007; de Beeck et al. 2008; Haushofer et al. 2008; Drucker and Aguirre 2009) or identity (Hsieh et al. 2010; Gorlin et al. 2012)]. For example, Hsieh et al. presented Mooney images to subjects before and after they saw the intact versions of the images. Thus, they could compare between the activation patterns when the Mooney images were or were not recognized with the patterns in response to the intact images. Surprisingly, they found an effect of recognition in early visual cortex (EVC; i.e., foveal confluence) and in LOC when the experiment was block-designed. Using a similar image degradation technique, Gorlin et al. found that primed images are more accurately decoded in anatomically defined left pericalcarine cortex. However, unlike Hsieh et al., they did not find significant priming effects in any *functionally* defined regions of interest (ROIs) within the EVC. It is therefore unclear if and how perceptual changes are reflected by changes in the *patterns* of activation in the visual cortex.

Our focus here was to test if changes in the patterns of brain activation that result from a perceptual shift occur in a *predictable* direction. Specifically, we hypothesized that the activation patterns would become more similar to the patterns of

response to the intact versions of the same images. To that end, subjects viewed highly degraded images of objects, before and after a short learning phase which boosted the participants' ability to identify those objects. Learned recognition was associated with a change in the multivoxel response pattern to degraded object images, such that the response pattern had become significantly more correlated with that evoked by the intact version of the same image. No such change was evident in early visual cortical areas. These results further indicate that LOC reflects the perceptual level of representation of visual objects.

Materials and Methods

Subjects

Fourteen healthy subjects gave their informed consent to participate in the fMRI study, which was approved by the Helsinki Ethics Committee of Hadassah Hospital, Jerusalem, Israel. Two subjects were excluded from further analysis due to excessive movement during the scan, and an additional subject was excluded due to the absence of sufficient activation (<50 voxels) during the localizer scan. Thus, the fMRI data from 11 subjects (mean age = 25 ± 3, 3 females) served as the database for this study.

Stimuli

Four animate and 4 inanimate colored natural images were cropped and downscaled to 512 × 512 pixels (6.9 × 6.9°; see example in Fig. 1A; all other images are shown in Fig. 1E). Using a phase scrambling method, we created images that were degraded to a varying degree by a linear interpolation between the original image and a noise image (Rainer and Miller 2000; Rainer et al. 2004): First, we calculated the average Fourier spectrum of all 8 images. Next, we combined each image's Fourier phase with the average spectrum and applied the inverse Fourier transform to yield Fourier normalized images (see example image in Fig. 1B). We also combined the average spectrum with a uniformly distributed random Fourier phase, thus creating the "noise" image (Fig. 1C). We then linearly combined the noise image
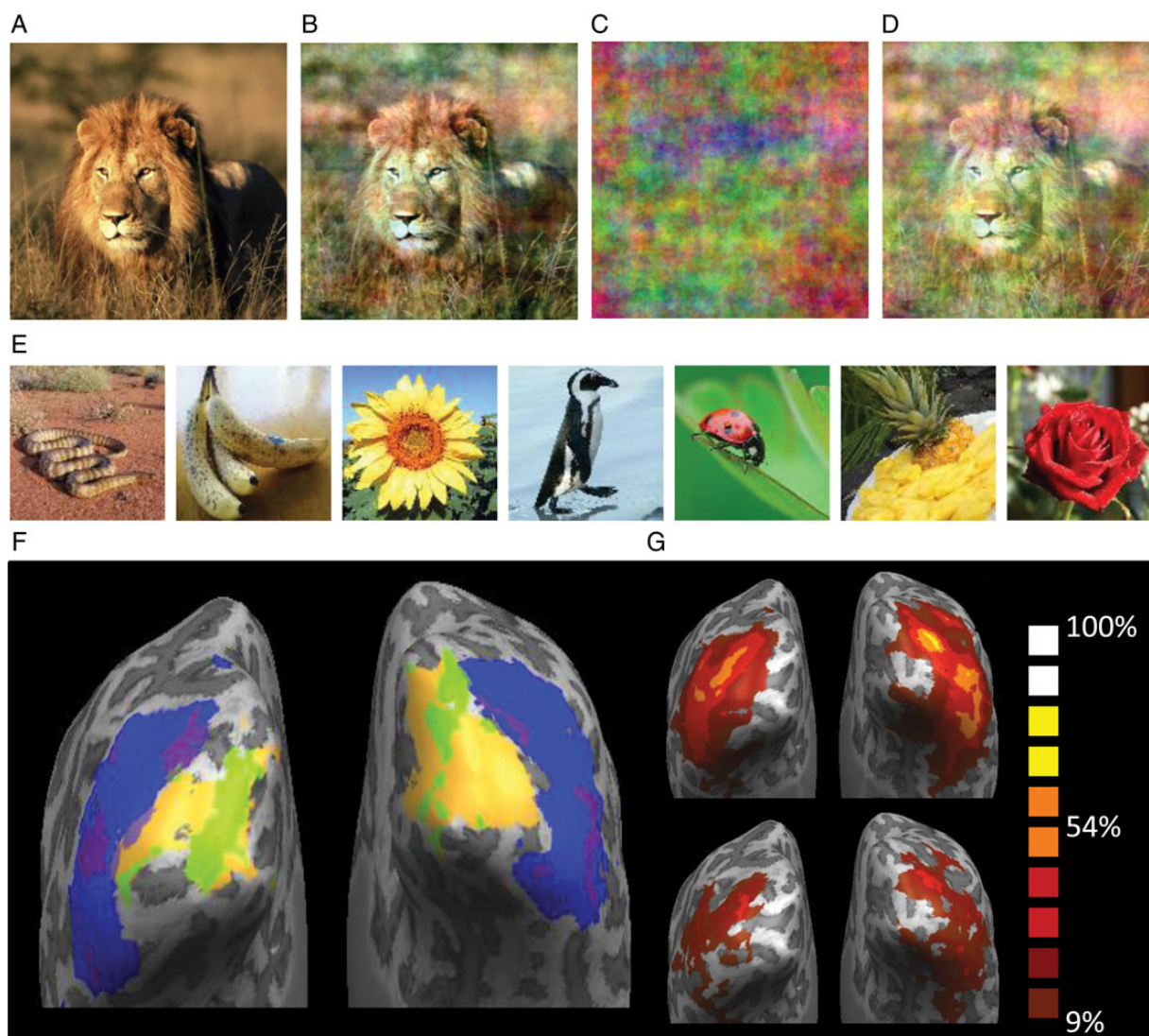
Figure 1. Example of the stimuli used in the experiment. (A) An original intact object image. (B) The same image after Fourier power spectrum normalization. (C) Noise image, having the same power spectrum as the normalized image (in B), with randomized Fourier phase values. (D) Degraded object image, generated by a linear combination of B and C. This specific image is shown for illustration purposes only; actual images used in the experiment were more degraded. (E) All 7 other intact images. (F) The location of the 2 ROIs (LOC colored in blue and EVC in orange) and the 150 most active voxels within each ROI (LOC in purple and EVC in green) projected on the posterior view of the cortical surface, in one representative subject. (G) Probability maps for LOC ROI (top) and the 150 most active voxels from LOC (bottom) projected on the posterior view of the cortical surface. The color coding reflects the probability that a specific Talairach voxel was within the LOC ROI (top), or within the top 150 voxels in LOC (bottom).

with each of the 8 normalized images (see example image in Fig. 1D), at 36 even-spaced noise levels (from 0.10–0.16 stimulus coherence to 0.22–0.28 coherence, depending on the specific image visibility assessed in a pilot study), thus generating a family of images with a monotonically increasing signal level. For each subject, we used a different noise image. These images were used in the 2 runs of the experiment. In between these runs, we had a short learning procedure. To assist learning, we generated "Hysteresis clips" displaying a gradual transformation of the noise image into an intact Fourier normalized image, and back to the noise. Each of the Hysteresis clips was shown at 15 frames per second and consisted of 120 frames (8 s).

Scrambled images (used in the localizer scan) were created by dividing each image into 8 × 8 pixel squares and randomly scrambling them.

Stimuli were presented using the Presentation software (Neurobehavioral Systems, Albany, CA, USA). Visual stimuli were projected via an MR shielded projector onto a screen located 114 cm from the participants. The screen was made visible to the subjects via a tilted mirror, positioned above the subjects' faces. The screen's dimensions were 53 cm wide and 30 cm high (26.4° × 15.1°). The display resolution was 1920 × 1080 pixels.
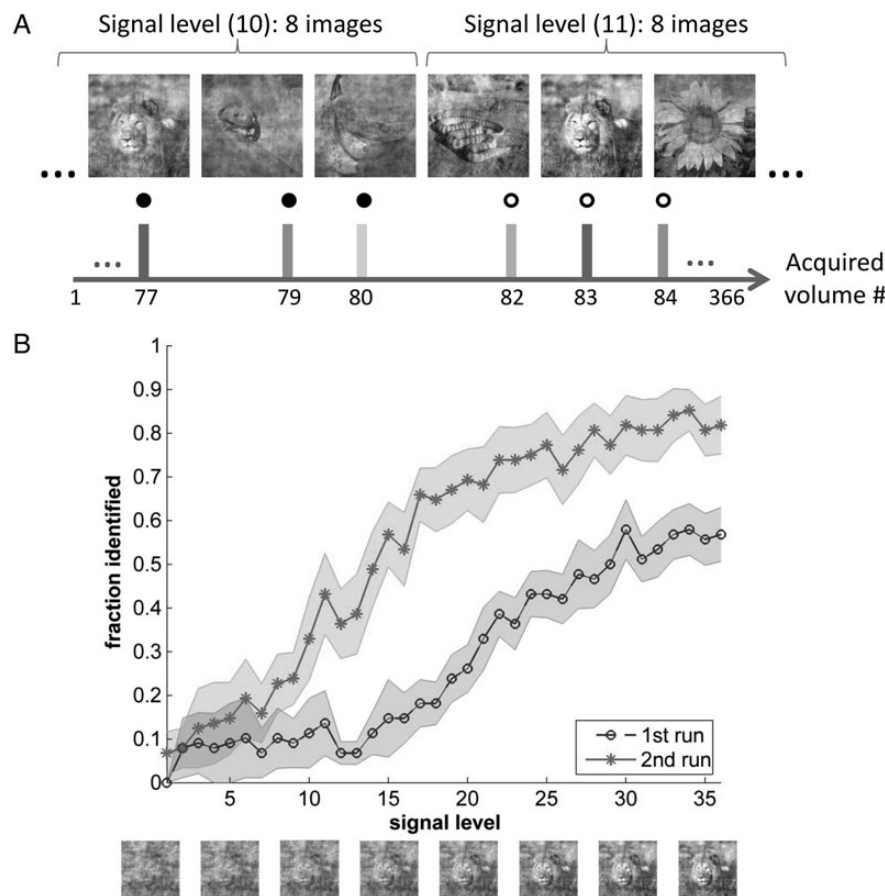
Experimental Procedure

The fMRI experiment consisted of 2 test runs (comprised of degraded images), in addition to an intact-image run and a functional localizer. For the first 4 subjects, the scans were in this order, and for the other 7 the intact-image run preceded the second test run. This change in the protocol was planned in an effort to maximize gains in the recognition of the degraded images, by exposing participants to the intact version of the same images. But in fact, post hoc, we found that this change had little effect on performance. The percentage of recognition reports was not statistically different between the 2 groups (two-way ANOVA, $P = 0.16$). During all scans, subjects were instructed to fixate on a central fixation point, which was displayed during the entire experiment.

Immediately prior to the scanning session, subjects were briefly trained on the recognition of degraded images of man-made objects, to allow them to be acquainted with the experimental procedure during the scan. This initial training consisted of 12 degraded images of man-made objects and food, and 5 null trials, and was repeated until subjects were sure they understood and felt comfortable with the task. These object images were not used in the main experiment, and the subjects were not told about the identity of the objects used in the experiments.

The event-related test runs consisted of image trials and null trials (Fig. 2A). In each image trial, a degraded image was presented for 200 ms followed by 1800 ms during which the subject had to press one button if the image was identified and a second button if it was not. In the null trials, lasting 2 s, no images were presented and no response was required. Each of the test runs lasted 12:12 min (366 volumes), consisting of 2 initial noise-image trials, 36 test blocks, and 4 final null trials. Each block consisted of 10 trials: 8 image trials (of all the animal/nonanimal images, at a specific coherence level) and 2 null trials, in a randomized order. The signal level increased monotonically with the blocks from level 1 (most degraded) to 36 (least degraded), so that during the last block the images were most easily identified.

Figure 2. Experimental procedure and behavioral results. (A) Experimental procedure: During the event-related fMRI scan, the signal level of all 8 images gradually increased by incremental steps from level 1 (most degraded) to level 36 (highest signal). Bars schematically represent the time of presentation of different images at 2 adjacent signal levels (levels 10 and 11 in this specific example). Subjects reported by button press if they did or did not identify each image (empty and filled circles, correspondingly, in the schematic time course). (B) The mean fraction of identified degraded images in the first and second test runs (averaged across subjects) as a function of the signal level in the image. Shaded areas denote the standard error of the mean (SEM). Images at different degradation levels are shown below the graph for illustration purposes only; actual images were in color and more degraded.

Next, we had an event-related intact-image run, consisting of 16 repetitions of each of the 8 original, intact images (without Fourier normalization) and the noise image. Each image was presented for 200 ms followed by 1800 ms of fixation. After adding 64 null trials (2 s with no image), the order of trials was pseudorandomized using Optseq (http://freesurfer.net/optseq/). An initial null trial and 4 final null trials brought the run duration to a total of 7:06 min (213 volumes). In this run, subjects were instructed to press a button every time the same intact image was repeated (1-back task).

Before the second test run, subjects viewed a hysteresis video clip (see Stimuli) of each specific object, 3 times. To avoid misguided learning of a specific noise pattern associated with the image, each of the 3 hysteresis video clips for a given object was generated with a different noise image. The second test run was identical to the first, apart from the intrablock order which was randomized anew.

ROI Selection: Functional Localizer

The final run was a localizer scan, comprised of 7 blocks of intact images and 7 blocks of the scrambled versions of those images, in an alternating order. Each block consisted of 16 images (8 animate, 8 inanimate, including the 8 images used for the main experiment), presented for 1 s each. An initial blank 2 s, and final blank 8 s brought the scan time to 3:54 min (117 volumes). As in the "intact images" run, subjects were instructed to respond by button press whenever a specific image was repeated (1-back task). We defined 2 ROIs: (1) The object-related, LOC, using the standard General Linear Model (GLM) contrast [intact images > scrambled images; q [false detection rate (FDR)] < 0.05; size range across subjects: 180–1383 functional voxels, mean ± standard deviation = 636 ± 450] and (2) EVC (which is more responsive to the added local contrast generated by the image scrambling process), using the opposite contrast (scrambled images > intact images; q(FDR) < 0.05; size range: 193–859 voxels, mean = 546 ± 219). The bilateral LOC ROI was located in the occipito-temporal cortex, and typically included the fusiform gyrus, occipito-temporal sulcus, inferior temporal gyrus, inferior temporal sulcus, middle occipital gyrus, lateral occipital sulcus, and superior occipital gyrus (Malach et al. 1995). The EVC ROI typically included the calcarine sulcus, occipital pole, and occipital gyri.

To further refine our choice of voxels, and select the same number of voxels from each subject, we chose a subset of the voxels within the 2 ROIs (LOC, and EVC, see Fig. 1F,G). These voxels were the most responsive [in terms of their signal-to-noise (SNR) ratio] in the intact-image run. Therefore, the choice of voxels was independent of the activation in the test runs, avoiding the problem of double dipping (Kriegeskorte et al. 2009). Specifically, we first averaged the activation t-values (across the 8 intact object images) of each voxel and then chose the 150 voxels with the highest mean t-value in each ROI (see further details in the section Data Processing, below). We present below the results of both the whole ROI analysis and the 150 voxel analysis.

MRI Scanning Parameters

The blood oxygen level-dependent (BOLD) fMRI measurements were obtained using a whole-body 3-T Magnetom Trio Siemens scanner and a 32-channel head coil. The functional MRI protocols were based on a multislice gradient-echo planar imaging and obtained under the following parameters: time repetition (TR) = 2 s, time echo = 30 ms, flip angle = 90°, imaging matrix = 64 × 64, field of view = 192 mm; 36 slices with 3-mm slice thickness and 15% gap (0.45 mm) and were oriented in the oblique position, covering the whole brain, with functional voxels of 3 × 3 × 3 mm. In addition, high-resolution, T_1-weighted magnetization-prepared rapid acquisition gradient-echo (MPRAGE) images were acquired (1 × 1 × 1 mm resolution).

Data Processing

Data analysis was conducted using the Brain Voyager QX software package (Brain Innovation) and in-house analysis tools developed in Matlab (MathWorks). Preprocessing of functional scans included 3D motion correction, slice scan time correction, and removal of low frequencies up to 3 cycles per scan (linear trend removal and high-pass filtering). The anatomical and functional images were transformed to the

Talairach coordinate system using trilinear interpolation. The cortical surface was reconstructed from the high-resolution anatomical images using standard procedures implemented by the BrainVoyager software.

Voxel time courses were generated using BrainVoyager and were then analyzed using the Matlab custom-made software. Specifically, we first transformed each voxel's time course to obtain a z-score value, by subtracting the mean activation and dividing by the standard deviation of the BOLD response across the whole run. Then, for the intact run, we used a standard GLM analysis with 9 regressors—one for each of the 8 intact images, and one for the noise image, assuming the standard (2 gamma) hemodynamic response function (Friston et al. 1998). This resulted in one activation parameter (beta weight) per regressor, for each voxel. Next, in each of the 3 analyses of the test runs (detailed below), we subdivided the experimental conditions into different categories, thereby obtaining a different number of regressors. To each regressor, we assigned a specific beta weight, based on its activation level. We then transformed the beta weights into t-values, by subtracting each voxel's mean beta weight (across all conditions; 9 conditions in the case of the intact images run) and dividing by each beta's estimator's standard deviation.

Mathematically, the transformation is given by (Worsley and Friston 1995):

$$t_i = \frac{b_i = \langle b \rangle}{\sqrt{\mathrm{var}(e)[(X^{\mathrm{T}}X)^{-1}]_{i,i}}}.$$

where b_i is the beta weight for condition i, $\langle b \rangle$ is the average beta weight over all conditions, X is the design matrix, X^{T} is its transpose, the i, i subscript indicates the ith diagonal element, and e is the voxel's residual (the difference between the voxel's actual time course and the model's predicted time course). Since the residual generally corresponds to the noise (or unexplained variance) in the measurement, the t-value reflects the signal-to-noise ratio (SNR) of each voxel.

We then rank-ordered all voxels according to their mean t-value across images of the intact run and chose the 150 top voxels in each subject's ROI. Finally, after selection of the top voxels, we normalized their t-values, so that when computing the vector of activation for a given image, each voxel's contribution was the *difference* from its mean t-value across conditions.

Multivoxel Pattern Analysis (MVPA) of Learning-Associated Changes

MVPA is based on the notion that the representation of each specific object is *distributed*, and captured by a *unique pattern* of activation across the relevant elements (i.e., voxels, see Fig. 3A; Norman et al. 2006). If this is indeed the case, we posit that recognition of a previously unrecognized (degraded) image of an object should be mirrored by a change in the pattern of responses evoked by that image, such that this pattern becomes more similar to one evoked by the corresponding intact image (see Fig. 3B).

To assess this working hypothesis, we used 3 different analyses (described and numbered below) applied to the data acquired during the test runs. These analyses provided 3 different sets of coefficients (beta weights), which were then transformed to t-values. After calculating Pearson's correlations between the patterns of t-values, the resulting correlation coefficients (Pearson's r-values) were converted using Fisher's Z transform to z-values, which were then averaged over the 8 images. All t-tests were performed on the mean z-values. Any trial in which the subject did not respond with a button press was ignored in all further analyses.

(1) *The Image-Presentation analysis*: we assessed the degree to which the multivoxel patterns of activation for the degraded images had changed between the first and second runs in relation to the multivoxel pattern evoked by the corresponding intact images presented in the intact-image run. To that end, each of the object images (in each test run) was modeled by one predictor time course in the GLM, which simply corresponded to the timing of the 36 presentations of that specific image, in its various signal levels (Fig. 4A). We then convolved each predictor time course with a standard hemodynamic response function (HRF; sum of 2 gamma functions; Friston et al. 1998) and applied linear
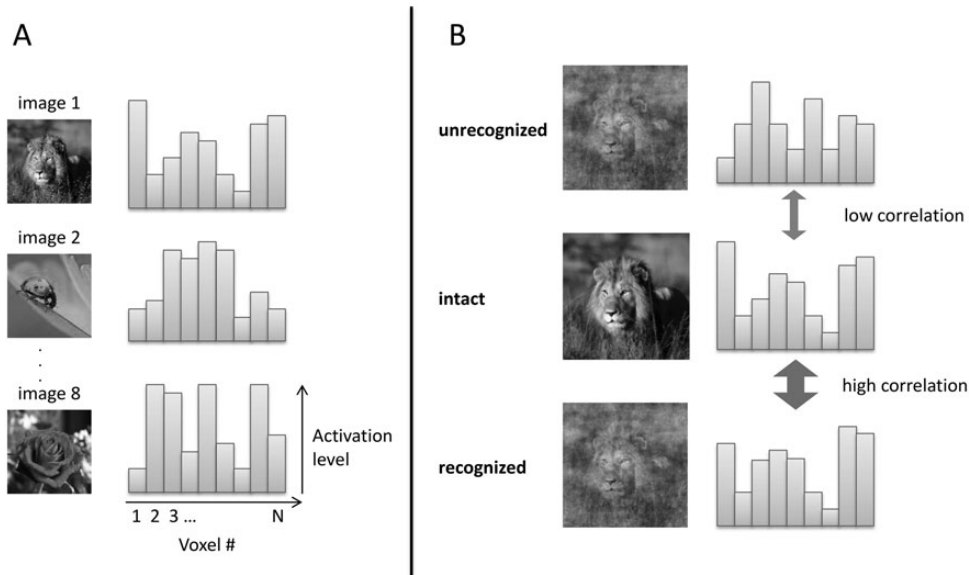
Figure 3. Schematic representation of our working hypothesis. (*A*) MVPA is based on the core idea of a distributed coding: Each specific object is represented by a *unique pattern* of activation across the relevant elements (e.g., voxels within a specific ROI). (*B*) Our assumption is that recognition of a particular visual object (e.g., lion) is reflected by a specific pattern of activation in LOC. Accordingly, recognition of a previously unrecognized (degraded) image of an object should be mirrored by a change in the pattern of responses evoked by that image, such that this pattern becomes more correlated with that evoked by the corresponding intact image (depicted by a thicker arrow in the bottom compared with the top).

regression, to extract the beta weights that best match the raw data. As a result, in each run, each voxel had one coefficient (beta) per image, which was then transformed into a *t*-value, as described with regard to the intact run. We then calculated the Pearson correlation coefficient between the voxels' pattern of response (i.e., their *t*-values) to the intact image (in the "intact" run) and their pattern of response to the degraded image (separately in each of the 2 test runs). This resulted in 16 correlation coefficients, 8 in each run, and their mean across images were calculated for each subject.

Next, we computed the correlation between the multivoxel response pattern to each degraded image with the response to the noise image. This again resulted in 16 correlation coefficients, 8 in each run.

To compare the resulting correlation coefficients to a benchmark, we also calculated the correlations between the first run's response pattern and the second run's response pattern for the same images. As an additional control, we also correlated the pattern of response to the degraded image with the responses to all 7 other intact images.

(2) *The Image-Identification analysis*: Each degraded image (across all 36 degradation levels) was modeled by 2 regressors: one for all the instances a degraded image was identified, and a second one for the times it was not identified (Fig. 5*A*). In addition, we correlated the voxels' *t*-values with their *t*-values in the intact run, only this time it was done separately for the "identified image" *t*-values and for the "unidentified image" *t*-values. This analysis enabled us to compare patterns of activity during trials that were identified and trials that were not, both before and after the learning phase. However, notice that the stimuli during identified trials were not the same stimuli as during the unidentified trials, and the less degraded images (i.e., higher stimulus level) were more likely to be identified. In addition, the number of trials included in the "unidentified" predictors was typically different from the number of trials in the "identified" predictors, depending on the subjects' responses, for each image. Both of these problems do not exist in the third analysis.

(3) *The Learned Images analysis*: Here, we focused only on the specific levels of degraded images that were not recognized in the first run but following learning, had become recognizable in the second run. Concentrating on this "learned range" (see Fig. 6) enabled us to

compare activity to identical stimuli which elicited different perceptual states. For this "image-learning" analysis, each of the 8 degraded images was modeled by 3 regressors. For the first run, these corresponded to: (i) the degraded image presentations that were not identified in both the first and second run, (ii) the presentations that were not identified during the first run but were identified during the second run, and (iii) those that were identified during the first run, regardless of their status in the second run (Fig. 6, top). For the second run, the regressors corresponded to (i) the degraded image presentations that were not identified in the second run, regardless of their status in the first run, (ii) the presentations that were not identified during the first run but were identified during the second run, and (iii) those that were identified during both runs (Fig. 6, bottom). We then concentrated on the beta weights of the second (ii) condition (constituting the learned range), converted them to *t*-values, and then calculated the correlations as described previously.

In some rare cases, the predictor time course for a certain subject and image had no corresponding trials (e.g., if the subject did not identify any degraded versions of a certain image). In such a case, the regressor was excluded from all further analysis.

Results

Our main goal in this study was to detect modifications in brain activity that mirror changes in visual perception following fast learning. To that end, subjects viewed highly degraded images of objects, before and after a short learning phase that boosted the participants' ability to identify those objects.

Behavioral Results

In general, during the second run, subjects identified the object images at an earlier phase (lower signal level) than during the first run. This improvement was significant (Wilcoxon signed-rank test, $P < 0.05$) from signal level 11 and onwards. Figure 2*B* depicts the average reported recognition level (across images and subjects) as a function of the image
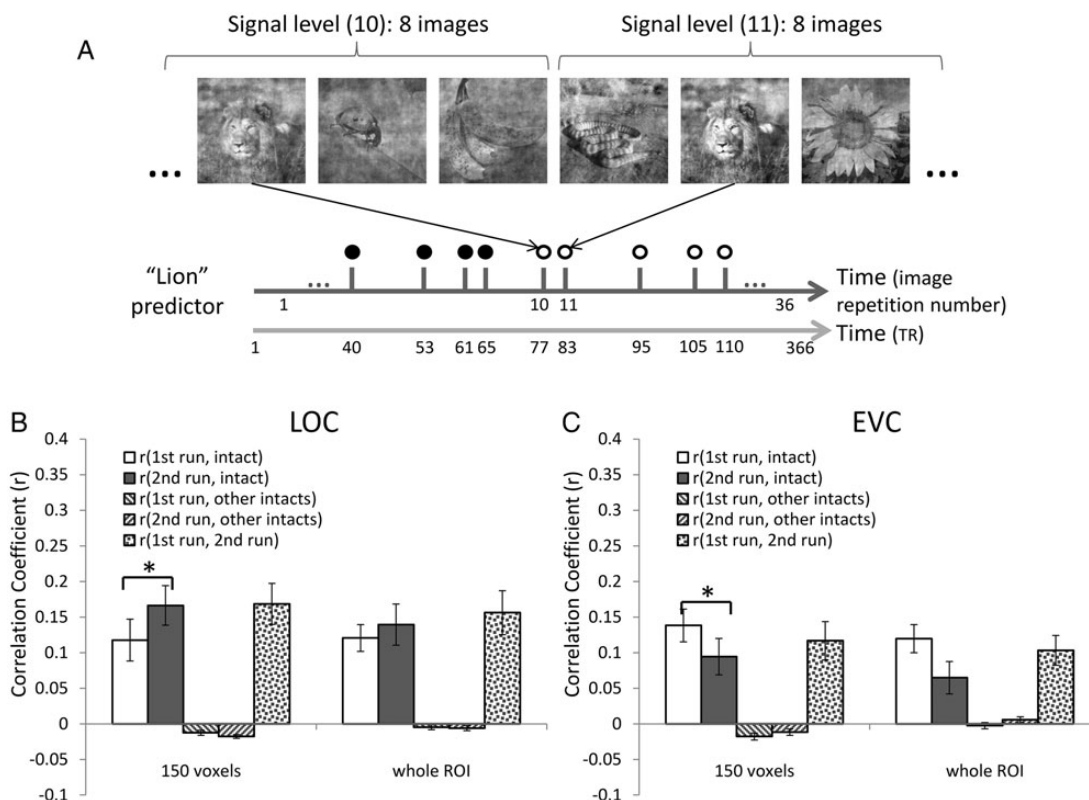
Figure 4. The image-presentation analysis. (*A*) Each image (e.g., lion) was modeled by a predictor time course reflecting the timing of the 36 presentations of that specific image within a test run (see image repetition time axis; note that sequential same image-presentation times can be quite apart from one another; lower time axis, in TR). Predictor time courses were then convolved with the HRF and linear regression was applied to each voxel's time course to obtain an activation parameter (beta weight) per image, for each run. The beta weights were then converted to *t*-values. Correlations were calculated between the multivoxel *t*-values for each (degraded) image, and the multivoxel *t*-values for the intact versions of the same image and other images. (*B*) Mean correlation (across images and subjects) between the *t*-values of LOC voxels' response to the degraded versions of an image in a specific run, with the same voxels' *t*-values evoked by the intact image (white bar, first run and gray bar, second run), or the intact version of the other object images (downward diagonal, first run and upward diagonal, second run). The dotted bars represent the correlations between the patterns evoked by the *same* degraded images in the 2 runs. The left panel describes the results using the top 150 voxels within LOC, while the right panel is the results obtained using the pattern throughout LOC. Statistically significant effects are denoted by *$P < 0.05$. Error bars here and in all following figures are SEMs. (*C*) The results of the same analysis, when applied to EVC.

signal level. While we could not verify that the subjects have indeed recognized correctly the images during the scan, pilot studies outside the scanner, in which subjects had to *verbally* report the recognized object of other images indicate that these reports are highly accurate. Overall, behavioral performance rose with an increasing signal level (e.g., decreasing degradation levels), although it did not rise monotonically.

fMRI Data

Next, we focus on object-related brain areas in the LOC within the occipito-temporal cortex. Our working hypothesis was that if the pattern of response in LOC is related to the perceptual experience (rather than just to the physical aspects of the visual stimulus), this should be mirrored by a *predicted* change in the pattern of activation, following learning. Specifically, the activation pattern elicited by a degraded image which becomes recognizable (following repeated exposure) should become more similar to one evoked by the corresponding intact image.

To test this at length, we present below 3 complementary analysis methods: at a coarse level: full first and second run performance comparison ("image-presentation" analysis); at an intermediate level: comparison of the patterns in the first and second run, splitting the data between images that were recognized, and those that were not ("image-identification"

analysis); and at a fine level: using only instances of the learned images, which were unidentified in the first run and recognized in the second ("learned image" analysis).

The Image-Presentation Analysis

This analysis compares the patterns of activation elicited for each image in the first and second run as a whole, regardless of whether that image was identified by the subject or not. It bears on the fact that, on average, performance was clearly better in the second run (as seen by the leftward shift of the performance function in Fig. 2). Following learning, this should be reflected in a better correspondence between the multivoxel pattern of activity evoked by the degraded presentations of an image, to those evoked by the intact image of the same object.

To that end, for each test run, we modeled each degraded image, presented at 36 different degradation levels, with one regressor (Fig. 4*A*). For every subject, we correlated the multi-voxel pattern of response in LOC to each degraded image with the response to the intact image. In each run, Pearson's correlation coefficients were used to quantify the similarity between the 2 different patterns of activity evoked by a degraded image and its corresponding intact image

In the LOC, the active voxels' response pattern to degraded images in the second run were indeed significantly more
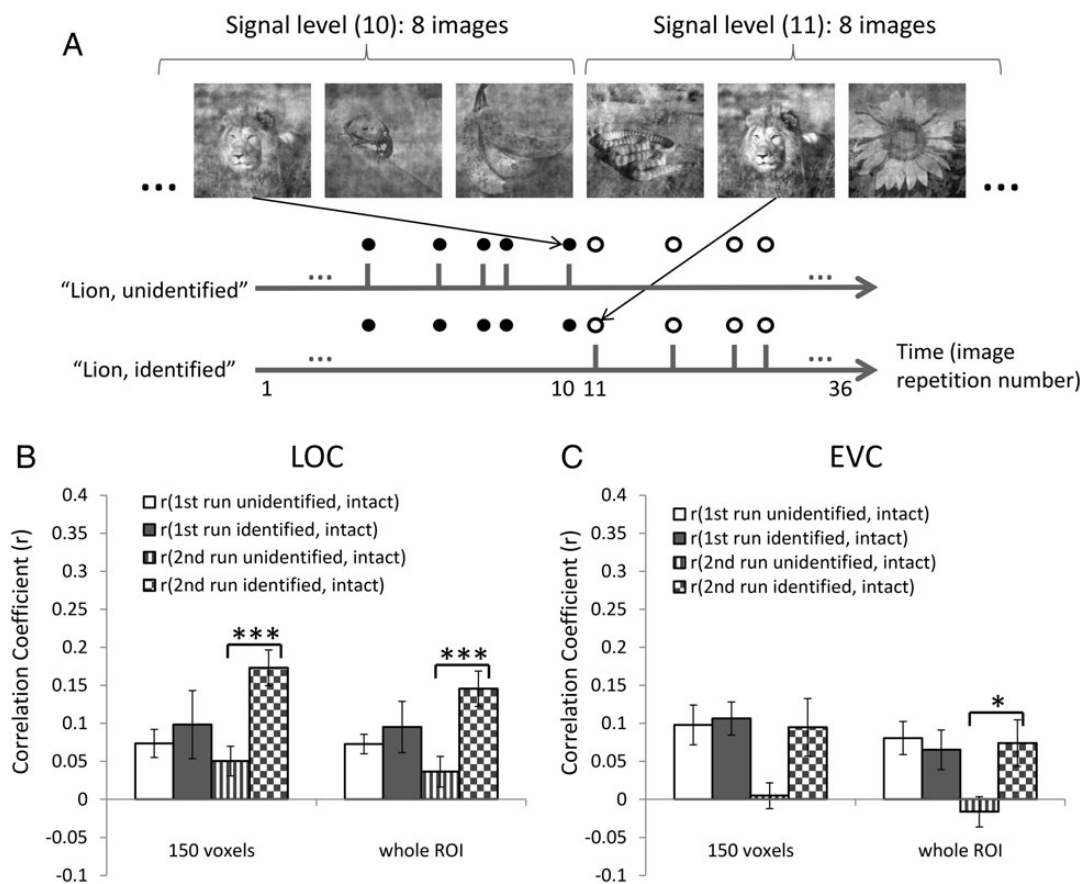
Figure 5. The image-identification analysis. (*A*) In each test run, 2 time courses were generated per image: one for the degraded images that were *identified* (empty circles) and another for those that *were not identified* (filled circles). Each time course was then convolved with the HRF. Linear regression resulted in 2 beta weights for each image, per voxel, which were converted to *t*-values. Correlations were calculated between the pattern of *t*-values for each image, identified and unidentified, with the *t*-values for the intact versions. (*B*) Mean correlations (across subjects and images) of the pattern of responses to identified and unidentified degraded levels of an image with the response to the same intact image in LOC (using 150 voxels: left; or the whole ROI: right). Significantly greater similarity to the intact pattern is seen for the *identified* levels than for the *unidentified* levels during the second run (***$P < 0.001$, paired *t*-test), but not during the first run. (*C*) The results of the same analysis applied to EVC.

correlated with the responses to the same intact images, than when the same comparison was based on the responses during the first run [see Fig. 4*B*, left panel. r(first run, intact) = 0.12 ± 0.03; r(second run, intact) = 0.16 ± 0.03; paired *t*-test: $t_{(10)} = 2.52$, $P < 0.01$], whereas in the EVC, the correlation was significantly higher for the first run [Fig. 4*B*, right panel; r(first run, intact) = 0.14 ± 0.02; r(second run, intact) = 0.1 ± 0.02; $t_{(10)} = 2.49$, $P < 0.05$]. When using all the voxels in each ROI, the trends remained visible, but were statistically insignificant in both LOC and EVC.

In both LOC and EVC, correlations with the response to the noise image were not significantly different between the 2 runs. We also correlated the pattern of response (of the most responsive voxels) to each degraded image with the patterns to each of the 7 *other* intact images. In both runs, this resulted in significantly reduced correlations than for the matching intact image, both in LOC and in EVC [LOC: r(first run, other intacts) = −0.02 ± 0.02; paired *t*-test comparing to correlation with matching intact: $t_{(10)} = 3.88$, $P < 0.01$; r(second run, other intacts) = −0.02 ± 0.02; $t_{(10)} = 5.78$, $P < 0.001$; EVC: r(first run, other intacts) = −0.02 ± 0.01; $t_{(10)} = 5.74$, $P < 0.001$; r(second run, other intacts) = −0.01 ± 0.02; $t_{(10)} = 3.86$, $P < 0.01$]. Finally, the mean activation level (beta weights) elicited by the degraded images, averaged across voxels and subjects, did not differ significantly between the runs, in both LOC and EVC. If

anything, the trend was for a lower beta weight in the second run. Thus, the greater match of the pattern of activation evoked by the degraded images with that evoked by the corresponding intact image in the second run is *not* merely due to an increase in the SNR of the BOLD response, leading to a more reliable pattern of responses. Rather, it seems to reflect a subtle change in the pattern itself, such that it becomes (somewhat) similar to that evoked by the intact stimulus.

The Image-Identification Analysis

In this analysis, we divided the 36 different degradation levels of each image, post hoc, into 2 groups: Those that were identified and those that were not (Fig. 5*A*). Then, for each image (e.g., "lion"), we correlated the response pattern evoked by the *identified* degradation levels of that image with the response to the intact image, and also with the response to the noise image. The same procedure was repeated, correlating the pattern of response elicited by the *unidentified* degradation levels of that image with the responses to the intact image, and the noise image. The rationale was that in areas whose activity is related to the perceptual experience, noisy but identifiable images should have greater similarity to the intact image (as both are recognized) than unidentifiable images.
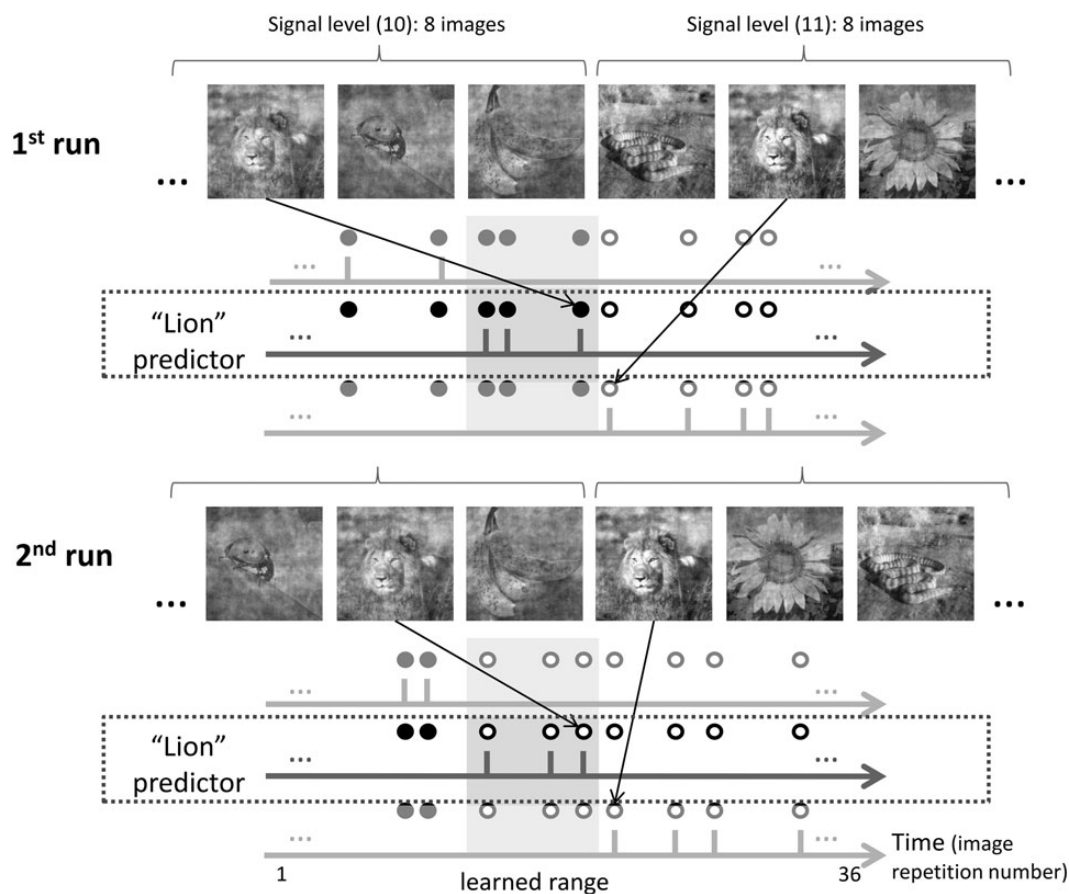
Figure 6. The image-perception analysis. Each image had 3 predictor time courses: one for the signal levels that were identified (empty circles) in both tests (top time course of each run), a second for the levels that were not identified in the first run, but were identified in the second (shaded area, representing the learned range), and a third for those that were not identified in the second run (bottom time course). Each time course was then convolved with the HRF. Linear regression resulted in 3 beta weights for each image, per voxel, which were converted to t-values. Correlations between multivoxel patterns were calculated only between the pattern of t-values from the second learned range condition in each of the 2 runs, and the t-values for the intact-image versions. The other 2 t-values for each image were not used for any further analyses.

In the LOC, during the second test run, the average correlation between the identified image response and the intact-image response was indeed significantly higher than the correlation between the unidentified image and the intact-image responses [Fig. 5B, left panel; 150 most responsive voxels: Pearson's r(second run unidentified, intact) = 0.05 ± 0.02; r(second run identified, intact) = 0.17 ± 0.02; paired t-test; $t_{(10)}$ = 5.76, $P < 0.001$ and whole ROI: r(second run unidentified, intact) = 0.04 ± 0.02; r(second identified, intact) = 0.15 ± 0.02; $t_{(10)}$ = 5.02, $P < 0.001$]. This was not the case in the first run [150 voxels: r(first run unidentified, intact) = 0.07 ± 0.02; r(first run identified, intact) = 0.10 ± 0.04; $t_{(10)}$ = 1.55, insignificant and whole ROI: r(first run unidentified, intact) = 0.07 ± 0.01; r(first run identified, intact) = 0.10 ± 0.03; $t_{(10)}$ = 1.60, insignificant]. One possible reason for this is that exposure to the intact image, which was presented between the 2 runs, removed the inherent doubt about the image identity which might have been prevalent in the first run (see Discussion), leading to a greater correlation between the noisy representation of the image and its template (the pattern of activation for the intact image).

When the same analysis was performed on the data from EVC, we found an effect in the same direction as LOC that was significant only when using all the ROI's voxels, but not when

we chose 150 voxels [Fig. 5B, right panel; 150 voxels: r(second run unidentified, intact) = 0.00 ± 0.02; r(second run identified, intact) = 0.09 ± 0.04; $t_{(10)}$ = 2.11, insignificant and whole ROI: r(second run unidentified, intact) = −0.02 ± 0.02; r(second run identified, intact) = 0.07 ± 0.03; $t_{(10)}$ = 2.90, $P < 0.05$]. For the first test, there was no visible trend [150 voxels: r(first run unidentified, intact) = 0.10 ± 0.03; r(first run identified, intact) = 0.10 ± 0.02; $t_{(10)}$ = 1.54, insignificant and whole ROI: r(first run unidentified, intact) = 0.08 ± 0.02; r(first run identified, intact) = 0.07 ± 0.03; $t_{(10)}$ = 0.33, insignificant].

As in the first analysis, the mean beta weights did not significantly differ between the identified and unidentified images, for both runs and both ROIs, whether using the whole ROI or the 150 most active voxels.

The Learned Image Analysis

The previous image-identification analysis suffers from one major drawback: Images that are identified typically have a higher signal level (i.e., are less degraded) than the ones which are not recognized. Thus, the fact that the activation evoked by identified images is more correlated with the intact image may simply stem from a greater physical resemblance between these categories, compared with the highly degraded versions of the image (which are not recognized). While this is a logically

sound argument, the results of the first image-presentation analysis, in which activation evoked by the stimuli in the second run was more correlated with the intact images than on the first run, although they were exactly the same physical stimuli, strongly argue against this possibility.

Still, to account for this potential confounding effect, in the third analysis, presentations of each image were split into 3 groups, such that one regressor corresponded to signal levels that were not identified during the first test run, but following the learning phase were identified during the second run (see Fig. 6). Thus, we compared responses to identical stimuli which were perceived differently. An added value, compared with the image presentation analysis, is that unlike the previous analysis, we select only the specific degraded images that have become recognizable, thereby focusing our analysis on the direct correlates of fast perceptual learning.

We find, similar to the results of the image-presentation analysis, that with learning, the pattern of responses in LOC to the newly recognized images had become more alike the one evoked by the intact stimuli: The activation patterns of the second run were significantly more correlated with the intact-image responses than in the first run [Fig. 7A, left panel; 150 voxels: r(first run, intact) = 0.07 ± 0.03; r(second run, intact) = 0.15 ± 0.02; $t_{(10)} = 4.15$, $P < 0.01$], though both are based on exactly the same images. In contrast, no significant effect of perceptual state-change was found in EVC for any of the comparisons (Fig. 7B).

A complementary view is the correspondence to the pure noise stimulus: Since the analysis focuses on degraded images that were uninterpretable in the first run (e.g., viewed as noise) and have become recognizable in the second run, the correlation of their evoked response with the pure noise image should *decrease* in the second run. Indeed, the pattern of response in LOC to the newly recognized images had become more dissimilar to that evoked by the *noise* stimuli: Correlations with the noise image were lower in the second run than the first run [Fig. 7C, left panel; 150 voxels: r(first run, noise) = 0.00 ± 0.02; r(second run, noise) = -0.05 ± 0.02; $t_{(10)} = 3.03$, $P = 0.0126$]. Similar trends were seen when the voxel representation was based on the whole LOC ROI, although the effects were not statistically significant (Fig. 7C, right panel).

As with the previous 2 analyses, the mean beta weights did not differ significantly between the runs, in both ROIs.

Correlations with the Other Intact Images

It is possible that, following learning, the pattern of activation in LOC for a given degraded object image has changed in a *nonspecific* way, so that it would match *any* object (but see Grill-Spector and Kanwisher 2005). This may be the expected outcome, if the activation pattern in LOC was related to the general recognition of an object (e.g., "animal"), without specific knowledge of its exact nature (e.g., "lion"). If this was indeed the case, the correlation between the response to the degraded images of that specific object and the responses evoked by *other* intact images should also increase following learning. However, in the chosen voxels from LOC, the average correlation between the response to the degraded image and the *other* intact images was significantly *lower* in the second run than in the first run [Fig. 4B, left; r(first run, other intacts) = -0.015 ± 0.004; r(second run, other intacts) = -0.022 ± 0.003; $t_{(10)} = 2.84$, $P < 0.05$]. Note that this is opposite to the *increased*

correlation for the *matching* image. Thus, the possibility that the pattern of activation in LOC simply reflects nonspecific object-like quality is ruled out.

Finally, we repeated the analysis separately for the animate and inanimate image categories, since these 2 distinct categories have unique neural representations (for a review, Martin 2007). For both categories, the correlations between the degraded images and the intact images during the second run were significantly greater than that with all other intact images from the same category (animate: 150 voxels and whole LOC: $P < 0.001$; inanimate: 150 voxels and whole LOC: $P < 0.05$), and in none of the cases was the mean correlation with the other intacts during the second run significantly greater than during the first run. Thus, we can conclude that the patterns became more similar to the specific corresponding intact image and not to a categorical prototype.

Discussion

Our results suggest that the coding in LOC reflects the perceptual level of representation of visual objects, which is strongly modulated by recognition.

Specifically, in LOC, the first analysis indicated that fast learning induces higher correlations with the pattern of activity evoked by the intact stimulus (Fig. 4B). One could posit that the change results solely from an unrelated change in state between the first and second run (such as the general arousal level). However, the results of the second analysis which compared the correlation of identified and unidentified images *within* each run rule this out (see Fig. 5B). The third analysis, which focused on the activation evoked by images in the *learned* range, generally repeated the same results as in the first analysis (namely, greater correlation with the intact images, for the same degraded images, in the second run, than in the first run; see Fig. 7A). This effectively eliminates the concern that the results in the second analysis were solely due to a difference in the degradation level between recognized and unrecognized images: Correlations are computed here for responses evoked by *identical* degradation levels, before and after learning. Finally, the third analysis also revealed that, in LOC, the patterns of activation evoked by images that have become recognizable in the second run have become *less* correlated with those evoked by the *noise image* (Fig. 7C). This mirrors the fact that, following learning, the same images are interpreted as objects, rather than mere noise as in the first run.

Do the activation patterns evoked by the degraded images actually *change* following learning, to better match the patterns evoked by the matching intact images? A plausible alternative may be that the pattern of activation across voxels has actually remained the same, but the average *level* of activation has become more pronounced following learning (see Grill-Spector et al. 2000). Greater overall activation would translate to a higher SNR, leading to less variance in the pattern of response across trials. This will ultimately result in higher correlations with the intact-image response pattern. To check this, we compared the average activation level (e.g., beta weights) in the first and second runs (globally, between identified and unidentified trials of both runs, and between the learned range of the first and second runs). No consistent trend could be found in this direction, and none of the comparisons
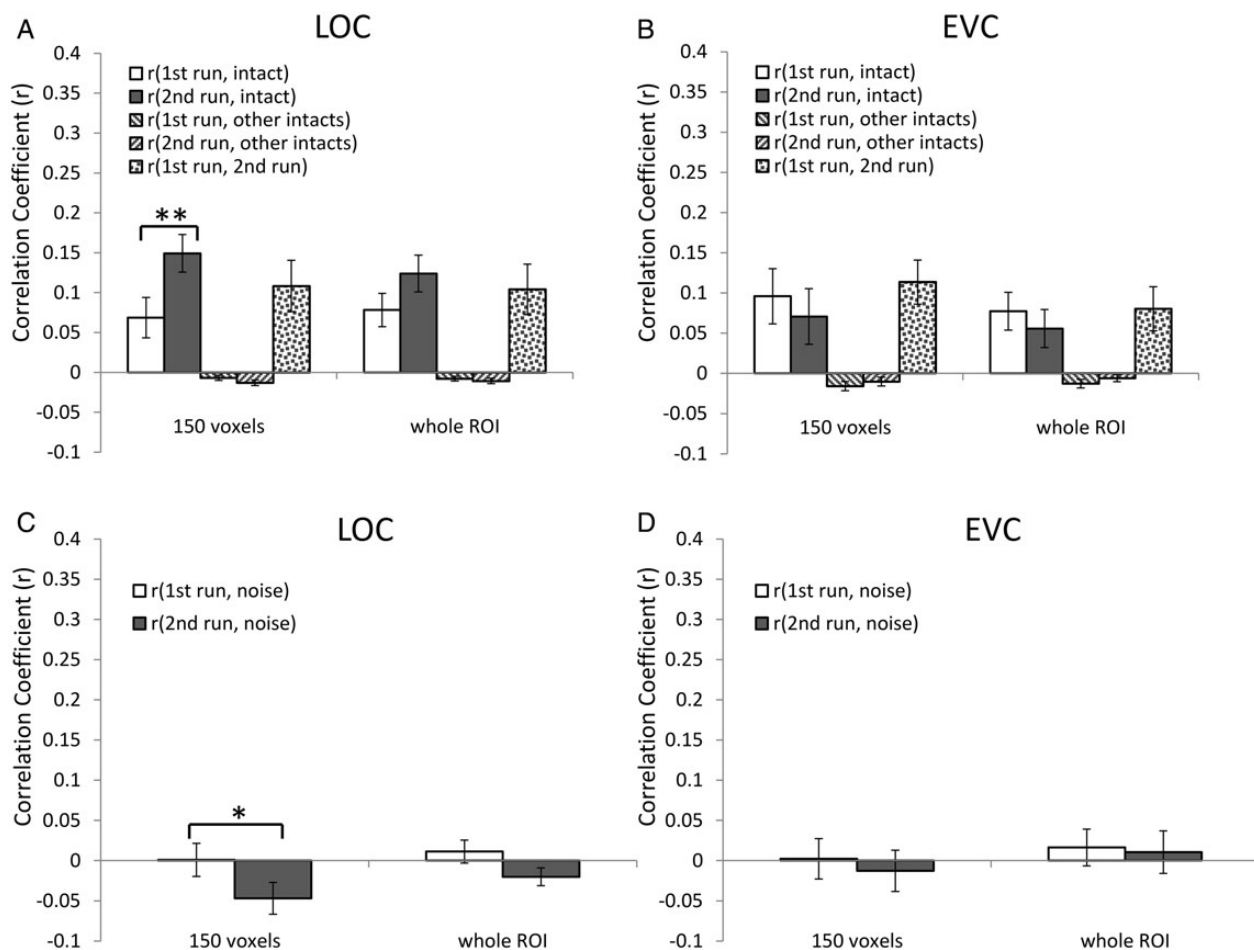
Figure 7. Results of image-perception analysis. (A) Mean correlations in LOC (across subjects and images) for the pattern of responses to the degraded images in the learned range (see Fig. 6) when identified (second run) and unidentified (first run), with the response to the same intact image (white and gray bars, respectively), with the *other* intact images (upward diagonal and downward diagonal bars, respectively), and between the 2 patterns, evoked by the same physical stimuli, in the 2 runs (dotted bar). Left and right panels are for multivoxel patterns based on the most responsive 150 voxels or all voxels within LOC, respectively. This analysis replicates the results shown in Figure 4A, when using data only from the learned range of degraded images, despite the weaker statistical power due to the smaller sample size. Note that in this case, the correlation between the patterns evoked by physically identical stimuli in the 2 runs (dotted bar) is even lower than between very different images of the same object (i.e., highly degraded and intact) that are *recognized* as the same object. (B) The same analysis, applied to EVC. No fast learning effect is evident. (C) Mean correlations of the LOC patterns of activation to the learned range images with the response to the noise image when using 150 voxels (left) and the whole ROI (right). When using the select voxels within LOC, significantly *more negative* correlation (*P < 0.05) was found in the second run, when these degraded images were recognized compared with the first run when the same images were not recognized. (D) EVC correlations with the noise image showed no significant difference between the runs.

yielded a significant change. We conclude that higher activation levels were not the cause of the higher correlations.

Another explanation could be that the response pattern to an identified image becomes more similar to a *nonspecific* "object" template pattern, rather than to the corresponding intact image. If indeed, with learning, such a nonspecific pattern emerges, it should be positively correlated with the activation evoked by *all* the *intact* images, not just the *matching* one. We tested this possibility by calculating the mean correlation between the pattern elicited by a degraded image and the patterns elicited by the *other* 7 intact images. A comparison between the first and second runs yielded a significant effect in LOC, but in the *opposite* direction: the mean correlation to the other intact images was significantly *lower* during the second run than in first run. This was also the case when analysis was restricted to the learned range (see details in Results section).

Recognition is therefore associated with greater correlation between the evoked response to the degraded image and its

corresponding intact image. Importantly, it is also associated with a *diminished* correlation with the patterns evoked by the *other* intact images. This makes perfect sense: If the different intact patterns are not correlated between themselves (as indeed is the case, the mean correlation between intact images was significantly lower than zero; $r = -0.12$, $P < 0.001$), a higher correlation with the match would necessarily lead to lower correlation with the other intact images.

Finally, classification of the images ("recognized" or not) was based on the subjects' *subjective* binary report (through button presses). Subjective reports can potentially differ from some forms of objective behavior (Hesselmann et al. 2011). But this only strengthens our results, since any mismatch between the subjective report and the actual image identity (i.e., false recognition) would be expected to *diminish* the observed effect, not accentuate it. Furthermore, in our case, pilot studies outside the scanner, in which subjects had to *verbally* report the recognized object, suggest that the subjective

response closely matched recognition performance. Still, a task requiring vocal classification of the degraded images is obviously preferable, if technically possible.

In the image-identification analysis of LOC, the activation evoked by identified images was significantly more correlated with that evoked by the intact images, than the unidentified images. But this was the case only for the second test run. Our interpretation of this is that since subjects become familiar with the images during the learning phase, when they identify an image during the second run, they can retrieve more details than are actually visible in the degraded image. This results in a richer perceptual experience than during the first run, which is mirrored by the greater correlation with the corresponding intact images in the second run.

In EVC, we found that with regard to the second run, correlations between patterns evoked by the intact images and those evoked by degraded images were lower for *uni*dentified images than for identified images (see Fig. 5*C*). This can be understood by noting that learning shifts the recognition curve to the left (see Fig. 2*B*), so that on average, unidentified images will be of higher degraded levels in the second run compared with the first. This will obviously lead to a lower correlation with the intact object images in run 2 compared with run 1. The same pattern is true for LOC, though to a lesser degree.

What leads to the increased identification rate during the second run, and what could be the source of the changes in the brain activation patterns seen in LOC? An initial guess may be that subjects learned to direct their gaze at the more informative regions of the images during the second run. But this is unlikely, since the images were shown for very brief times (200 ms), effectively barring the possibility for a saccade during image presentation. We think that the experience gained during the first run provides the subjects with 3 main advantages during the second run.

First, subjects may learn *where to direct their attention*. Although they do not have enough time to fixate their eyes on an informative region, they may manage to figuratively "fixate" their attention on such a region (see Egeth and Yantis 1997). These regions may be specific for each image, or they could comprise an area statistically informative over all images. If the distribution of attention fixations is more similar to the distribution when viewing the intact image, this could result in the LOC activations becoming more similar too. However, this possibility seems unlikely, since we would expect patterns in EVC to change according to such fixations as well and become more similar to responses to the intact images.

Secondly, after the first run, which clearly revealed the identity of the images in the experiment, subjects could effectively narrow their prior expectation from all possible object images in the world, down to only 8 relevant images. Clearly, this makes identification a much easier task. Specifically, it has been suggested that feedback signals from prefrontal cortex provide possible "interpretations" of presented stimuli to LOC, thus guiding the bottom-up activity to correspond to reasonable percepts (Bar et al. 2006; Cheung and Bar 2012). When the number of possible stimulus interpretations decreases, the top-down signals should become more specific and precise, resulting in patterns of activation in LOC that correspond more strongly to the correct object. In addition, following exposure to the specific images in the first run, it is plausible that during the second run, subjects were able to retrieve these images from memory and imagine them. Mental imagery has been

shown to result in patterns of activation in LOC that are similar to those present when viewing a stimulus (Stokes et al. 2009; Reddy et al. 2010; Lee et al. 2012). Some form of implicit imagery (Albright 2012) could thus explain the higher correlations during the second run.

Thirdly, presentation of the images at high signal (e.g., low degradation) levels possibly allowed gaining insights into the way noise interacts with the specific images to obscure their recognition. By noticing which image-specific contours are most robust to noise, subjects might have learned to better recognize the image at lower signal levels. This strategy would have become even more dominant had the order of presentation been reversed (from high signal levels to lower ones), leading to a hysteresis in recognition performance.

Alternatively (though not necessarily mutually exclusive), a fourth possibility is that the behavioral change may reflect low-level neural changes, such as reweighting of specific connections' strength from EVC to LOC. Such modifications in the synaptic weighing could potentially lead to a change in the firing patterns of the receiving region (LOC) without concurrent changes in the pattern of activity (at the voxel scale) in the transmitting region (EVC), as was in our case. As a result, activation patterns in LOC should become more similar to intact-image activation patterns, thus enabling recognition.

If learning is based on a change in bottom-up activity (e.g., from earlier visual areas), the activation changes should take effect in the *initial* response to the stimuli. On the other hand, if feedback signals are responsible for the change in the pattern of activation in LOC, we might expect the initial pattern of responses (in the first ~100–150 ms) to remain the same. The activity pattern should become similar to the intact-image pattern only when the feedback signals "kick in," later in time. Unfortunately, the slow and temporally smoothed BOLD signal is insufficient to allow differentiation between these 2 temporal dynamics. Future experiments incorporating methods with finer temporal resolution (e.g., transcranial magnetic stimulation and magnetoencephalography) may possibly enable us to better understand the mechanisms by which learning affects the activation patterns.

How do our results correspond to earlier studies? Using a masking paradigm with briefly presented object images, Grill-Spector et al. (2000) trained subjects to identify such images for several daily sessions. This resulted in higher recognition performance and significantly higher activation levels (i.e., beta weights) in LOC, specifically for images that were initially unrecognized and have become identifiable with training. In contrast, we did not find higher activation levels in LOC during the second test run, despite the higher recognition levels. The cause of this difference between the results may lie in the extensive 5–7 days of training subjects underwent in the masking study. In our study, learning was much faster: a brief 3-min learning phase, in addition to the first test run, was sufficient to lead to recognition of many degraded images that were previously unidentifiable. Possibly, an increase in the average activation levels requires longer time periods, or greater numbers of repetitions of the same stimulus. Indeed, in a similar experiment using a "one-shot" fast learning procedure, Hsieh et al. (2010) found that there was no significant change in the mean voxel activation levels following learning, despite a clear behavioral improvement.

Similar to our experimental design, Hsieh et al. (2010) had subjects learn to recognize Mooney figures, following a brief exposure to their original images. They measured the degree to which the patterns of activation evoked by the Mooney figures matched those elicited by the intact images, before and after learning. Unlike us, they report changes in the activation patterns elicited by the Mooney figures, such that they better match the intact images, in *early* visual cortical areas (but only in voxels representing the central visual field; i.e., the foveal confluence). They suggested that these changes reflect a top-down signal from higher cortical representations (although they do not report similar changes in LOC during their event-related fMRI study). One possibility why the 2 studies show opposite effects may be that Mooney figures have clear contours, and recognition is mainly hampered by the problem of scene segmentation (e.g., assigning edges to be part of the figure or background). Top-down information may be very useful for this purpose, and indeed in monkeys, the responses of neurons in early visual areas (V1/V2) are often stronger if a specific edge is interpreted as being part of the figure, than when it perceived as part of the background (Lamme 1995; Zhang and von der Heydt 2010). In our case, some of the contours of the original images are missing from the degraded images and need to be retrieved from memory—a process which may depend on the activity of higher level areas. It is also possible that the different functional definitions of the early visual areas ROIs (e.g., voxel selection based on a different GLM contrast) may be the cause of the contradictory findings.

To summarize, using multiple (though not independent) analyses, we demonstrate here that recognition of previously unidentified visual images is associated with a *predicted* change in the pattern of activation in LOC. The change is not in the overall level of activation but rather in a greater similarity to the pattern of activation elicited by the matching intact stimulus. Our results add to a body of previous studies, suggesting that the pattern of activity in LOC, observed at a coarse scale using fMRI, reflects *perceptual experience* (e.g., object recognition) rather than merely the *physical attributes* of the stimulus.

Funding

Notes

We thank Tanya Orlov, Yuval Porat, and Tal Seidel for helpful suggestions on earlier drafts of this manuscript. *Conflict of Interest*: None declared.

References

Albright TD. 2012. On the perception of probable things: neural substrates of associative memory, imagery, and perception. Neuron. 74:227–245.

Andrews TJ, Schluppeck D. 2004. Neural responses to Mooney images reveal a modular representation of faces in human visual cortex. NeuroImage. 21:91–98.

Andrews TJ, Schluppeck D, Homfray D, Matthews P, Blakemore C. 2002. Activity in the fusiform gyrus predicts conscious perception of Rubin's vase–face illusion. NeuroImage. 17:890–901.

Avidan G, Harel M, Hendler T, Ben-Bashat D, Zohary E, Malach R. 2002. Contrast sensitivity in human visual areas and its relationship to object recognition. J Neurophysiol. 87:3102–3116.

Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, Hämäläinen M, Marinkovic K, Schacter D, Rosen B. 2006. Top-down facilitation of visual recognition. Proc Natl Acad Sci USA. 103:449–454.

Bonneh YS, Cooperman A, Sagi D. 2001. Motion-induced blindness in normal observers. Nature. 411:798–801.

Cheung OS, Bar M. 2012. Visual prediction and perceptual expertise. Int J Psychophysiol. 83:156–163.

de Beeck HPO, Torfs K, Wagemans J. 2008. Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. J Neurosci. 28:10111–10123.

Dolan RJ, Fink GR, Rolls E, Booth M, Holmes A, Frackowiak RSJ, Friston KJ. 1997. How the brain learns to see objects and faces in an impoverished context. Nature. 389:596–599.

Donner TH, Sagi D, Bonneh YS, Heeger DJ. 2008. Opposite neural signatures of motion-induced blindness in human dorsal and ventral visual cortex. J Neurosci. 28:10298–10310.

Drucker DM, Aguirre GK. 2009. Different spatial scales of shape similarity representation in lateral and ventral LOC. Cereb Cortex. 19:2269–2280.

Eger E, Ashburner J, Haynes J-D, Dolan RJ, Rees G. 2008. fMRI activity patterns in human LOC carry information about object exemplars within category. J Cogn Neurosci. 20:356–370.

Egeth HE, Yantis S. 1997. Visual attention: control, representation, and time course. Ann Rev Psychol. 48:269–297.

Friston K, Fletcher P, Josephs O, Holmes A, Rugg M, Turner R. 1998. Event-related fMRI: characterizing differential responses. NeuroImage. 7:30–40.

Gorlin S, Meng M, Sharma J, Sugihara H, Sur M, Sinha P. 2012. Imaging prior information in the brain. Proc Natl Acad Sci USA. 109:7935–7940.

Grill-Spector K, Kanwisher N. 2005. Visual recognition: as soon as you know it is there, you know what it is. Psychol Sci. 16:152–160.

Grill-Spector K, Kourtzi Z, Kanwisher N. 2001. The lateral occipital complex and its role in object recognition. Vis Res. 41: 1409–1422.

Grill-Spector K, Kushnir T, Hendler T, Malach R. 2000. The dynamics of object-selective activation correlate with recognition performance in humans. Nat Neurosci. 3:837–843.

Grill-Spector K, Malach R. 2004. The human visual cortex. Ann Rev Neurosci. 27:649–677.

Hasson U, Hendler T, Bashat DB, Malach R. 2001. Vase or face? A neural correlate of shape-selective grouping processes in the human brain. J Cogn Neurosci. 13:744–753.

Haushofer J, Livingstone MS, Kanwisher N. 2008. Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. PLoS Biol. 6:e187.

Haynes JD, Rees G. 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci. 8:686–691.

Hesselmann G, Hebart M, Malach R. 2011. Differential BOLD activity associated with subjective and objective reports during "blindsight" in normal observers. J Neurosci. 31:12936–12944.

Hesselmann G, Kell CA, Eger E, Kleinschmidt A. 2008. Spontaneous local variations in ongoing neural activity bias perceptual decisions. Proc Natl Acad Sci USA. 105:10984–10989.

Hsieh PJ, Vul E, Kanwisher N. 2010. Recognition alters the spatial pattern of fMRI activation in early retinotopic cortex. J Neurophysiol. 103:1501–1507.

Konkle T, Oliva A. 2012. A real-world size organization of object responses in occipitotemporal cortex. Neuron. 74:1114–1124.

Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI. 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci. 12:535–540.

Lamme VA. 1995. The neurophysiology of figure-ground segregation in primary visual cortex. J Neurosci. 15:1605–1615.

Lee S-H, Kravitz DJ, Baker CI. 2012. Disentangling visual imagery and perception of real-world objects. NeuroImage. 59:4064–4073.

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB. 1995. Object-related

activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci USA. 92:8135–8139.

Martin A. 2007. The representation of object concepts in the brain. Annu Rev Psychol. 58:25–45.

Norman KA, Polyn SM, Detre GJ, Haxby JV. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci. 10:424–430.

Rainer G, Lee H, Logothetis NK. 2004. The effect of learning on the function of monkey extrastriate visual cortex. PLoS Biol. 2:e44

Rainer G, Miller EK. 2000. Effects of visual experience on the representation of objects in the prefrontal cortex. Neuron. 27:179–189.

Reddy L, Tsuchiya N, Serre T. 2010. Reading the mind's eye: decoding category information during mental imagery. NeuroImage. 50:818–825.

Sawamura H, Georgieva S, Vogels R, Vanduffel W, Orban GA. 2005. Using functional magnetic resonance imaging to assess adaptation and size invariance of shape processing by humans and monkeys. J Neurosci. 25:4294–4306.

Sterzer P, Kleinschmidt A, Rees G. 2009. The neural bases of multi-stable perception. Trends Cogn Scie. 13:310–318.

Stokes M, Thompson R, Cusack R, Duncan J. 2009. Top-down activation of shape-specific population codes in visual cortex during mental imagery. J Neurosci. 29:1565–1572.

Tong F, Meng M, Blake R. 2006. Neural bases of binocular rivalry. Trends Cogn Sci. 10:502–511.

Tong F, Nakayama K, Vaughan JT, Kanwisher N. 1998. Binocular rivalry and visual awareness in human extrastriate cortex. Neuron. 21:753–759.

Williams MA, Dang S, Kanwisher NG. 2007. Only some spatial patterns of fMRI response are read out in task performance. Nat Neurosci. 10:685–686.

Worsley KJ, Friston KJ. 1995. Analysis of fMRI time-series revisited—again. NeuroImage. 2:173–181.

Zhang NR, von der Heydt R. 2010. Analysis of the context integration mechanisms underlying figure-ground organization in the visual cortex. J Neurosci. 30:6482–6496.