Published in final edited form as:
*Neuroimage*. 2010 April 1; 50(2): 818–825. doi:10.1016/j.neuroimage.2009.11.084.

# Reading the mind's eye: decoding category information during mental imagery

**Leila Reddy**[1,2,3,*], **Naotsugu Tsuchiya**[4,5], and **Thomas Serre**[6]

[1] Université de Toulouse; UPS; Centre de Recherche Cerveau et Cognition; France

[2] CNRS; CerCo; Toulouse, France

[3] Computation & Neural Systems, California Institute of Technology, Pasadena, CA

[4] Humanities and Social Sciences, California Institute of Technology, Pasadena, CA

[5] Brain Science Institute, Tamagawa University, Tokyo, Japan

[6] Mc Govern Institute, MIT, Cambridge, MA

## Abstract

Category information for visually presented objects can be read out from multi-voxel patterns of fMRI activity in ventral temporal cortex. What is the nature and reliability of these patterns in the absence of any bottom-up visual input, for example, during visual imagery? Here, we first ask how well category information can be decoded for imagined objects, and then compare the representations evoked during imagery and actual viewing. In an fMRI study, four object categories were either visually presented to subjects, or imagined by them. Using pattern classification techniques we could reliably decode category information (including for non-special categories, i.e., food and tools) from ventral temporal cortex in both conditions, but only during actual viewing from retinotopic areas. Interestingly, in temporal cortex when the classifier was trained on the viewed condition and tested on the imagined condition, or vice-versa, classification performance was comparable to within the imagined condition. The above results held even when we did not use information in the specialized category-selective areas. Thus, the patterns of representation during imagery and actual viewing are in fact surprisingly similar to each other. Consistent with this observation, the maps of "diagnostic voxels" (i.e., the classifier weights) for the perception and imagery classifiers were more similar in ventral temporal cortex than in retinotopic cortex. These results suggest that in the absence of any bottom-up input cortical back projections can selectively re-activate specific patterns of neural activity.

### Keywords

imagery; perception; fMRI; multi-voxel pattern analysis; occipito-temporal cortex; object recognition

Corresponding author: Leila Reddy, Centre de Recherche Cerveau et Cognition, Faculte de Medecine, Rangueil, 31062 Toulouse CEDEX, FRANCE, Leila.reddy@gmail.com.

## Introduction

The contents of visual perception can be decoded from fMRI activation patterns in visual cortex. In retinotopic regions, an ideal observer can predict features of a viewed stimulus (e.g., the orientation of a grating) (Kamitani and Tong, 2005), the attentional state of the observer (Kamitani and Tong, 2005), properties of a stimulus that was not consciously perceived (Haynes and Rees, 2005), and the identity of viewed natural images (Kay et al., 2008). In higher-tier areas, object-category information can be gleaned from groups of category-selective voxels, as well as from more distributed representations (Carlson et al., 2003; Cox and Savoy, 2003; Haxby et al., 2001; Norman et al., 2006; Reddy and Kanwisher, 2007; Spiridon and Kanwisher, 2002). These regions have also been implicated in processing driven solely by top-down control, in the absence of bottom-up input – i.e., during mental imagery (Finke, 1985; Ishai and Sagi, 1995). Accordingly, both visual perception and imagery activate earlier areas, particularly when subjects judge fine details of a stimulus (Ganis et al., 2004; Kosslyn et al., 1995). Area MT is activated by imagery of moving stimuli (Goebel et al., 1998), and in ventral-temporal cortex, imagery of preferred categories (faces and houses) activates the corresponding category-selective regions (O'Craven and Kanwisher, 2000). More recently, lateral occipital complex was shown to be involved when subjects viewed and imagined the letters ‘X’ and ‘O’ (Stokes et al., 2009).

This observed spatial overlap during imagery and perception does not necessarily imply that the corresponding representations are qualitatively the same. Indeed, our subjective experience of imagining something and seeing it are undoubtedly different. Thus, the neural representations of visual perception and imagery might be expected to be substantially different from each other – after all, the former process is driven by bottom-up input, while the latter is initiated by top-down signals. Accordingly, in single neurons, responses during imagery are found to be significantly longer than during perception, with latencies differing by approximately 100 ms, and peak neuronal response times by as much as 800 ms (Kreiman et al., 2000). fMRI studies have also shown differences in activation during the two conditions. For example, reliable deactivation in auditory cortex has been observed during visual imagery, but not during visual perception (Amedi et al., 2005), the overlap between activation patterns during imagery and perception is much larger in frontal regions, than in ventral-temporal cortex (Ganis et al., 2004), and finally, even in category-selective regions, imagery activates fewer voxels at a given statistical threshold (Ishai et al., 2000; O'Craven and Kanwisher, 2000), with a lower overall response.

Here we directly test whether imagery and perception of object categories share common visual representations. In an fMRI study, 10 participants viewed or imagined four object categories (Figure 1). Linear support-vector machines (SVM) were trained on fMRI activation patterns in a distributed set of object-responsive (OR) voxels in the perceptual (P) and imagery (I) conditions. First, we tested whether these P and I classifiers could decode the contents of perception and imagery respectively. Second, to determine whether the two states share common representations, each classifier was tested on the *other* condition: i.e., the P-classifier decoded the category of imagined objects and vice versa. The logic here is as follows: if the representations are largely independent of each other, performance in decoding category information from the other state should be no different from chance. Conversely, if the two processes share common representations performance should be above-chance in the test condition.

A recent study by Stokes et al. (2009) used a similar logic in testing imagery and perceptual representations in LOC. In that study subjects were tested with two stimuli (the letters X and O) that were presented at fixation. In contrast to this relatively simpler classification between two elementary shapes, the present study probes a greater degree of abstraction in visual

representations by implementing a 4-way classification of a larger and diverse set of colored natural photographs (see Methods).

One version of these results has previously been presented in abstract form (Reddy et al., Society for Neuroscience, (Washington D.C), 2008).

## Methods

### Subjects

Ten healthy subjects participated in the fMRI study. All subjects gave signed consent and had normal, or corrected-to-normal vision. The study was approved by the Caltech IRB.

### Experimental Design

Each subject participated in 7 or 8 fMRI scanning runs. Each run consisted of 5 fixation blocks, 8 blocks of a visual presentation condition, and 8 blocks of a visual imagery condition (see Figure S1 for an illustration of the design of an example run). Each block lasted 16s accounting for a total scan time per run of 5.6 minutes. During the visual presentation condition, subjects were visually presented with 4 categories of objects in different blocks. The four categories were food (common fruits and vegetables), tools, famous faces and famous buildings. Each category was presented twice per run, in separate blocks. Each block consisted of 4 trials – one trial per category exemplar (Figure 1). 4 exemplars per category were used in one half of the runs, and another 4 exemplars per category were used in the other half of the fMRI runs. Thus, in total we had 8 exemplars per category (Figure S2). Each trial consisted of 2s of visual presentation and 2s for task response. The trial order was randomized within the block.

During the visual imagery condition the block design was similar to that of the visual presentation condition. On each 4-second trial of a block, headphones were used to give subjects the name of which category exemplar they were to imagine (e.g., in a “food” block the instructions could have been “apple”, “pear”, “grapes”, “tomato”). As in the visual presentation condition 4 exemplars were used in one half of the runs and another 4 exemplars were used in the other half of the runs. Although only well-known exemplars were used for all categories, we also made sure that the subjects were familiar with all the category exemplars for the visual imagery condition. Thus, the night before the scan session, subjects were provided with the set of 32 images (4 categories × 8 exemplars) that would be used during the visual presentation blocks, and their associated names, and asked to familiarize themselves with the stimuli. Additionally, 15 minutes prior to the scan, subjects were again asked to examine the stimuli. No subjects reported being unfamiliar with any of the stimuli, as was expected since only common and highly familiar category exemplars were used. Before the scan session, subjects were instructed to try to generate vivid and detailed mental images as similar as possible to the corresponding images seen in the visual presentation condition.

Note that, as with most previous imagery studies, the critical task for our subjects was to either visually examine the stimuli presented in the visual perception condition, or to create vivid mental images during the imagery condition. Additionally, in order to make sure that subjects were attending to the images, they were asked to perform a secondary task during both conditions, and press a button if the color of two successive images was the same (i.e., a one-back task; see legend of Figure S2 for further details of the subjects' tasks). For the famous face category we had examples of African-American and Caucasian celebrities and subjects performed the task on the color of these faces. Subjects performed the color-task for the entire stream of 16 images (4 images × 4 categories) in each “perception” and “imagery” condition (i.e., both within and across categories; see Figure S1 for a depiction of these blocks). All images were presented at fixation and subtended approximately 6 degrees of visual angle.

Using auditory instructions, subjects were asked to close their eyes prior to the visual imagery blocks, and to open them prior to the start of the visual presentation blocks (we confirmed that subjects followed these instructions with online monitoring with an ASL eyetracker). Because of these instructions, the visual presentation (P) and visual imagery (I) conditions were presented in sequences of 4 blocks, separated by a fixation block (e.g., fixation, P-face, P-building, P-tool, P-food, fixation, I-tool, I-building, I-face, I-food, fixation, …). Within each P or I sequence the block order was randomized. The order of the sequences followed an ABBA design in each run. Each run started with a P sequence.

It should be noted that, as shown in Figure S1, our perception and imagery conditions were presented in distinct blocks, separated by 16 second long fixation intervals. Additionally, most of the time, perception of category X was followed by imagery of another category Y (with the 16 second fixation interval in between). This design minimizes any priming effects between the perception and imagery conditions; indeed, any priming effect of perception on imagery would only have been detrimental to decoding performance (at least in the majority of blocks where perception of X was followed by imagery of Y). Additionally the order of imagery and perception blocks was counter-balanced on each run.

### Regions of Interest (ROIs)

In separate localizer runs subjects were presented with blocks of faces, scenes, objects and scrambled images. Based on the data obtained in these localizer runs a set of object responsive voxels (OR) was defined. This OR ROI was the set of *distributed* voxels in the ventral temporal cortex that were more strongly activated to faces, objects, or scenes compared to scrambled images ($p < 10^{-4}$, uncorrected). OR thus included the FFA, PPA and LOC, as well as other object responsive voxels in ventral temporal cortex. See Figure S3a for a map of OR. In control analyses we also considered OR with the exclusion of the FFA and PPA and refer to this ROI as OR-FFA&PPA. The FFA was defined as the set of contiguous voxels in the fusiform gyrus that showed significantly stronger activation ($p < 10^{-4}$, uncorrected) to faces than to other objects. The PPA was defined as the set of voxels in the parahippocampal gyrus that showed stronger activation to scenes versus objects ($p < 10^{-4}$, uncorrected).

### Retinotopy

Meridian mapping was performed by alternately presenting a horizontal or vertical flickering checkerboard pattern for 18 seconds at each location. The horizontal and vertical meridians were stimulated 8 times each per run (total run time = 288s). Two such runs were acquired per subject. The average retinotopic ROI across subjects is shown in Figure S3b.

### fMRI data acquisition and analysis

fMRI data was collected on a 3T Siemens scanner (gradient echo pulse sequence, TR = 2s, TE = 30 ms, 32 slices with a 8-channel head coil, slice thickness = 3 mm, in-plane voxel dimensions = 3mm × 3mm) at the Caltech Brain Imaging Center. High- resolution anatomical images were also acquired per subject. Data analysis was performed with FreeSurfer and FS-FAST (http://surfer.nmr.mgh.harvard.edu), fROI (http://froi.sourceforge.net) and custom Matlab scripts. Before statistical analysis, all images were motion corrected (using AFNI with standard parameters), intensity normalized and smoothed with a 5 mm full width at half maximum Gaussian kernel. (Note that to check the effect of smoothing on the final results several different kernel sizes were also applied (Figure S4)). For defining the ROIs average signal intensity maps were then computed for each voxel using FS-FAST. For each subject, we created a design matrix that included the fixation condition and the four conditions of the localizer runs. The predictor for each stimulus condition (0 or 1 at each time point) was convolved with a gamma function, and the general linear model was used to compute the response of each voxel in each condition. This response was expressed as the percent signal change, i.e., the response in each

condition minus the response in the fixation condition, normalized by the mean signal in each voxel.

### Multivariate analysis

Preprocessing for the multivariate analysis was conducted using the Princeton Multi-Voxel Pattern Analysis (MVPA) toolbox (http://www.csbmb.princeton.edu/mvpa) as well as custom Matlab functions. Following the MVPA processing stream, after motion correction and smoothing, for each subject, the BOLD signal was detrended by fitting a second-degree polynomial for each voxel and each run. After detrending, a z-score transform was applied to the data (for each voxel in each run). Finally to correct for the hemodynamic lag, the regressor for each presentation condition (i.e., the matrix of values that denotes at each timepoint which condition was active) was convolved with a gamma hemodynamic response function. The regressors matrix was then used in the classification procedure as category labels.

The multi-class classification results reported here are based on the Support Vector Machine (SVM) classification algorithm and the machine learning Spider toolbox developed at the Max Planck Institute (http://www.kyb.mpg.de/bs/people/spider). In all experiments we used a linear kernel and the one-versus-all multi-class classification scheme. Because of the small number of examples available for training and testing we did not attempt to optimize the ‘C’ constant (default value ‘C=Inf’). In a post-hoc analysis, we nonetheless verified that the performance obtained for the resulting classifier remained robust to the exact parameter value. Very similar classification results were obtained using non-linear kernels (linear vs. polynomial vs. Gaussian), other classification schemes (one-versus-all vs. all-pairs) and other classification algorithms (SVM vs. boosting vs. regularized least-square). Using a leave-one-run-out procedure, we trained classifiers on N-1 runs and computed the mean classification performance on the remaining Nth run for each subject. Mean performance values across subjects are reported here. For further details see the supplemental information section.

### Analysis of SVM weight maps

Note that the Support Vector Machine (SVM) analysis was conducted individually for each subject in his or her respective ROIs and the performance values across subjects were then averaged (Figure 2). However, to plot the average weights of the SVM analyses in OR across subjects (Figure S7), each subject's brain was aligned to the FreeSurfer ‘fsaverage’ brain. FreeSurfer was first used to reconstruct the original surface for each participant from the high-resolution anatomical scan. Individual brains were then aligned to each other in FreeSurfer by spatially normalizing the cortical surfaces to a spherical surface template using an automated procedure to align the major sulci and gyri (Fischl et al., 1999). For each subject, a map of SVM weights was computed by taking the z-score across voxels of the weight maps per run from each leave-one-run-out procedure. The average of these maps across runs and then subjects was computed and overlaid on the average brain. To compute the correlation values shown in Figure 4 we calculated the z-score of the weight maps from each leave-one-run-out procedure in each subjects' functional space. These weight maps were then averaged across runs and correlations between the weight maps for the perception and imagery classifiers were computed for all pairs of categories for each subject. The results were then averaged across subjects.

### Meridian mapping

Meridian mapping analysis was performed on the reconstructed cortical surface for each subject by contrasting the horizontal and vertical stimulation periods to define the borders between visual areas. Areas V1 and V2 were included in the “Retinotopic Voxels ROI” described in the Results section.

### Statistical tests

All the ANOVAs reported in this study are repeated measures ANOVAs. All post-hoc tests were Bonferroni corrected for multiple comparisons.

## Results

Participants were tested in two fMRI experimental conditions (Figure 1). In the visual presentation (P) condition they viewed different exemplars of four categories of stimuli (food (common fruits or vegetables), tools, famous faces and famous buildings) in separate blocks. In the visual imagery (I) condition they were given auditory instructions with the names of these exemplars and asked to imagine them. For each category four exemplars were used in one half of the fMRI runs, and another four exemplars were used in the second half of runs. As mentioned in the Methods section, prior to the fMRI scans participants were asked to familiarize themselves with all the stimuli and their corresponding names so that they could generate mental images of these stimuli in the I condition. The average activation during the perception and imagery conditions across subjects is shown in Figure S5. Consistent with previous studies, the imagery condition evoked activation in smaller clusters compared to the perception condition (Ishai et al., 2002).

### Classification performance during Perception and Imagery in OR

For each subject we defined a set of object responsive (OR) voxels in ventral temporal cortex that responded more strongly to images of faces, scenes or random objects compared with scrambled images. The multivariate pattern of responses in the distributed set of object responsive voxels in ventral temporal cortex has previously been shown to provide information about object category (Haxby et al., 2001; Reddy and Kanwisher, 2007; Spiridon and Kanwisher, 2002). Consistent with these studies a multivariate analysis of the responses in OR allowed us to read out category information during the visual presentation condition. Using a leave-one-run-out procedure, a linear support vector machine (SVM) was trained and tested on the OR fMRI activity patterns corresponding to the four object categories in the P condition. The performance of this classifier in OR is shown in Figure 2A. The top-left confusion matrix in Figure 2A shows the probability with which an input pattern (along the rows) was classified as each of the 4 alternative choices (along the columns). The higher probabilities along the diagonal and the lower off-diagonal values indicate successful classification for all categories. For the P-P classification test (i.e., trained and tested on the P condition), average performance was 67% (chance performance: 25%).

Having obtained above-chance classification performance in the P condition, we next asked whether category information could also be read out when participants were imagining the objects, in the absence of any visual input. To address this question, a classifier was trained and tested on activation patterns associated with the mental imagery conditions. As shown in the confusion matrix for this I/I classification (Figure 2A), above chance performance was obtained across all categories (50%, with chance at 25%). Thus, regions that carry information about perceived object category also seem to contribute to the representation of these categories in the absence of bottom-up visual inputs.

The successful performance of the P and I classifiers on the P/P and I/I classification tests, allowed us to next address the main question of this study – whether viewing an object and imagining it evoked similar representations in OR. To this end, we tested the P classifier on the I condition, and the I classifier on the P condition (i.e., cross-generalization, across conditions). Above-chance classification performance in these cases would indicate that the two representations share common features that allow the classifiers to generalize from one condition to the other. As shown in Figure 2A (bottom left and right matrices respectively),

classification performance was on average 47% and 52% (with chance at 25%). Note that similar classification performance was observed when the set of voxels activated during mental imagery was considered as the ROI (Figure S8).

To test how classification performance in OR depended on category and classification test (i.e., P/P, I/I, I/P, P/I), a 2-way repeated measures ANOVA of category X classification test was performed. The ANOVA revealed a significant main effect of category ($F(3,27) = 10.27$; $p<.001$), a significant main effect of classification test ($F(3,27) = 11.51$; $p < .0001$), and a significant interaction effect ($F(9,81)=3.93$; $p<0.005$). Post-hoc tests, Bonferroni corrected for multiple comparisons, revealed that classification performance for faces and buildings was significantly higher than for food and tools. A 2-way discrimination of tools versus food also revealed above chance performance for all 4 classifier tests in OR. The results of this discrimination performance can be seen in the lower right portion of the confusion matrices in Figure 2a, where we directly see how often foods and tools were correctly predicted versus how often foods were confused for tools and vice versa (incorrect predictions). A 2-way ANOVA of the performance values (correct prediction vs. incorrect prediction) x classifier test revealed a significant main effect of performance ($F(1,72)=42.14$; $p< 0.0001$) but no significant effect of classifier test, nor a significant interaction effect. A post-hoc Bonferroni-corrected test revealed that the performance on correct predictions was significantly larger than on incorrect predictions.

The successful performance obtained in OR was not solely driven by face and scene selective voxels in the FFA and PPA respectively. Similar performance values were also obtained when the FFA and PPA were removed from OR (Figure 2B): 65% for P/P, 47% for I/I, 44% for P/I and 48% for I/P. Similar to the results in Figure 2A, a 2-way ANOVA in this OR-FFA&PPA ROI of category x classification type revealed significant main effects of category ($F(3,27) =7.48$; $p<0.001$), classification type ($F(3,27)=10.51$; $p<0.0005$), and a significant interaction effect ($F(9,81)=2.95$; $p< 0.005$). Post-hoc tests indicated a category advantage in the order faces>buildings=tools>food.

In terms of the type of classification performed, for both the ROIs considered in Figure 2A and B, the post-hoc tests revealed that performance for P/P classification was significantly higher than for the other three types. Importantly, the post-hoc test showed no significant difference between the I/I classification and both the P/I and I/P classification tests. In other words, classification performance across the P and I conditions was just as good as performance within the I condition. Performance in the I/I classification test serves as an upper bound for the expected performance of the P/I and I/P classifications – this is because classification within each condition must theoretically be better than, or just as good as, classification across conditions. Thus, the finding that across-condition classification was not significantly different from classification within the I condition indicates that, overall, the activation patterns obtained on perception and imagery runs are at least as alike as patterns obtained on different runs of visual imagery in both OR and OR-FFA&PPA.

Do these observed results rely on the actual pattern of individual voxel activations in the ROIs, or is the relevant information provided equally well by some global property of each ROI, such as the mean response? To address this question, we performed two tests: first, we scrambled the voxel order of the test data relative to the training dataset – this procedure amounts to keeping the mean BOLD response in each ROI constant across training and test, but removes information carried in the multi-voxel pattern. Second, we shuffled the labels associated with each category in the training data, thus removing any consistent category-specific information in the activation patterns. Classification performance in these scrambled controls, based on 50 shuffles of the labels and voxel order, is shown in Supplementary Figure 6. The ability to decode category information was severely reduced in the scrambled controls indicating that

the successful classification in Figure 2A and B relied on the fine-scale pattern of voxel activations in the fMRI response. In OR, a 2-way repeated measures ANOVA of scrambling type (intact (original) ROI, scrambled voxels or shuffled labels) X classifier test revealed significant main effects of scrambling ($F(2,18) = 88.3$; $p < 0.0001$), and classification tests ($F(3,27) = 10.29$; $p < 0.005$). Post-hoc tests revealed that performance of the P/P classification was significantly larger than the other three, and performance in the intact ROI was larger than in both scrambled controls. The interaction effect of the ANOVA was also significant ($F(6,54) = 14.94$; $p<0.0001$), consistent with the higher performance of the P/P classification in the intact versus scrambled ROIs. Similarly when the FFA and PPA were removed from OR, the 2-way repeated measures ANOVA revealed significant main effects of scrambling ($F(2,18) = 50.56$; $p < 0.0001$), and classification tests ($F(3,27) = 9.26$; $p < 0.005$), and a significant interaction effect ($F(6,54)=9.97$; $p<0.0001$). Post-hoc tests revealed that performance of the P/P classification was significantly larger than the other three, and performance in the intact ROI was larger than in both scrambled controls.

We used the results from the shuffle-label control in a non-parametric bootstrap analysis to determine whether classification performance for each individual category was significantly above chance in the four classification tests. Surrogate classification performance values for each subject were obtained by randomly drawing from one of the 50 re-shuffles of the shuffle-label control, and averaging these values across subjects. This procedure was repeated $10^6$ times with different random drawings of each subject's surrogate performance, and each time the true performance values were compared with the average of these surrogates. Based on this analysis classification performance for each category was significantly above chance at a threshold of $p< 5*10^{-6}$ in OR and $p< 5*10^{-5}$ in OR-FFA&PPA in all four classification tests. Above-chance classification performance for faces and houses might be expected from previous studies that have shown that mental imagery of these categories elicits a significant increase in the average BOLD response in the FFA and PPA respectively. However, here we show 1) that these results also hold when the FFA and PPA are not included in the analysis, and 2) that imagery of non-“special” categories (i.e., food and tools) also generates reliable activation patterns in object responsive cortex.

## Classification performance during Perception and Imagery in Retinotopic Regions

The four classification tests were also performed in early retinotopic voxels (V1+V2). In the intact (i.e., original, non-scrambled) retinotopic ROI only the P/P classification performed was above chance (Figure 2C). A 2-way repeated measures ANOVA of category X classification type in the intact ROI revealed a significant main effect of category ($F(3,27) = 4.06$; $p<0.02$), a significant main effect of classification type ($F(3,27) = 14$; $p < 0.0001$), and a significant interaction effect ($F(9,81)=4.94$; $p<0.0001$). Post-hoc tests, Bonferroni corrected for multiple comparisons, revealed that classification performance for faces and food was significantly larger than for tools (performance for landmarks was not different from either group) and performance of the P/P classification was significantly larger than the other three tests. Note that it is not surprising that the P/P test performs well in the intact retinotopic ROI – all stimuli were presented at the center of the screen and classification could merely rely on lower level properties of these stimuli (e.g., the similarity in shapes for faces, or the spatial frequencies for landmarks). Importantly, classification performance of the P/P test in the intact retinotopic ROI was significantly lower than in the intact OR ROI ($p < 0.005$). Furthermore, classification performance of the imagery classifier was at chance in the intact retinotopic ROI (see also Figure 3).

Scrambling tests were also performed in the retinotopic ROI (Supplementary Figure 6). A 2-way ANOVA of scrambling type X classifier test revealed significant main effects of scrambling ($F(2,18) = 20.49$; $p < 0.0001$), and classification tests ($F(3,27) = 10.99$; $p < 0.005$),

and a significant interaction effect ($F(6,54)=16.87$; $p<0.0001$). Again, post-hoc tests revealed higher performance in the intact ROI, and for the P/P classification.

These results of the four classification tests performed in all three ROIs are summarized in Figure 3 which reports performance pooled over categories. As mentioned earlier, performance in OR and OR-FFA&PPA was above chance for all classification tests for the intact ROI, but at chance for both scrambling controls (Figure 3A and B). In the retinotopic ROI (Figure 3C), performance was only above chance for the P/P classification in the intact ROI. A 3-way ANOVA of ROI (OR, OR-FFA&PPA, retinotopic) x scrambling type (intact ROI, scrambled voxels, scrambled labels) x classifier test supported these observations. We obtained significant main effects of ROI ($F(1,9) = 39.81$; $p < 0.0001$), classifier test ($F(3,27) = 13.93$; $p < 0.0001$) and scrambling type ($F(2,18) = 72.13$; $p< 0.0001$). Post-hoc tests revealed higher performance in the P/P versus the other three tests and an ROI advantage in the order: OR > OR-FFA&PPA > Retinotopic. A significant interaction effect of classification test x scrambling type ($F(6,54) = 19.26$; $p < 0.0001$) was consistent with the higher performance of the P/P classification in the intact ROIs. Classification performance of the P/P test in the intact retinotopic ROI was significantly lower than in the intact OR ROI ($p < 0.005$) and the intact OR-FFA&PPA ROI ($p<0.05$). Finally, the 2-way interaction of ROI x scrambling type was also significant ($F(2,18) = 31.03$; $p< 0.0001$). There was no significant interaction of ROI x classification test ($F(3,27) = 1.18$; $p = 0.33$).

To summarize, our results show that ventral temporal activation patterns obtained during both visual perception and mental imagery provide information about the object categories being imagined. Furthermore, the activation patterns obtained in the two conditions overlap substantially, thus allowing for successful cross-generalization across the two conditions. We next examine this similarity in representations in greater detail.

## Overlap of representations during perception and imagery

The above-chance classification performance of the P and I classifiers in generalizing across conditions (i.e., imagery and perception respectively), suggests that there is a significant overlap in the representations of these two states in object responsive voxels in ventral temporal cortex. Figure S7 shows one way to visualize this overlap by considering the pattern of weights assigned by the SVM procedure to each voxel -- these weight maps essentially indicate the importance of each voxel's contribution to the discrimination between categories.

In particular, the successful cross-generalization performance argues that the weight maps of the P and I classifiers for a given category should be more similar to each other than the weight map of the P classifier for one category and the I classifier for a different category (or vice versa). To statistically test this prediction, we computed correlations of the weight maps for all pairs of categories, in both ROIs. The correlations were first computed for each subject's weight maps (i.e. on the individual rather than the “average” brain), and then averaged across subjects. The results are shown in Figure 4A for OR and Figure 4B for the retinotopic ROI. A 2-way repeated measures ANOVA of weights (within category/across category) x ROI revealed a significant main effect of weights ($F(1,9) = 54.22$; $p < 0.0001$), a significant main effect of ROI ($F(1,9) = 52.67$; $p < 0.0001$) and a significant interaction effect ($F(1,9) = 48.94$; $p < 0.0005$). Post-hoc Bonferonni corrected comparisons revealed that the weight map overlap within category was higher than across category, and higher in OR than in retinotopic voxels.

In particular, this comparison of weight maps indicated a significant overlap for each category between the representations involved in perception and imagery in OR. Separate statistical tests for each category showed that the weight maps were significantly more correlated within category than across categories (i.e., a comparison of the diagonal versus off-diagonal elements in Figure 4A; $p<0.001$ for each category, paired t-tests). A similar effect was not observed in

the retinotopic voxels for any category, except landmarks (p=0.02). Note also that for each category the within-category correlations in OR were significantly larger than the corresponding within-category correlations in the retinotopic ROI ($p < 0.05$ for tools and $p<0.001$ for the other categories). These results thus indicate that the category representations during perception and imagery share the same "diagnostic voxels" in OR.

## Discussion

In this study we asked three questions: first, if we could reliably decode the content of mental images, second, if visually perceived and imagined objects were coded for in similar regions, and finally, if the representations in both conditions shared equivalent neural substrates at the level of multi-voxel patterns in ventral temporal cortex.

In response to the first question, we found that category-level information for imagined objects (including non-special objects i.e., tools and fruits) could be successfully read out from object responsive voxels in ventral temporal cortex. Second, consistent with other studies (Ishai et al., 2000; Mechelli et al., 2004; O'Craven and Kanwisher, 2000), the same voxels were also involved in the coding of visually perceived stimuli. In the last few years, multi-voxel pattern analysis techniques have been extensively used to not only decode the information available in visual areas, but to also investigate the effects of top-down modulating signals on visual processing. For example, Kamitani and Tong (Kamitani and Tong, 2005) showed that when two stimuli were simultaneously presented to subjects, it was possible to read out which of the two stimuli subjects were attending to. Similarly, Serences and Boynton (Serences and Boynton, 2007) demonstrated that it was possible to decode the attended direction of motion from area MT. More recently, Harrison and Tong showed that orientation information held in working memory could be read out from early visual areas (Harrison and Tong, 2009). Finally, Stokes and colleagues (Stokes et al., 2009) recently found that imagery of the letter X versus the letter O could be decoded from LOC. The present study extends this body of work and suggests that top-down driven visual information of natural object categories can be robustly readout in the complete absence of bottom-up input, during mental imagery.

Although OR was activated in both top-down and bottom-up driven visual processes, as argued earlier, a spatial overlap of voxel activations does not imply shared fine-grained representations at the level of individual voxels. The use of pattern classification techniques in the current study allowed us to conclusively address our third question, and indicated that actual viewing and mental imagery shared the same representations at the level of fine-grained multi-voxel activation patterns in object responsive ventral temporal cortex. When using such classifiers trained on perception to decode imagery and vice versa, we found reliable cross-generalization performance, which in fact was similar to the performance achieved within the imagery condition. Furthermore, the SVM weight maps indicated that the same voxels participated in discriminating between object categories during perception and imagery. Thus the present study demonstrates a high level of similarity between the fine-grained representations involved in perception and imagery of natural object categories. An interesting question for future research would be to assess the similarity of activation patterns during automatic retrieval and visual perception.

A recent study by Stokes et al., (Stokes et al., 2009) reported similarity of multi-voxel patterns in LOC during perception and imagery of two letters (X and O). The authors showed that a classifier trained on activation patterns in anterior LOC during visual perception could decode above chance which of the two items participants were imagining (however, the study did not report classification performance when training on imagery and testing on perception; this condition, which in our study shows reliable classification performance, is important to assess the similarity between the multi-voxel patterns recorded during perception vs. imagery). In

contrast to classifying two elementary shapes, our study reports a 4-way classification of high-level, category-level information within a large and diverse set of colored natural photographs. As detailed in the Methods section, we used four categories of stimuli, with eight exemplars per category, and each half of our fMRI study was based on independent sets of these stimuli. This design thus served to increase the overall variability of the stimulus sets in the classification procedure, and consequently, the generalizability of our findings about high-level representations during perception and imagery.

Perceptual processing involves interactions between top-down signals and bottom-up inputs (Koch and Poggio, 1999; Lamme et al., 1998; Murray et al., 2002; Rao and Ballard, 1999; Williams et al., 2008). The present set of results indicates that feedback signals in the absence of bottom-up input can be sufficient to evoke category-specific representations in ventral temporal cortex. Although these "mental imagery" representations do not induce the same vivid percept as during actual viewing, they were still reliable enough to be decoded with multi-voxel pattern analysis techniques. In contrast, the corresponding information could only be read out from retinotopic voxels when the stimuli were actually viewed, i.e., when bottom-up inputs were present. The role of primary visual cortex during mental imagery is still debated (for a meta-analysis see (Kosslyn and Thompson, 2003)). On the one hand, several recent studies have shown that V1 can be activated when subjects imagine stimuli or retrieve them from memory (Cui et al., 2007; Ishai et al., 2002; Kosslyn et al., 1999; Kosslyn et al., 1995; Slotnick et al., 2005). For instance, Kosslyn and colleagues have argued that mental imagery of objects (Kosslyn et al., 1995) and other simpler stimuli (Kosslyn et al., 1999) activates primary visual cortex, and that performance on the imagery task is impaired after applying r-TMS to these areas (Kosslyn et al., 1999). Cui et al, found that early visual areas were activated during imagery, and further that the activity in these voxels was correlated with participants' subjective report of the vividness of their mental images (Cui et al., 2007). Finally, very recently Harrison & Tong (Harrison and Tong, 2009) showed that orientation information held in working memory could be decoded from fMRI activity in areas V1-V4. However, in contrast to these studies, several other authors have found no evidence for the role of V1 in generating mental images (D'Esposito et al., 1997; Formisano et al., 2002; Ishai et al., 2000; Knauff et al., 2000; Trojano et al., 2000; Wheeler et al., 2000). Consistent with this work, here we also show that patterns of V1 activation do not predict category information for imagined stimuli, but that this information can still be gleaned from higher-level areas. Thus, our results indicate that while V1 may get activated during imagery, it is not a necessary condition for the generation of mental images.

Indeed, V1 activation during imagery may only be called for when participants have to access high-resolution information during visual imagery (Kosslyn and Thompson, 2003), or retrieve information from short-term memory (Ishai et al., 2002). A very simple model of mental imagery (Serre, 2006) predicts that mental images are created via feedback to different areas across the visual hierarchy. According to this model, lower level areas (i.e., V1 vs. OR) will only be activated for difficult tasks that require more time to generate the mental images, or tasks that rely on fine discriminations. Thus, while our findings in this study indicate that the same patterns of neural activity generated during visual perception get reactivated during mental imagery, whether and when lower areas also get involved, remains an open question.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We gratefully acknowledge funding from the Fondation pour la Recherche Medicale and the Fyssen Foundation to L.R., DARPA (IPTO and DSO) and NSF to T.S., and the Japan Society for the Promotion of Science to N.T. We also thank John Serences for the use of his meridian mapping code, and Rufin VanRullen, Tomaso Poggio and Michèle Fabre-Thorpe for comments on the manuscript. Finally, we are grateful to Christof Koch for providing financial support for this study.

## References

Amedi A, Malach R, Pascual-Leone A. Negative BOLD differentiates visual imagery and perception. Neuron 2005;48:859–872. [PubMed: 16337922]

Carlson TA, Schrater P, He S. Patterns of activity in the categorical representations of objects. J Cogn Neurosci 2003;15:704–717. [PubMed: 12965044]

Cox DD, Savoy RL. Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 2003;19:261–270. [PubMed: 12814577]

Cui X, Jeter CB, Yang D, Montague PR, Eagleman DM. Vividness of mental imagery: individual variability can be measured objectively. Vision Res 2007;47:474–478. [PubMed: 17239915]

D'Esposito M, Detre JA, Aguirre GK, Stallcup M, Alsop DC, Tippet LJ, Farah MJ. A functional MRI study of mental image generation. Neuropsychologia 1997;35:725–730. [PubMed: 9153035]

Finke RA. Theories relating mental imagery to perception. Psychol Bull 1985;98:236–259. [PubMed: 3901061]

Fischl B, Sereno MI, Tootell RB, Dale AM. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Mapp 1999;8:272–284. [PubMed: 10619420]

Formisano E, Linden DE, Di Salle F, Trojano L, Esposito F, Sack AT, Grossi D, Zanella FE, Goebel R. Tracking the mind's image in the brain I: time-resolved fMRI during visuospatial mental imagery. Neuron 2002;35:185–194. [PubMed: 12123618]

Ganis G, Thompson WL, Kosslyn SM. Brain areas underlying visual mental imagery and visual perception: an fMRI study. Brain Res Cogn Brain Res 2004;20:226–241. [PubMed: 15183394]

Goebel R, Khorram-Sefat D, Muckli L, Hacker H, Singer W. The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. Eur J Neurosci 1998;10:1563–1573. [PubMed: 9751129]

Harrison SA, Tong F. Decoding reveals the contents of visual working memory in early visual areas. Nature. 2009

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 2001;293:2425–2430. [PubMed: 11577229]

Haynes JD, Rees G. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci 2005;8:686–691. [PubMed: 15852013]

Ishai A, Haxby JV, Ungerleider LG. Visual imagery of famous faces: effects of memory and attention revealed by fMRI. Neuroimage 2002;17:1729–1741. [PubMed: 12498747]

Ishai A, Sagi D. Common mechanisms of visual imagery and perception. Science 1995;268:1772–1774. [PubMed: 7792605]

Ishai A, Ungerleider LG, Haxby JV. Distributed neural systems for the generation of visual images. Neuron 2000;28:979–990. [PubMed: 11163281]

Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. Nat Neurosci 2005;8:679–685. [PubMed: 15852014]

Kay KN, Naselaris T, Prenger RJ, Gallant JL. Identifying natural images from human brain activity. Nature 2008;452:352–355. [PubMed: 18322462]

Knauff M, Kassubek J, Mulack T, Greenlee MW. Cortical activation evoked by visual mental imagery as measured by fMRI. Neuroreport 2000;11:3957–3962. [PubMed: 11192609]

Koch C, Poggio T. Predicting the visual world: silence is golden. Nat Neurosci 1999;2:9–10. [PubMed: 10195172]

Kosslyn SM, Pascual-Leone A, Felician O, Camposano S, Keenan JP, Thompson WL, Ganis G, Sukel KE, Alpert NM. The role of area 17 in visual imagery: convergent evidence from PET and rTMS. Science 1999;284:167–170. [PubMed: 10102821]

Kosslyn SM, Thompson WL. When is early visual cortex activated during visual mental imagery? Psychol Bull 2003;129:723–746. [PubMed: 12956541]

Kosslyn SM, Thompson WL, Kim IJ, Alpert NM. Topographical representations of mental images in primary visual cortex. Nature 1995;378:496–498. [PubMed: 7477406]

Kreiman G, Koch C, Fried I. Imagery neurons in the human brain. Nature 2000;408:357–361. [PubMed: 11099042]

Lamme VA, Super H, Spekreijse H. Feedforward, horizontal, and feedback processing in the visual cortex. Curr Opin Neurobiol 1998;8:529–535. [PubMed: 9751656]

Mechelli A, Price CJ, Friston KJ, Ishai A. Where bottom-up meets top-down: neuronal interactions during perception and imagery. Cereb Cortex 2004;14:1256–1265. [PubMed: 15192010]

Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL. Shape perception reduces activity in human primary visual cortex. Proc Natl Acad Sci U S A 2002;99:15164–15169. [PubMed: 12417754]

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 2006;10:424–430. [PubMed: 16899397]

O'Craven KM, Kanwisher N. Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. J Cogn Neurosci 2000;12:1013–1023. [PubMed: 11177421]

Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 1999;2:79–87. [PubMed: 10195184]

Reddy L, Kanwisher N. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. Curr Biol 2007;17:2067–2072. [PubMed: 17997310]

Serences JT, Boynton GM. Feature-based attentional modulations in the absence of direct visual stimulation. Neuron 2007;55:301–312. [PubMed: 17640530]

Serre, T. Brain and Cognitive Sciences. Massachusetts Institute of Technology; Cambrdige, MA: 2006. Learning a dictionary of shape-components in visual cortex: Comparison with neurons, humans and machines.

Slotnick SD, Thompson WL, Kosslyn SM. Visual mental imagery induces retinotopically organized activation of early visual areas. Cereb Cortex 2005;15:1570–1583. [PubMed: 15689519]

Spiridon M, Kanwisher N. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. Neuron 2002;35:1157–1165. [PubMed: 12354404]

Stokes M, Thompson R, Cusack R, Duncan J. Top-down activation of shape-specific population codes in visual cortex during mental imagery. J Neurosci 2009;29:1565–1572. [PubMed: 19193903]

Trojano L, Grossi D, Linden DE, Formisano E, Hacker H, Zanella FE, Goebel R, Di Salle F. Matching two imagined clocks: the functional anatomy of spatial analysis in the absence of visual stimulation. Cereb Cortex 2000;10:473–481. [PubMed: 10847597]

Wheeler ME, Petersen SE, Buckner RL. Memory's echo: vivid remembering reactivates sensory-specific cortex. Proc Natl Acad Sci U S A 2000;97:11125–11129. [PubMed: 11005879]

Williams MA, Baker CI, Op de Beeck HP, Shim WM, Dang S, Triantafyllou C, Kanwisher N. Feedback of visual object information to foveal retinotopic cortex. Nat Neurosci 2008;11:1439–1445. [PubMed: 18978780]

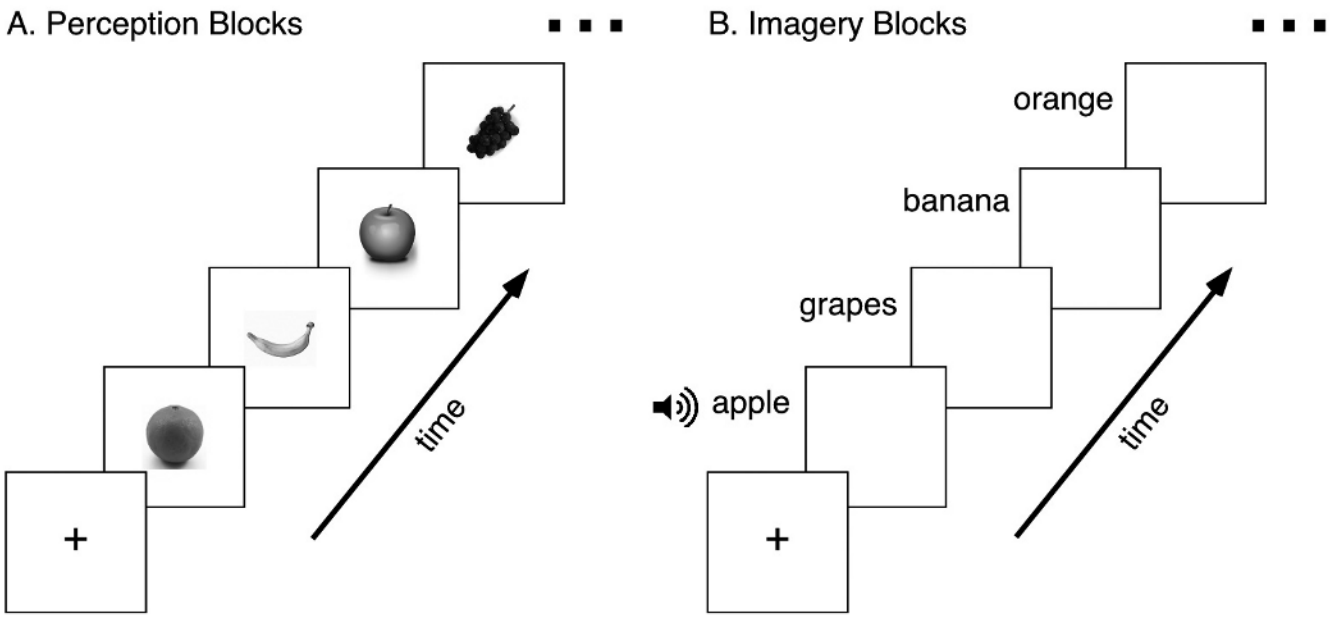


**Figure 1.**

Experimental Design. The experiment consisted of two conditions. A). In the visual perception (P) condition subjects viewed different exemplars of 4 categories of objects (tools, food (common fruits and vegetables), famous faces and famous buildings). B). In the visual imagery (I) condition subjects were given auditory instructions with the names of the stimuli and asked to generate vivid and detailed mental images corresponding to these names. Note that in the actual experiment colored stimuli were used (see Figure S2).

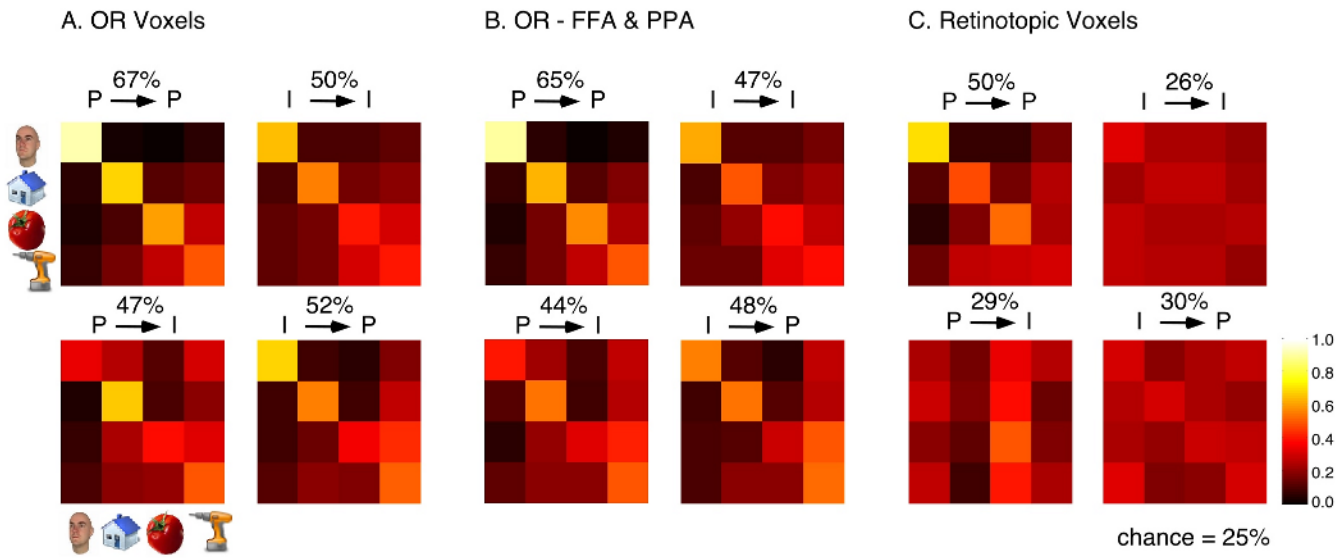


**Figure 2.**
Confusion matrices for classification in A) the intact OR ROI, B) OR ROI with the FFA and PPA excluded and C) the Retinotopic voxels. Each confusion matrix shows the probability with which an input pattern presented along the rows would be classified as one of the 4 alternative outcomes (along the columns). P/P and I/I correspond to classification performance when both training and testing was performed on the visual presentation conditions or the mental imagery conditions respectively. P/I corresponds to training on visual presentation and testing on imagery (and vice versa for I/P).

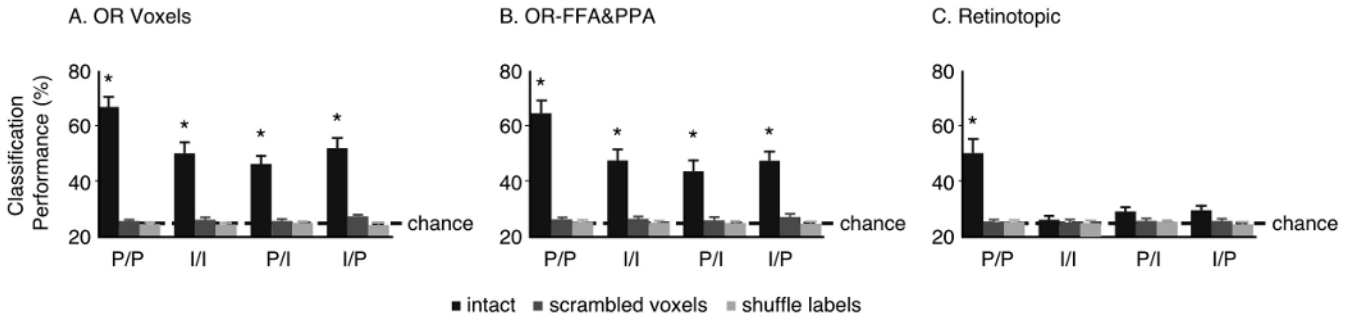


**Figure 3.**

Classification performance for the 4 types of classification pooled over categories in A) object responsive voxels, B) OR-FFA&PPA and C) in retinotopic voxels. "Scrambled voxels" corresponds to scrambling the voxel order for the test data in comparison to the training data, and "Shuffle labels" corresponds to shuffling the labels of the training examples. The performance values plotted here correspond to the mean of the diagonal values in the corresponding matrices in Figures 2 and S3 (* = $p < 0.005$). Note that since a 4-way classification was performed, chance performance is at 25%. Error bars represent S.E.M.

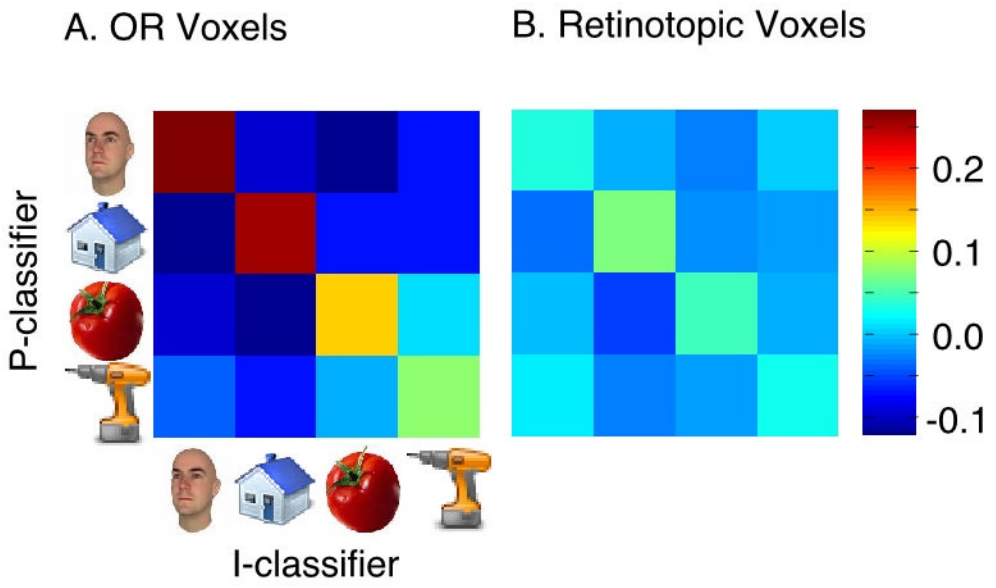


**Figure 4.**
Correlation of the SVM weight maps of the P and I classifiers for all pairs of categories in OR (A) and the retinotopic voxels (B).