Fine-scale activity patterns in high-level visual areas encode the category of invisible objects

Philipp Sterzer

Wellcome Trust Centre for Neuroimaging,
Institute of Neurology, United Kingdom,
Institute of Cognitive Neuroscience,
University College London, United Kingdom, &
Department of Psychiatry, Charité University Hospital,
Humboldt University, Berlin, Germany

John-Dylan Haynes

Bernstein Center for Computational Neuroscience,
Berlin, Germany

Geraint Rees

Wellcome Trust Centre for Neuroimaging,
Institute of Neurology, United Kingdom, &
Institute of Cognitive Neuroscience,
University College London, United Kingdom

When incompatible images are presented to the two eyes, one image can dominate awareness while the other is suppressed and invisible. We used high-resolution functional neuroimaging in humans to investigate the neural representation of such suppressed stimuli. Overall responses of high-level ventral visual areas to two different types of invisible object stimuli (faces and houses) were very weak and did not differ in amplitude. Despite this, fine-grained spatial activity patterns within these areas allowed us to predict significantly better than chance whether an observer was presented with face or house stimuli not only when these stimuli were visible but also when they were suppressed and entirely invisible. These findings demonstrate the presence of category-specific information in high-level visual areas during profound interocular suppression of object stimuli that can only be retrieved when the fine-scale pattern of activity within these areas is taken into account.

Introduction

When dissimilar images are presented to each eye, one image can dominate awareness while the other is suppressed and invisible (Blake & Logothetis, 2002; Tong, Meng, & Blake, 2006). It has been suggested that this phenomenon, known as binocular rivalry, reflects the outcome of competitive neuronal interactions at multiple levels of the visual system (Blake & Logothetis, 2002; Tong et al., 2006; Wilson, 2003). That such interactions occur even at the earliest monocular stages of visual processing is supported by recent evidence from functional magnetic resonance imaging (Haynes, Deichmann, & Rees, 2005; Tong & Engel, 2001; Wunderlich, Schneider, & Kastner, 2005). The outcome of binocular rivalry can also be influenced by complex information from suppressed stimuli (Alais & Parker, 2006; Andrews & Blakemore, 1999; Jiang, Costello, Fang, Huang, & He, 2006; Jiang, Costello, & He, 2007; Kovács, Papathomas, Yang, & Féher, 1996), suggesting competitive interactions

between neural representations of such information at higher levels of the visual system. However, neurophysiological evidence regarding the representation of suppressed stimuli in high-level visual areas remains inconclusive. In line with earlier electrophysiological findings (Kreiman, Fried, & Koch, 2002; Sheinberg & Logothetis, 1997), functional magnetic resonance imaging (fMRI) revealed percept-related activity fluctuations during rivalry in high-level visual areas of the ventral visual pathway similar to those during actual stimulus alternations (Sterzer & Rees, 2008; Tong, Nakayama, Vaughan, & Kanwisher, 1998), suggesting near-complete suppression of neural activity representing the non-dominant stimulus. While some studies that explicitly probed responses to suppressed visual stimuli indeed failed to observe significant activity in ventral high-level visual areas (Fang & He, 2005; Pasley, Mayes, & Schultz, 2004; Williams, Morris, McGlone, Abbott, & Mattingley, 2004), more recently some fMRI activations have been observed in these areas in response to interocularly suppressed stimuli (Jiang & He, 2006). However, it is

not known whether the well-documented preferences for particular object categories in ventral visual areas, such as the fusiform face area (FFA) or parahippocampal place area (PPA; Epstein, Harris, Stanley, & Kanwisher, 1999; Kanwisher, McDermott, & Chun, 1997), are preserved when stimuli are suppressed during binocular rivalry.

The sensitivity of fMRI measurements can be substantially increased by the use of recently developed multivariate pattern analysis (MVPA) techniques, which, compared to conventional univariate approaches, have the advantage that weak information from single voxels can be accumulated in an efficient way across many voxels (Haynes & Rees, 2006; Kriegeskorte, Goebel, & Bandettini, 2006; Norman, Polyn, Detre, & Haxby, 2006). MVPA was first used to demonstrate extended category-specific cortical activity patterns in response to visible object stimuli (Haxby et al., 2001), but recently also revealed that activity patterns in early visual cortex can encode information about stimuli even if they are entirely invisible (Haynes & Rees, 2005). Here, we tested whether high-resolution fMRI in conjunction with MVPA (Kriegeskorte et al., 2006) can reveal information about suppressed stimuli at higher levels of the human visual system, where conventional techniques have repeatedly failed to identify category-specific responses (Pasley et al., 2004; Williams et al., 2004).

Methods

Data acquisition and experimental design

Five healthy volunteers aged 21–35 years (mean age 29, 4 male) with normal vision gave written informed consent to participate in the study, which was approved by the local ethics committee. Functional MRI was performed on a 3T Siemens Allegra system using a standard head coil. We obtained high-resolution (25 slices, voxel size $1.5 \times 1.5 \times 1.5$ mm) blood-oxygenation-level dependent (BOLD) echo-planar image volumes every 2.55 s (Haynes et al., 2005). As the focus of our study was on responses in the FFA and PPA, the scan volume was positioned to fully cover the mid-fusiform and the parahippocampal gyri, which due to the small scan volume did not allow for reliable coverage of other brain structures that might also have been of interest, such as primary visual cortex. Stimuli were projected from an LCD projector (NEC LT158, refresh rate 60 Hz) onto a screen at the head-end of the scanner that was viewed via a mirror attached to the head coil directly above the participants' eyes (viewing distance 62 cm). The size of the screen was 10.5×8.4 deg of visual angle. For dichoptic visual stimulation, the screen was vertically split into two halves by a black cardboard divider that was placed between the screen and the mirror, thus separating images presented to left and right eyes. Perceptually, all participants reported that the two halves of the screen readily fused into one vertically oriented rectangle. The visual stimuli were projected onto the centers of the two half screens. Stimuli were presented and key presses were recorded with MATLAB (Mathworks, Sherborn, MA) using the COGENT 2000 toolbox (www.vislab.ucl.ac.uk/cogent/index.html).

Target stimuli were low-contrast black-and-white photographs of faces and houses on a gray background presented monocularly (Figure 1). These categories were chosen because they are well known to elicit robust and reliable responses in anatomically well-defined areas (Epstein et al., 1999; Kanwisher et al., 1997), and because they have been used in previous studies addressing related questions (Moutoussis & Zeki, 2002; Tong et al., 1998; Williams et al., 2004). Note that this study did not focus on category-*selectivity* of neural processing (i.e., stronger responses to a given category than to any other category).
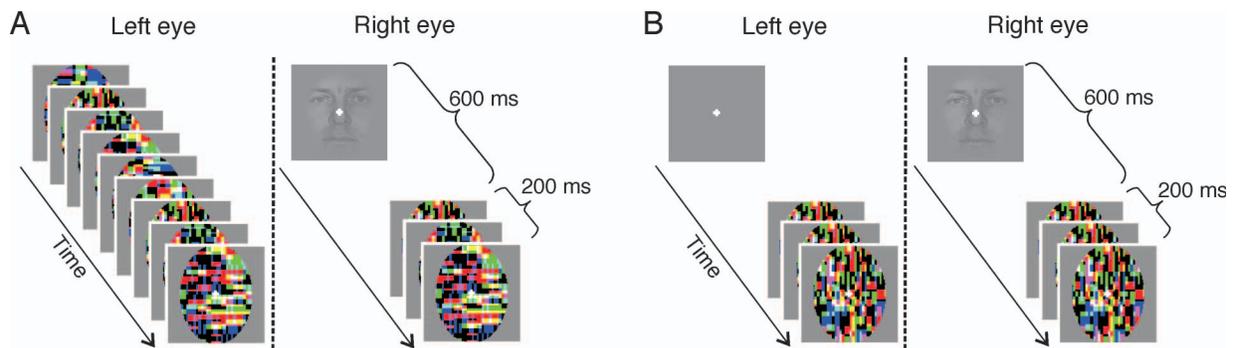

Figure 1. Stimulus presentation. In the invisible conditions (A), a low-contrast target stimulus (a face, as shown here, or a house) presented to one eye for 600 ms was rendered invisible by showing rapidly changing (30 Hz) high-contrast Mondrian-like patterns to the contralateral eye (continuous-flash suppression, CFS). CFS stimuli were presented to both eyes for the following 200 ms to prevent afterimages to the target. In the visible conditions (B), a blank screen was presented to the contralateral eye, resulting in conscious perception of the target, again followed by binocular presentation of the CFS stimuli for 200 ms. Eight stimuli of one type (visible faces, visible houses, invisible faces, or invisible houses) were presented in 12.8 s blocks, directly followed by an awareness assessment and a rest period of 7.65 s duration.

Category-*specificity* here refers to responses or response patterns that differ significantly comparing the two stimulus categories used. Mean luminance of the face and house pictures was equal to the background luminance. Eight different exemplars of each category were used. Both face and house stimuli were of oval shape (height 4 deg of visual angle, width 2.8 deg) to preclude overall stimulus shape as confound, and the edges were blurred into the gray background. The other eye was either presented with high-contrast Mondrian-like patterns flashed at a frequency of 30 Hz for continuous flash suppression (CFS; Tsuchiya & Koch, 2005), i.e., to render the target stimuli invisible (Figure 1A); or with a blank gray background that resulted in dominance of the target stimulus (Figure 1B). To facilitate fixation and binocular alignment, both the target and CFS stimuli were overlaid with a white central fixation cross and surrounded by a white square (4 × 4 deg). Stimuli were presented for 800 ms with an inter-stimulus interval of 800 ms, during which only the white fixation cross and the white square were presented. To avoid afterimages, the target stimuli were replaced after 600 ms by the same CFS stimuli that were presented to the other eye (see Figure 1). In the visible conditions, where the eye contralateral to the target stimulus was presented only with the gray background for the first 600 ms, this resulted in the perception of a target stimulus followed by a short sequence of CFS stimuli.

Visible faces, visible houses, invisible faces, or invisible houses were presented in blocks of 12.8 s (8 stimuli of one category per block). Stimulus order and the assignment of target and CFS stimuli to each eye within the blocks were completely randomized, as were the block sequences. The whole experiment comprised 40 blocks of each condition and was subdivided in 10 short runs of 5 min 39 s duration. The blocks were interleaved with a 'baseline' condition of 7.65 s duration, at the beginning of which awareness of the stimuli was assessed. The first question represented a 2-alternative forced choice task ('face or house?') for the 'objective' assessment of awareness (Pessoa, 2005). Participants had to indicate which stimulus category they were presented with in the preceding block (visible conditions), or to guess when they did not see the face or house stimuli (as usually the case in invisible conditions). Detectability (d') was estimated according to signal detection theory as the difference between *z*-transformed hit and false-alarm rates (Macmillan & Creelman, 1991). To assess 'subjective' awareness, participants were asked to rate how confident they were of their choice on the first question on a rating scale from 0 to 3 (0 = 'no clue'; 1 = 'did not see stimuli but had a certain feeling'; 2 = 'saw bits of the stimuli but not enough to be sure'; 3 = 'could clearly identify stimulus category'). Blocks where the subjective confidence rating was >0 occurred rarely (on average 4.1% ± 2.3 *SEM* of all blocks) and were excluded from further analysis. Responses were given via right hand button presses. One participant reported after the experiment that he was unable to completely binocularly fuse the stimuli due to sleepiness during one experimental run, which rendered some stimuli in the invisible conditions partly visible. This run was discarded from the analysis.

In addition to the main experiment, we performed a functional localizer experiment (2 × 205 scan volumes, TR = 2.21 s) at standard resolution (voxel size 3 × 3 × 3 mm) to define each subjects FFA and PPA. In brief, participants were presented with 22 s blocks of either black-and-white faces or buildings presented on a white screen (500 ms each with 500 ms inter-stimulus interval), interleaved with 22 s blocks during which scrambled versions of these stimuli were presented. A T1-weighted structural scan was also performed (voxel size 1 × 1 × 1 mm).

Imaging data analysis

Data preprocessing

Data were preprocessed using statistical parametric mapping software (SPM2, www.fil.ion.ucl.ac.uk/spm). After discarding the first five image volumes from each run to allow for T1 equilibration effects, functional image volumes were realigned to the first of the remaining volumes, unwarped (Andersson, Hutton, Ashburner, Turner, & Friston, 2001), and co-registered to the individual participants' structural scans. The functional FFA- and PPA-localizer scans were smoothed using a 5 mm full-width-at-half-maximum Gaussian kernel. Data from the main experiment were not spatially smoothed unless otherwise stated (see below). We removed low frequency fluctuations with a high-pass filter with a cut-off at 128 s and used an autoregressive model of order one (AR(1) + white noise) to correct for temporal autocorrelation in the data. The time series of each voxel was normalized to the mean across all scan volumes of each run to reduce variability between experimental runs.

Voxel-wise univariate analyses

First, FFA and PPA regions of interest (ROI) were determined for each participant individually from the functional localizer experiment according to standard criteria (Kanwisher et al., 1997). Activated voxels in each experimental condition were identified using a general linear model (GLM) containing boxcar waveforms representing each of the four experimental conditions, convolved with a canonical hemodynamic response function (HRF). Parameter estimates for each regressor at every voxel were determined using multiple linear regression and scaled to the global mean signal of each run across conditions and voxels. Contiguous voxels in the fusiform gyrus that responded significantly ($p < 0.001$, uncorrected) more to faces than to houses were defined as the FFA. Likewise, contiguous voxels in the parahippocampal gyrus that responded more to houses than to faces were defined as the PPA. Both the FFA and the PPA could be

identified with confidence bilaterally in all 5 participants and comprised on average 480 ± 62 (*SEM*) high-resolution voxels (Figure 2A).

The data from the main experiment were first similarly analyzed using a GLM. Again, boxcars convolved with a canonical HRF were used to model regressors for the four main conditions of interest (visible faces, visible houses, invisible faces, invisible houses), for the awareness assessment following each block, and for the inter-block baseline. Blocks for which the participants' confidence rating was >0 were modeled as separate regressors that were not used for any of the statistical comparisons. All further analyses were performed using the 200 voxels within the FFA and PPA masks (as derived from the localizer experiment) that were most responsive to our stimulus paradigm (Haynes & Rees, 2005). Responsiveness was determined using an (unsigned) F-contrast including the four conditions of interest. Importantly, this contrast tests for effects in any direction for any included condition and therefore imposes no bias on the likelihood of finding any differences *between* conditions.

To maximize sensitivity for the detection of category-specific effects in the invisible conditions, we used various strategies to identify such responses:

1. *ROI averages of parameter estimates from unsmoothed data:* The response of each ROI (as determined by the functional localizer experiment) in each condition was determined by averaging the condition-specific parameter estimates from all voxels within an ROI. Each participant's ROI averages for each condition were then subjected to *t*-tests for statistical inference at the group level (see Figure 3A).

2. *Peak responses from smoothed data:* A separate analysis was performed after spatially smoothing the data (3 mm full-width-at-half-maximum Gaussian

kernel). We determined the response maxima for visible faces > houses within the FFA and for visible houses > faces in the PPA, and then tested for response differences in the invisible conditions at these maxima. Smoothing creates a weighted average of nearby voxels determined by the smoothing kernel. This might slightly differ from the average across all ROI voxels (as determined by the standard-resolution functional localizer) and could therefore increase the sensitivity for response differences in FFA and PPA subregions that responded most strongly to visible stimuli in the main experiment (at high resolution). The resulting parameter estimates were again subjected to *t*-tests for statistical inference at the group level (see Figure 3B).

3. *Mean fMRI signal across blocks from unsmoothed data:* For this analysis, for which preprocessed (unsmoothed) fMRI data rather than parameter estimates from a GLM were used, each voxel's fMRI signal was averaged across blocks of visual stimulation as was done for multivariate pattern classification (see below; Kamitani & Tong, 2005), to allow for better comparability of the results from univariate and multivariate analyses. Single data points for each individual visual stimulation block and each voxel were generated by shifting the preprocessed fMRI time series by 2 scan volumes (5.1 s) to account for the hemodynamic delay, and then averaging each voxel's fMRI signal intensity across the following 4 scan volumes (Kamitani & Tong, 2005). For ROI-wise univariate statistical inference, the mean signal across each ROI as determined by the functional localizer experiment was subjected to *t*-tests at the group level (see Figure 3C). In addition, voxel-wise *t*-tests were performed for the analysis of response preferences (Haynes & Rees, 2005).
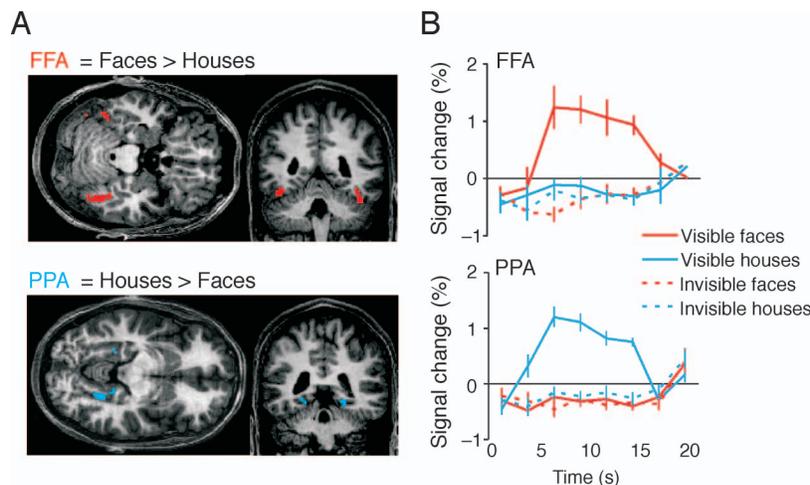

Figure 2. (A) Fusiform face area (FFA) and parahippocampal place area (PPA) of one participant determined from a functional localizer experiment at standard resolution (3 mm isotropic voxels). (B) Signal time courses (error bars = *SEM*) for visible faces, visible houses, invisible faces, and invisible houses in the main experiment (1.5 mm isotropic voxels) from FFA and PPA averaged across participants.
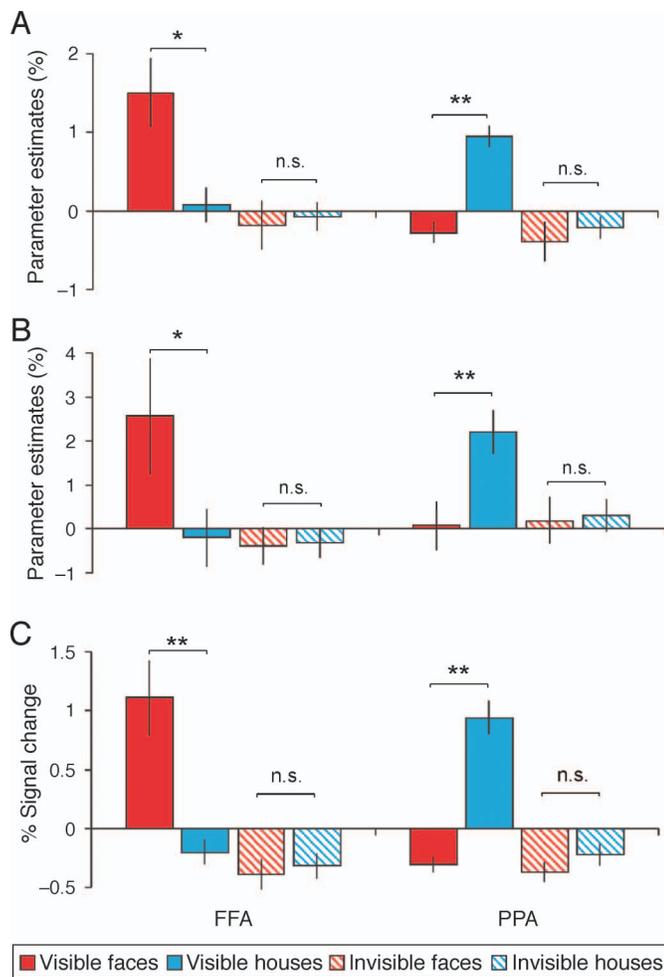
Figure 3. None of the several types of univariate analyses yielded significant category-specific responses to invisible stimuli in the FFA or PPA. (A) ROI averages of parameter estimates from unsmoothed data. (B) Parameter estimates at local maxima for contrasts of visible stimuli from smoothed data. (C) ROI averages of mean signal across visual stimulation blocks. *Note*: $*p < 0.05$; $**p < 0.005$; n.s. = non-significant; FFA = fusiform face area; PPA = parahippocampal place area.

Multivariate pattern analysis

MVPA was used to predict from distributed response patterns in the FFA and PPA ROIs, which of two given stimulus categories was currently being presented. The average BOLD signal across each visual stimulation block of each condition (see above) was extracted for 200 voxels within each ROI, forming a set of pattern vectors $\mathbf{a}$, which were transformed to normalized activation vectors $\mathbf{x}$ with unit length following $\mathbf{x} = \mathbf{a}/\|\mathbf{a}\|$. Pattern classification was performed using linear support vector machines (SVM; Vapnik, 1995) in the implementation of Gunn (http://www.isis.ecs.soton.ac.uk/). A linear classifier finds a hyperplane

$$\mathbf{w}^T\mathbf{x} + b = 0 \qquad (1)$$

defined by weight vector $\mathbf{w}$ and offset b separating the training points $\mathbf{x}$ with two different given labels. The principle of SVM is to find the optimally separating hyperplane that maximizes the margin (given by $2/\|\mathbf{w}\|$) with respect to both training classes (for detailed algorithm, see Christianini & Shawe-Taylor, 2000; Vapnik, 1995). Under the presence of noise (as here), the response vectors of both stimuli might not be linearly separable and a so-called "soft-margin classifier" can be used, which allows for a certain proportion of misclassifications by minimizing

$$\|\mathbf{w}\|/2 + C\sum \xi_i \qquad (2)$$

subject to

$$y_i(\mathbf{w}\mathbf{x}_i + b) \geq 1 - \xi_i \quad i = 1, 2, ..., N \quad \xi \geq 0, \qquad (3)$$

where ξ_i is a slack variable representing misclassification error for the ith pattern $\mathbf{x}_i$ with label $y_i \in \{1, -1\}$ and C a regularization parameter (default $C = 1$; Fu et al., 2008) determining the trade-off between the largest margin and lowest number of misclassifications.

Classification performance was assessed using a standard N-leave-one-out cross-validation procedure. The data set of N pattern vectors per condition (max. 40, if the confidence rating was always 0; see above) was subdivided into a training set of $N - 1$ pattern vectors and a test set (1 pattern vector), with N possible assignments of independent training and test data sets. For SVM classification, the classifier was iteratively trained on the N training data sets and tested on the respective test data set. Importantly, the 200 most responsive voxels (see above) were determined for each iteration independently by performing a separate F-test on each iteration's training data set to not impose an a priori bias on SVM classification (Haynes & Rees, 2005). Classification accuracies were averaged across all different training and test data assignments. The average classification accuracies from each participant were then for each pairwise comparison subjected to two-tailed one-sample t-tests at the group level testing for significant deviation from chance level (50%).

SVM classification was performed for the following pairwise comparisons: visible faces vs. visible houses from FFA voxels; invisible faces vs. invisible houses from FFA voxels; visible faces vs. visible houses from PPA voxels; invisible faces vs. invisible houses from PPA voxels; visible faces vs. visible houses from all FFA and PPA voxels together; invisible faces vs. invisible houses from all FFA and PPA voxels together. Finally, we also probed the performance of SVM classifiers trained on visible conditions when tested on invisible conditions and *vice versa* ('cross-classification'). To that end, SVM classification was performed using the N-leave-one-out procedure as described above where the test data set in

each iteration was replaced by data from the other respective condition (i.e., training on visible faces vs. visible houses and testing on invisible faces vs. invisible houses, and *vice versa*).

In addition to the FFA and PPA, SVM classification was also performed in three control regions: ventromedial prefrontal cortex, retinotopic visual cortex, and occipital face area (OFA). The ventromedial prefrontal was chosen as a control ROI outside visual cortex because it was covered by the high-resolution scan volume in all participants and because activity in this structure is not typically thought to reflect category-specific processing of object stimuli. It was therefore expected to lack signals enabling the SVM classifier to correctly predict stimulus category. Signal was extracted from all voxels within a spherical ROI of 15 mm diameter (512 voxels), the location of which was chosen to cover mostly gray matter but was otherwise arbitrary. The retinotopic cortex ROI was determined from the local maximum in the occipital pole region that was responsive to all four stimulus types (all stimuli vs. baseline). Signal was again extracted from all voxels within an ROI of 15 mm diameter centered on this maximum. It should be noted that the relatively small high-resolution scan volume was placed to optimally cover FFA and PPA and therefore only covered the infracalcarine portion of the occipital pole region in two participants. Moreover, the foveal representation of the stimuli precluded the attribution of fMRI responses to specific retinotopic areas. The resulting retinotopic cortex ROI can therefore only be a rough approximation to the actual representation of the stimuli in early retinotopic visual areas. Finally, the OFA was defined from the same functional localizer scan that was used to define the FFA, using the same procedure except that in two participants the threshold for determining the OFA had to be lowered to $p < 0.01$, uncorrected, in order to obtain a sufficient number of voxels for MVPA. The OFA was defined as the activation cluster of contiguous voxels that responded more to faces than to houses and that was located in the ventral occipital cortex posterior to, and clearly distinct from, the FFA. It could be identified with confidence bilaterally in all 5 participants and comprised on average 514 ± 329 (*SEM*) high-resolution voxels. The algorithm for SVM classification in prefrontal cortex, retinotopic cortex, and OFA was identical to the one used for FFA and PPA, including the selection of the 200 most responsive voxels.

Results

Behavioral results

Online visibility assessment during scanning confirmed profound suppression of the monocular face or house

stimuli from awareness by CFS. Discrimination of invisible faces and houses was tested in a 2-alternative-forced-choice task performed directly after each block; performance was at chance level in all participants (average $d' = -0.05 \pm 0.14$ *SEM*, $p = 0.71$, one-sample *t*-test).

Univariate analyses

Conventional univariate analyses of brain activity evoked in the FFA and PPA of each participant confirmed strong and highly significant differential activations for visible faces vs. visible houses in the FFA and, conversely, for visible houses vs. visible faces in the PPA (Figures 2 and 3). However, in line with previous observations (Pasley et al., 2004; Williams et al., 2004), no differential activations could be detected in these areas in response to invisible face and house stimuli. Several different analyses were applied to maximize sensitivity. First, the response of each ROI (as determined by the functional localizer experiment) in each condition was determined by averaging the condition-specific parameter estimates from all voxels within an ROI (Figure 3A). Significant differences were observed for visible faces > houses in the FFA ($t(4) = 2.9$, $p = 0.02$) and visible houses > faces in the PPA ($t(4) = 6.5$, $p = 0.0002$). No such effect was observed for invisible stimuli ($t(4) = -0.3$, $p = 0.77$, and $t(4) = 0.6$, $p = 0.5$, respectively). The difference in processing of visible and invisible stimuli could be substantiated using two-way repeated-measures analyses of variance (ANOVA), which showed a significant interaction of the factors stimulus category (face or house) and visibility (visible or invisible) for both FFA and PPA ($F(1,4) = 18.6$, $p = 0.01$, and $F(1,4) = 30.7$, $p = 0.005$, respectively). A second analysis was performed using smoothed data (3 mm Gaussian kernel). Parameter estimates were determined from the peak voxels in each ROI from the comparison of visible conditions and then again subjected to *t*-tests at the group level (Figure 3B). As response maxima from visible conditions were used, there were significant effects for visible faces > houses in the FFA ($t(4) = 4.0$, $p = 0.017$) and visible houses > faces in the PPA ($t(4) = 13.2$, $p < 0.001$). Again, no such effects were observed for invisible stimuli ($t(4) = -0.2$, $p = 0.82$, and $t(4) = 0.35$, $p = 0.75$, respectively) and there was a significant category-by-visibility interaction in both FFA and PPA ($F(1,4) = 13.3$, $p = 0.02$, and $F(1,4) = 24.2$, $p = 0.008$, respectively). In a third analysis, preprocessed (unsmoothed) fMRI data rather than parameter estimates from a GLM were used (Figure 3C) and each voxel's fMRI signal was averaged across blocks of visual stimulation as was done for MVPA (Kamitani & Tong, 2005), to allow for better comparability of the results from univariate analysis and MVPA. Again paired *t*-tests revealed significant differences between visible stimuli (FFA: $t(4) = 3.9$, $p = 0.004$; PPA: $t(4) = 8.0$, $p = 0.00004$)

but no detectable effects for invisible stimuli (FFA: $t(4) = -0.43$, $p = 0.67$; PPA: $t(4) = -1.2$, $p = 0.27$). As in the previous two analyses, there was a significant category-by-visibility interaction in both FFA and PPA ($F(1,4) = 13.5$, $p = 0.02$, and $F(1,4) = 62.8$, $p = 0.001$, respectively).

Multivariate pattern analysis

Next, we used SVM classification (Vapnik, 1995), an established method for MVPA (Kamitani & Tong, 2005), to test whether FFA and PPA response patterns could predict whether participants were presented with face or house stimuli, both when stimuli were visible and when invisible, on a trial-by-trial basis. Each short block was treated as one trial and all trials of a condition pair (visible faces/visible houses or invisible faces/invisible houses) were divided into training and test data sets. The SVM classifiers were trained on the training trials and were then applied to the independent test trials in a standard *N*-leave-one-out cross-validation procedure.

As shown in Figure 4, single-trial prediction was significantly above chance level (50%) at correctly classifying visible stimuli as faces or houses both from FFA (on average 76.7% ± 4.0, *SEM*; $t(4) = 6.6$, $p = 0.003$) and the PPA voxels (75.6% ± 2.2 *SEM*; $t(4) = 11.5$, $p < 0.001$). Pooling of all voxels from FFA and PPA together resulted in highly accurate prediction for visible faces vs. houses (88.7% ± 2.9 *SEM*, $t(4) = 13.6$, $p < 0.001$). Strikingly, prediction accuracy for invisible stimuli (where univariate analyses had failed to show differences between activity evoked by face and house stimuli) was also significantly above chance level (FFA: 58.8 % ± 2.3 *SEM*, $t(4) = 3.8$, $p = 0.019$; PPA: 62.5% ± 3.3 *SEM*, $t(4) = 3.8$, $p = 0.019$; FFA + PPA: 63.5% ± 3.7 *SEM*, $t(4) = 3.6$, $p = 0.022$). Prediction accuracies were significantly higher for visible than for invisible stimuli (FFA: $t(4) = 4.1$, $p = 0.015$; PPA: $t(4) = 3.8$, $p = 0.018$; FFA + PPA: $t(4) = 7.4$,

$p = 0.002$). Thus, MVPA can recover the category (face versus house) of both visible (as expected) and also entirely invisible stimuli from category-specific areas (FFA or PPA) in the human ventral visual pathway.

Analysis of voxel response preferences

It has been suggested that MVPA takes advantage of signals that are too weak to give significant results in univariate analyses but are stable enough over time to constitute reliable pattern information (Haxby et al., 2001; Haynes & Rees, 2005; Kamitani & Tong, 2005). We therefore tested the reproducibility of response preferences (expressed as *t*-values) by dividing the data set into odd and even runs and computing the stimulus preference for each voxel separately for each half (Haynes & Rees, 2005). Significant correlations between the response biases in odd and even runs were found in all participants for visible (average correlation coefficient across subjects for FFA: $r = 0.31 ± 0.05$ *SEM*, $t(4) = 6.1$, $p = 0.004$; PPA $r = 0.30 ± 0.03$, $t(4) = 10.0$, $p = 0.001$) and remarkably also for invisible stimuli (FFA: $r = 0.21 ± 0.02$ *SEM*, $t(4) = 9.9$, $p = 0.001$; PPA $r = 0.25 ± 0.04$, $t(4) = 5.9$, $p = 0.004$). The weak and overall non-significant responses to invisible stimuli are thus similarly stable over time as those to visible stimuli and therefore also allow for above-chance prediction of stimulus category using MVPA.

Cross-classification

Finally, we explored the relationship between response patterns to visible and those to invisible stimuli. First, we tested whether each voxel's response preferences for visible face or house stimuli were retained when stimuli were suppressed from conscious awareness. However, neither in the FFA nor in the PPA did voxel-wise response
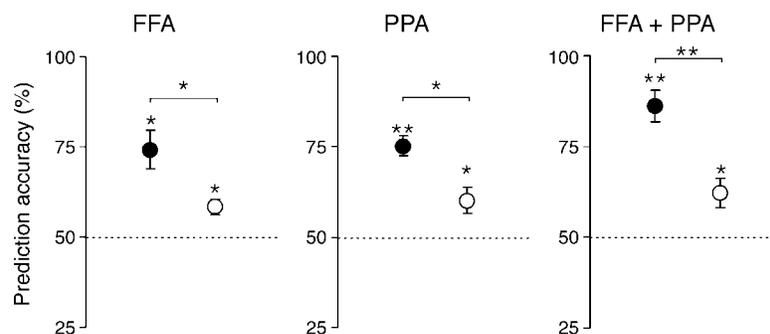

Figure 4. Performance of support-vector-machine (SVM) classifiers for pairwise classification of face and house presentations from fusiform face area (FFA), parahippocampal place area (PPA), and from FFA and PPA together. Average prediction accuracies across participants ($n = 5$) for visible faces vs. houses are denoted by filled circles (±*SEM*) and for invisible faces vs. houses by empty circles. The dotted lines denote chance level (50%). *Note*: *$p < 0.05$; **$p < 0.005$.
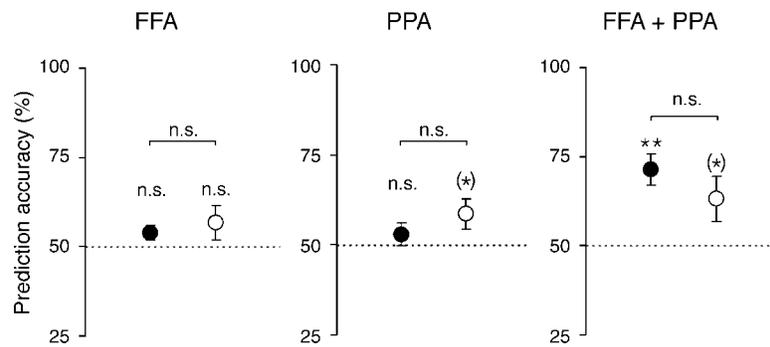
Figure 5. Performance of support-vector-machine (SVM) classifiers for cross-classification of invisible stimulus presentations after training on response patterns to visible stimuli, and *vice versa,* from fusiform face area (FFA), parahippocampal place area (PPA), and from FFA and PPA together. Average prediction accuracies across participants (*n* = 5) for prediction of invisible faces vs. houses after training on response patterns to visible stimuli are denoted by filled circles (±*SEM*) and for prediction of visible faces vs. houses after training on response patterns to invisible stimuli by empty circles. *Note*: **p* < 0.1; ***p* < 0.01; n.s. = non-significant (*p* > 0.1).

preferences consistently correlate between the visible and invisible conditions (average correlation coefficient across subjects for FFA: *r* = 0.09 ± 0.04 *SEM*, *t*(4) = 2.6, *p* = 0.06; PPA: 0.03 ± 0.04, *t*(4) = 0.9, *p* = 0.40). In line with the absence of significant cross-correlations, cross-classification, i.e., training on visible conditions and testing on invisible conditions, and *vice versa,* did not result in significant prediction when using voxels from either FFA or PPA (FFA: 53.8% ± 2.0 *SEM*, *t*(4) = 1.9, *p* = 0.13, and 56.7% ± 4.8 *SEM*, *t*(4) = 1.4, *p* = 0.23, respectively; PPA: 52.9% ± 3.2 *SEM*, *t*(4) = 0.9, *p* = 0.41, and 58.6% ± 4.1 *SEM*, *t*(4) = 2.1, *p* = 0.10, respectively; see Figure 5). However, when FFA and PPA voxels were pooled for cross-classification, prediction accuracy rose above chance level for training on visible and testing on invisible stimuli (71.3%, *t*(4) = 4.9, *p* = 0.008) and also showed a trend for the opposite direction (63.1%, *t*(4) = 2.1, *p* = 0.10). While there is thus generally little agreement between voxel preferences for visible and invisible stimuli, at least some voxels seem to retain their response preferences independently of visibility.

Multivariate analyses in control regions

To test whether our finding that MVPA can differentiate between face- and house-related activity patterns was specific to FFA and PPA, we additionally performed MVPA in three control regions, the ventromedial pre-frontal cortex, retinotopic visual cortex, and the occipital face area (Figure 6).

First, we wanted to rule out the possibility of a general bias toward above-chance classification related to the SVM algorithm used. To that end, we chose a control region in ventromedial prefrontal cortex that was equiv-alent in size to the FFA and PPA ROIs (see Methods section for details of selection) and that was not expected to be directly involved in processing of visual object categories. As expected, univariate analysis did not reveal any overall activity differences between visible faces and houses (*t*(4) = 0.02, *p* = 0.99) or between invisible faces and houses (*t*(4) = 0.13, *p* = 0.90). Likewise, MVPA did not permit correct classification of visible (55.5%, *t*(4) = 1.4, *p* = 0.22) nor of invisible stimuli (50.7%, *t*(4) = −0.2,
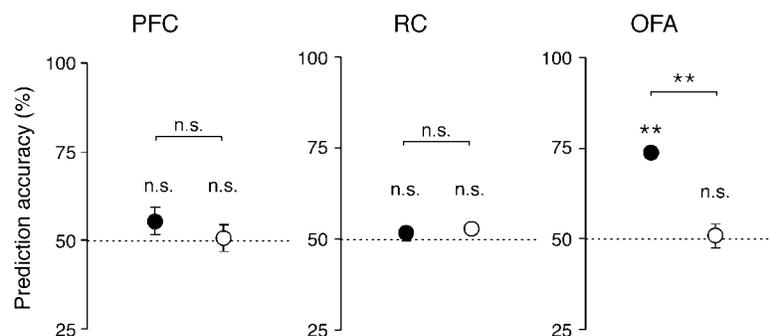

Figure 6. Performance of support-vector-machine (SVM) classifiers for pairwise classification of face and house presentations from control regions in prefrontal cortex (PFC), retinotopic cortex (RC), and from the occipital face area (OFA). Average prediction accuracies across participants (*n* = 5) for visible faces vs. houses are denoted by filled circles (±*SEM*) and for invisible faces vs. houses by empty circles. The dotted lines denote chance level (50%). *Note*: ***p* < 0.005; n.s. = non-significant (*p* > 0.1).

$p = 0.87$) from this prefrontal region, excluding a general bias in the SVM algorithm (see Figure 6).

Second, we explored the possibility that category-specific activity patterns in FFA and PPA might reflect processing of lower level visual features, i.e., the observed response patterns might have reflected propagated activity from low-level stimulus representations. We therefore probed MVPA performance in retinotopic visual cortex. As the analysis of retinotopic visual cortex responses had not been the primary goal of our study, we used the response maxima in the occipital pole region that were commonly activated by all stimulus types as an approximation to the stimulus representation in retinotopic visual cortex (Williams, Dang, & Kanwisher, 2007). Univariate analysis did not show any overall response differences between visible faces and houses ($t(4) = -0.91$, $p = 0.39$) or between invisible faces and houses ($t(4) = -0.14$, $p = 0.19$). Again, MVPA was neither able to differentiate between visible (51.6%, $t(4) = 1.4$, $p = 0.22$) nor between invisible face and house stimuli (50.7%, $t(4) = -0.2$, $p = 0.87$, see Figure 6).

Finally, we examined the pattern information at an intermediate level of cortical visual object processing, in the OFA. This region is located posterior to the FFA and also shows stronger responses to faces than to other object stimuli but is thought to process more the physical aspects of faces rather than more complex aspects such as configuration and identity (Kanwisher & Yovel, 2006). Similar to the FFA, overall OFA responses to visible faces were significantly greater than to visible houses ($t(4) = 3.1$, $p = 0.015$), whereas no such difference was present for invisible stimuli ($t(4) = 0.45$, $p = 0.67$). MVPA also yielded above-chance classification of OFA response patterns to visible faces vs. houses ($t(4) = 15.6$, $p < 0.001$, see Figure 6). However, MVPA performed at chance level for classification of invisible faces vs. houses ($t(4) = 0.25$, $p = 0.82$), in contrast to the FFA, where MVPA classification was significantly better than chance also for both visible and invisible stimuli (see above).

Discussion

Our data demonstrate that activity patterns in the FFA and the PPA differentiate two categories of object stimuli (faces and houses) even when the stimuli are rendered invisible by interocular suppression. Importantly, this information about invisible stimuli could only be retrieved when MVPA was used to take the fine-scale activity pattern into account. In contrast, differential responses to invisible stimuli could not be identified using standard univariate techniques alone. Other studies found stimulus- or category-specific responses to invisible stimuli in higher level visual areas even with univariate analyses (Dehaene et al., 2001; Moutoussis & Zeki, 2002) but had

used other means than binocular rivalry to render visual stimuli invisible. In contrast, previous studies explicitly addressing the representation of stimuli suppressed by binocular rivalry either failed to find such responses (Pasley et al., 2004; Williams et al., 2004) or they found activations in higher level regions of the dorsal stream and to a lesser extent in ventral visual areas, but without showing category-specificity in the latter (Fang & He, 2005; Jiang & He, 2006). Along with the results of our rigorous assessment of awareness, our failure to detect category-specific responses with various univariate methods indicates that the CFS technique used in our study was highly effective. Remarkably, even under these conditions of most profound interocular suppression there was still sufficient information present in the fMRI signal from high-level ventral visual cortex to predict the category of invisible stimuli with above-chance accuracy. Our results thus demonstrate that the absence of a differential signal when averaging across activity in a functionally defined brain region cannot be taken as evidence for the absence of neural selectivity. This is in line with a recent study investigating fMRI activity patterns in the lateral occipital complex where the spatial response pattern, but not the mean response across voxels, contained information about object categories (Williams et al., 2007). Interestingly, pattern information in the lateral occipital cortex is only present when the category to which a visual object belongs is correctly identified compared to trials with incorrect categorization, which is compatible with our result of higher classification accuracy for visible stimuli. In contrast, this earlier finding that activity patterns in lateral occipital cortex did not contain information about stimuli that were not consciously perceived (and hence not assigned to the correct category) is in apparent contradiction with our finding of category-specific pattern representations of invisible stimuli. However, some differences between this previous report and our current study should be noted. The earlier study investigated the information content of fMRI activation patterns as a function of behavioral performance. The focus of our present study was fundamentally different in that we asked whether we could detect any category specific information in brain responses to stimuli that were completely and reliably suppressed from awareness. The previous study used backward masking of briefly presented (33 ms) stimuli while stimuli in our study were presented for 600 ms and rendered invisible by CFS. Perhaps most importantly, we used ecologically relevant stimulus categories which are known to be differentially represented in the human visual system, while the earlier study used abstract object stimuli, the representations of which are likely to be less distinct.

We sampled neuronal responses from large populations using BOLD contrast fMRI, which is correlated both to the spiking activity of single neurons and to local field potentials (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001). While our findings are therefore consistent with a

representation of the suppressed stimulus at the level of single neurons, they do not fully constrain the precise physiological form of such a representation. Electrophysiological studies have shown that most neurons in monkey IT (Sheinberg & Logothetis, 1997) and human medial temporal lobe (Kreiman et al., 2002) are active only during conscious perception of their preferred stimulus. However, a small percentage of neurons in IT are not modulated by perception (Sheinberg & Logothetis, 1997) and such neurons may have contributed to category-specific patterns observed in our study. Moreover, it is also conceivable that other classes of neurons, such as the small (and difficult to record) inhibitory interneurons may also carry information about the suppressed stimulus and contribute to category-specific BOLD activity patterns. Irrespective of the exact neuronal processes that underlie our current findings, they provide a potential physiological basis for how neurons at high-level processing stages that represent suppressed category-specific information could engage in competitive interactions with neurons coding for the currently dominant percept (Blake & Logothetis, 2002; Tong et al., 2006; Wilson, 2003).

An important question raised by our results in FFA and PPA is whether the category-specific pattern signals we observed could reflect processing of lower level visual stimulus features. However, we found that activity patterns in retinotopic visual cortex neither allowed SVM classification of visible nor of invisible face vs. house stimuli. It should be noted that our experimental paradigm and scanning procedure was not originally designed to assess response patterns in early visual cortex and we cannot therefore attribute the patterns to specific retinotopic areas. However, irrespective of the exact retinotopic area, this finding speaks against the general possibility that the FFA and PPA response patterns could merely reflect activity propagated from low-level stimulus representations. Interestingly, activity patterns in the OFA allowed SVM classification of visible but not of invisible face vs. house stimuli. This suggests that the cortical representation of complex visual information that is binocularly suppressed may be limited to higher level visual areas. Such regions may thus be preferentially involved in mediating the influence of complex stimulus information on the resolution of binocular conflict.

Our detailed voxel-wise analyses of response preferences revealed only little agreement in the spatially distributed BOLD activity for visible and invisible stimuli. Accordingly, cross-classification across visibility conditions was only successful when pattern information was maximized by pooling FFA and PPA voxels and may thus have relied on a few voxels that carry category information irrespective of visibility (Sheinberg & Logothetis, 1997). The absence of a strong relationship between response patterns to visible and invisible stimuli may be due to the fact that considerably less information is present in response patterns to invisible stimuli, as indicated by significantly lower prediction accuracies.

Visible stimuli might drive responsive neurons more effectively, leading to a broader spatial spread of BOLD signal across voxels and making the response pattern more uniform. However, the information contained in distributed response patterns may differ not only quantitatively but also qualitatively, e.g., due to the processing of the masking stimuli which were only present in the invisible conditions. Moreover, at a fine spatial scale, binocular suppression might modify the category-specific preferences of individual neurons, giving rise to different patterns of biased sampling at the larger spatial scale of the BOLD signal. Finally, stimulus preferences of individual neurons might remain unchanged but additional neural signals reflecting awareness, e.g., recurrent processing (Lamme & Roelfsema, 2000), might change the overall pattern of biased sampling.

Conclusions

Our data show how information about a visual stimulus can be extracted from the spatial pattern of responses in higher visual cortex even when no differential signal is detectable using conventional analyses. Brain regions that show largely homogenous response properties during awareness of a visual stimulus can thus display altered but nevertheless informative response patterns during unawareness. Our results support the view that complex category-specific information from binocularly suppressed stimuli is represented at high levels of the visual processing hierarchy and provide a physiological basis for the contribution of such complex information to the resolution of binocular conflict.

Acknowledgments

Commercial relationships: none.
Corresponding author: Dr. Philipp Sterzer.
Email: philipp.sterzer@charite.de.
Address: Department of Psychiatry, Charité Campus Mitte, Charitéplatz 1, D-10117 Berlin, Germany.

References

Alais, D., & Parker, A. (2006). Independent binocular rivalry processes for motion and form. *Neuron, 52,* 911–920. [PubMed] [Article]

Andersson, J. L., Hutton, C., Ashburner, J., Turner, R., & Friston, K. (2001). Modeling geometric deformations in EPI time series. *Neuroimage, 13,* 903–919. [PubMed]

Andrews, T. J., & Blakemore, C. (1999). Form and motion have independent access to consciousness. *Nature Neuroscience, 2,* 405–406. [PubMed]

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews, Neuroscience, 3,* 13–21. [PubMed]

Christianini, N., & Shawe-Taylor, J. (2000). *Support vector machines and other kernel-based learning methods*. Cambridge: Cambridge University Press.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience, 4,* 752–758. [PubMed]

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron, 23,* 115–125. [PubMed] [Article]

Fang, F., & He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nature Neuroscience, 8,* 1380–1385. [PubMed]

Fu, C. H., Mourao-Miranda, J., Costafreda, S. G., Khanna, A., Marquand, A. F., Williams, S. C., et al. (2008). Pattern classification of sad facial processing: Toward the development of neurobiological markers in depression. *Biological Psychiatry, 63,* 656–662. [PubMed]

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430. [PubMed]

Haynes, J. D., Deichmann, R., & Rees, G. (2005). Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. *Nature, 438,* 496–499. [PubMed] [Article]

Haynes, J. D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience, 8,* 686–691. [PubMed]

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews, Neuroscience, 7,* 523–534. [PubMed]

Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 17048–17052. [PubMed] [Article]

Jiang, Y., Costello, P., & He, S. (2007). Processing of invisible stimuli: Advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychological Science, 18,* 349–355. [PubMed]

Jiang, Y., & He, S. (2006). Cortical responses to invisible faces: Dissociating subsystems for facial-information processing. *Current Biology, 16,* 2023–2029. [PubMed] [Article]

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8,* 679–685. [PubMed] [Article]

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311. [PubMed] [Article]

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 361,* 2109–2128. [PubMed] [Article]

Kovács, I., Papathomas, T. V., Yang, M., & Féher, A. (1996). When the brain changes its mind: Interocular grouping during binocular rivalry. *Proceedings of the National Academy of Sciences of the United States of America, 93,* 15508–15511. [PubMed] [Article]

Kreiman, G., Fried, I., & Koch, C. (2002). Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 8378–8383. [PubMed] [Article]

Kriegeskorte, N., Goebel, R., & Bandettini, P. A. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 3863–3868. [PubMed] [Article]

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23,* 571–579. [PubMed]

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412,* 150–157. [PubMed]

Macmillan, N. A., & Creelman, C. D. (1991). *Signal detection theory: A user's guide*. Cambridge: Cambridge University Press.

Moutoussis, K., & Zeki, S. (2002). The relationship between cortical activation and perception investigated with invisible stimuli. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 9527–9532. [PubMed] [Article]

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern

analysis of fMRI data. *Trends in Cognitive Sciences, 10,* 424–430. [PubMed]

Pasley, B. N., Mayes, L. C., & Schultz, R. T. (2004). Subcortical discrimination of unperceived objects during binocular rivalry. *Neuron, 42,* 163–172. [PubMed] [Article]

Pessoa, L. (2005). To what extent are emotional visual stimuli processed without attention and awareness? *Current Opinion in Neurobiology, 15,* 188–196. [PubMed]

Sheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Sciences of the United States of America, 94,* 3408–3413. [PubMed] [Article]

Sterzer, P., & Rees, G. (2008). A neural basis for percept stabilization in binocular rivalry. *Journal of Cognitive Neuroscience, 20,* 389–399. [PubMed]

Tong, F., & Engel, S. A. (2001). Interocular rivalry revealed in the human cortical blind-spot representation. *Nature, 411,* 195–199. [PubMed]

Tong, F., Meng, M., & Blake, R. (2006). Neural bases of binocular rivalry. *Trends in Cognitive Sciences, 10,* 502–511. [PubMed]

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron, 21,* 753–759. [PubMed] [Article]

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience, 8,* 1096–1101. [PubMed]

Vapnik, V. N. (1995). *The nature of statistical learning theory.* New York: Springer.

Williams, M. A., Dang, S., & Kanwisher, N. G. (2007). Only some spatial patterns of fMRI response are read out in task performance. *Nature Neuroscience, 10,* 685–686. [PubMed]

Williams, M. A., Morris, A. P., McGlone, F., Abbott, D. F., & Mattingley, J. B. (2004). Amygdala responses to fearful and happy facial expressions under conditions of binocular suppression. *Journal of Neuroscience, 24,* 2898–2904. [PubMed] [Article]

Wilson, H. R. (2003). Computational evidence for a rivalry hierarchy in vision. *Proceedings of the National Academy of Sciences of the United States of America, 100,* 14499–14503. [PubMed] [Article]

Wunderlich, K., Schneider, K. A., & Kastner, S. (2005). Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nature Neuroscience, 8,* 1595–1602. [PubMed] [Article]