Cerebral Cortex April 2006;16:500-508
doi:10.1093/cercor/bhi129
Advance Access publication July 13, 2005

Mistaking a House for a Face: Neural Correlates of Misperception in Healthy Humans

Christopher Summerfield[1], Tobias Egner[2], Jennifer Mangels[1] and Joy Hirsch[2]

[1]Department of Psychology, Columbia University, 406 Schermerhorn Hall, 1190 Amsterdam Ave, New York, NY 10027, USA and [2]Functional MRI Research Center, Department of Psychiatry and Radiology, Columbia University Neurological Institute Box 108 710 West 168th Street, New York, NY 10032, USA

Individuals with normal vision can sometimes momentarily mistake one object for another. In this functional magnetic resonance imaging study, we investigated how extrastriate visual regions respond during these erroneous perceptual judgements. Subjects were asked to discriminate images of houses and faces that were degraded such that they were close to an individually defined threshold for perception. On correct trials, voxels localized on the inferior occipital (OFA), fusiform (FFA) and parahippocampal (PPA) gyri exhibited selectivity for face and house images as expected. On incorrect trials, no face- or place-selectivity was observed for OFA or PPA. However, consistent with 'predictive coding' accounts of perception, we observed that the FFA also responded robustly on trials where a house was misperceived as a face, and concurrent activation was observed in medio-frontal and right parietal regions previously implicated in decision making under uncertainty. We suggest that FFA responses during misperception may be driven by a predictive top-down signal from these regions.

Keywords: extrastriate, misperception, neural correlates predictive coding, visual regions

Introduction

Recent approaches to perception have drawn upon the theory that part of the problem of deciding what it is that we are seeing can be solved before the stimulus is even presented. There exist intrinsic regularities in ongoing perception, such that, for example, each time you walk through the front door of your apartment, the configuration of objects, textures and colours which greets you is likely to be highly similar to the last. Most prominent among these regularities is a powerful temporal autocorrelation in the perceptual signal (what you are seeing now, it is very likely that you will be seeing in a few seconds' time). According to recent models, the brain can capitalize upon these regularities to generate, over time, a predictive code corresponding to perceptual events which are likely to occur (Mumford, 1992; Rao and Ballard, 1999; Bar, 2003; Friston, 2003; Murray *et al.*, 2004). The role of such a predictive signal would be to transfer part of the computational burden to the epoch preceding the stimulus, thereby limiting post-stimulus processing to the testing of a pre-established 'prior' hypothesis (and the further processing of residual prediction error). According to this view, once a stimulus has been presented, bottom-up sensory information is 'matched' to a predictive code rather than being processed *de novo* in feedforward succession. In other fields of psychology, such as reward learning, the existence of such predictive signals is well established, and it has been shown that reward-related neural activity will shift from reinforcer to predictive cue over the course of repeated pairings (Tobler *et al.*, 2005).

It has yet to be empirically demonstrated that the visual system is testing pre-established hypotheses in a Bayesian fashion. However, aside from its intuitive appeal, there exists considerable circumstantial evidence that predictive coding may be occurring. Firstly, the context of a visual event is highly influential in shaping perception decisions (Palmer, 1975; Biederman *et al.*, 1982; Henderson and Hollingworth, 1999; Bar, 2004). For example, a mailbox can be perceived as a loaf of bread if the context provided by surrounding objects indicates that it is to be found in the kitchen (Palmer, 1975). With respect to pre-stimulus code generation, recent evidence suggests that the brain is far from silent in the period preceding stimulation; rather, there is a tendency for neural synchrony to increase prior to onset of an expected stimulus (Brunia and Damen, 1988; Engel *et al.*, 2001; Tallon-Baudry *et al.*, 2005). Modelling work has suggested that feedback within a hierarchically organized system in which only prediction error is transferred between layers can account for the response properties of simple and complex cells in early visual cortex (Rao and Ballard, 1999). Moreover, one of the predictions of the model — that V1 activity will be suppressed when there is a good 'explanation' for the sensory data — has been borne out in functional magnetic resonance imaging (fMRI) studies in which subjects view matched gestalt and non-gestalt stimuli (Murray *et al.*, 2004). Predictive coding offers a framework for understanding a wide range of neurocognitive phenomena, including repetition suppression (Dolan *et al.*, 1997; Ishai *et al.*, 2004), change blindness (Rensink, 2000), vision occurring in 'reverse' (Ahissar and Hochstein, 2004), visual context effects (Bar, 2004) and perceptual hysteresis (Kleinschmidt *et al.*, 2002), as well as patterns of effective connectivity observed in the visual cortex (Pascual-Leone and Walsh, 2001).

Under normal viewing conditions, where perceptual information is rich and the visual environment is regular, there is little reason why perception should err. However, where visual information is limited (such as in the dark) this may not be the case. It follows from predictive coding accounts that perceptual errors (or 'misperceptions') may occur when higher-order visual regions incorrectly 'explain' impoverished information arriving via feedforward pathways from early visual regions. In other words, where the bottom-up visual signal is ambiguous but the top-down signal is strong (and wrong), the latter may gain precedence over the former, resulting in the generation of a false or erroneous percept. Interestingly, this view is highly reminiscent of recent models of hallucinatory or illusory experience in patient populations, which have described a mismatch between top-down and bottom-up information as crucial to non-veridical perceptual experiences (Grossberg, 2000; Collerton *et al.*, 2005). The phenomenon of 'misperceiving'

© The Author 2005. Published by Oxford University Press. All rights reserved.
For permissions, please e-mail: journals.permissions@oupjournals.org

one object as another is common to many neurological and psychiatric disorders (ffytche and Howard, 1999), as well as patients with damage to the posterior brain (Warrington and Shallice, 1984), but even healthy individuals frequently report 'illusory' perceptual experiences (McKellar, 1957). These misperceptions are most likely to occur when visual information is limited, such as at night or in the darkened interior of a room (Murgatroyd and Prettyman, 2001), as would be expected if such errors were due to erroneous 'explanation' of a weakened visual signal.

In the present study, we studied how the brain responds during erroneous perceptual decisions, with a view to understanding more about how predictive signals shape perception. It follows from predictive coding accounts of perception that on incorrect trials (where bottom-up information is weak or ambiguous), neural activity will be observed in visual regions tuned to the predicted stimulus, as well as more anterior structures from which the top-down 'prediction' originates. For example, where stimulus A is mistaken for ('explained as') stimulus B, extrastriate visual regions representing stimulus B will become active despite the fact that no bottom-up information corresponding to stimulus B has been presented. In other words, if predictive coding is occurring, then selectivity for the reported percept should be preserved during erroneous discrimination.

To test this hypothesis, we used a challenging perceptual task in which subjects discriminated rapid, visually degraded images of faces and houses. Discriminability was carefully controlled with psychophysical thresholding, such that all images were presented very close to individual thresholds for discrimination. fMRI responses were acquired from ventral posterior cortical regions known to respond preferentially to the face (fusiform and inferior occipital gyri) and house (parahippocampal gyrus) images used in our discrimination task (Dolan *et al.*, 1997; Kanwisher *et al.*, 1997; Aguirre *et al.*, 1998; Epstein and Kanwisher, 1998). By exploring how these regions responded on incorrect discrimination trials (for example, when a house is mistaken for a face), it was possible to assess whether perceptual selectivity is preserved under situations where one object is mistaken for another.

Materials and Methods

Subjects

Subjects (n = 8, four females) were neurologically normal individuals ranging in age from 19 to 34 years. All subjects gave informed consent in accordance with Columbia University Medical Center Institutional Review Board guidelines.

Stimuli

Stimuli were 400 × 400 pixel grayscale images of faces and houses. Faces came from the AR database (Martinez and Benavente, 1998) and houses were from photographs taken by the authors in Brooklyn, New York. Images were cropped within the borders of the face/house such that features (eyes/nose, or windows/door) were more prominent than overall shape (see Fig. 1*a*). House stimuli were additionally smoothed with a 2D Gaussian filter (width = 11 pixels) to match house and face stimuli for high-frequency information. All stimuli were normalized to a mean luminance of 0.5 (range 0–1). For psychophysical testing, contrast-modulated exemplars of the house/face images were generated by scaling luminance values by c within a range generated by 0.5 ± [c/2]. For example, 0.2 contrast images were generated by scaling all luminance values in the range 0.4–0.6. Mask stimuli were random checkerboards (400 × 400 pixels, each square 40 × 40 pixels) which were smoothed and deformed by the 'spherize' and 'ocean ripple' tools in Adobe Photoshop. Examples of contrast-modulated face and house images, and masks, can be seen in Figure 1*a*.

Localizer Task

In a pre-experimental fMRI run, subjects passively viewed 12 alternating blocks of unmodulated, unmasked faces and houses. House/face stimuli were presented for 750 ms with 250 ms inter-stimulus intervals, in blocks of 15 consecutive stimuli. A 10 s rest period (fixation) was interleaved between blocks.

Discrimination Task

In the discrimination task, each trial began with a blank screen for a variable period. This duration was varied in a Gaussian fashion such that total inter-trial interval varied between 2000 and 3000 ms. A fixation cross for 500 ms cued the onset of the stimulus. The face/house image was presented for 100 ms, followed immediately by a randomly selected mask for 250 ms. On each trial, the subject indicated with a button press whether the stimulus was a face or a house, and whether they had high or low confidence in their response. The mask duration was 250 ms, after which it was replaced by a fixation cross for 1000 ms. The sequence of events on each discrimination trial can be seen in Figure 1*b*.

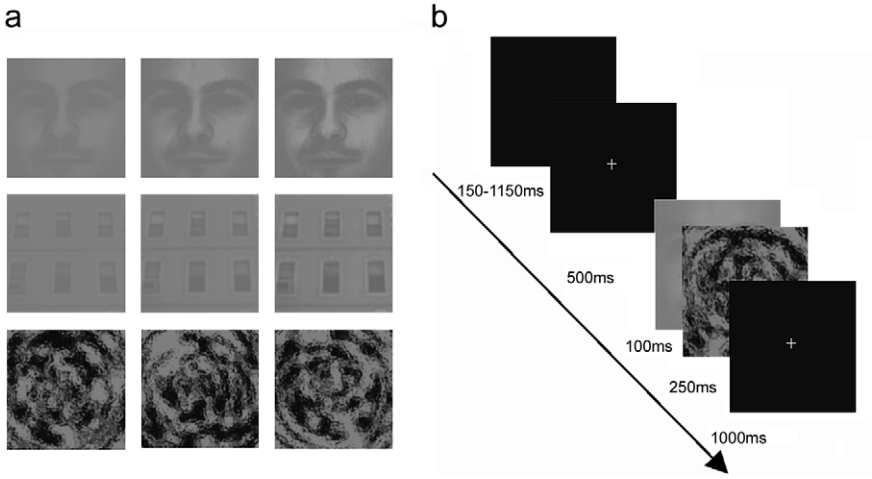

Figure 1. Example stimuli. (*a*) Three examples of contrast-modulated face and house images, and example masks (bottom). Images were contrast modulated by normalizing scalar intensity values within a range around 0.5. Face and house exemplars increase in contrast from left to right. (*b*) The sequence of stimuli presented in each trial. Inter-stimulus intervals are shown interposed between frames.

A genetic algorithm was used to generate pseudorandom sequences of faces/houses which were optimized for detection of contrast between trial types (Wager and Nichols, 2003). Levels of contrast modulation were determined individually for each subject with extensive pre-experimental testing. Subjects performed 540 practice trials (180 outside the scanner, 360 in the scanner) on which contrast modulation varied from 0.03 to 0.3 (in steps of 0.03) in an ascending and descending staircase fashion. Each subject's point of subjective equality (PSE) was defined as the contrast level closest to which she exhibited 75% discrimination performance. During the main task, 10 stimuli were drawn from a Gaussian distribution of contrast-modulated images falling within 0.05 of this level in steps of 0.01.

fMRI Data Acquisition

Images were acquired with a GE Twin-Speed 1.5 T scanner. All images were acquired parallel to the AC-PC orientation with a T2*-weighted EPI sequence of 24 contiguous axial slices [T_R = 2000, T_E = 40, flip angle = 60°, field of view (FoV) = 190 mm, array size = 64 × 64″] of 4.5 mm thickness and 3 × 3 mm in-plane resolution, providing whole-brain coverage. The region of interest (ROI) localizer task consisted of a single run of 155 scans, and the discrimination task consisted of four runs of 160 scans each. High-resolution anatomical scans were acquired with a T1*-weighted SPGR sequence (T_R = 19, T_E = 5, flip angle = 20, FoV = 220), recording 24 slices at a slice thickness of 1.5 mm and in-plane resolution of 0.86 × 0.86 mm.

fMRI Data Analysis

Spatial pre-processing and statistical mapping were carried out with SPM2 software (Wellcome Department of Imaging Neuroscience, University College London, UK, http://www.fil.ion.ucl.ac.uk/spm/spm2.html). Functional T2* images were slice-timing corrected, spatially realigned to the first volume acquired. The first five functional scans from each task were discarded prior to the subsequent analyses. A 128 s temporal highpass filter was applied in order to exclude low-frequency artifacts. Temporal correlations were estimated using restricted maximum likelihood estimates of variance components using a first-order autoregressive model. The resulting non-sphericity was used to form maximum likelihood estimates of the activations. Each subject's structural T1 image was co-registered to an individual mean EPI image. Transformation parameters were derived from normalizing the co-registered structural image to a template brain within the stereotactic space of the Montreal Neurological Institute (MNI), and the derived parameters were then applied to normalize each subject's EPI volumes (from both localizer and task runs). Normalized images were smoothed with a Gaussian kernel of 9 × 9 × 13.5 mm full-width half-maximum (i.e. three times the voxel dimensions as originally acquired).

Trial Classification

For the discrimination task, trials were classified according to whether a face was correctly perceived as a face (FF), a face was incorrectly perceived as a house (FH), a house was incorrectly perceived as a face (HF) or a house was correctly perceived as a house (HH). Correct responses were further subdivided into high (FFhc, HHhc) and low confidence (FFlc, HHlc) options (this subdivision was not possible for incorrect trials, for which the high confidence response was very rarely used), yielding a total of six conditions of interest. Regressors of stimulus events in the discrimination task (convolved with a canonical HRF) were created for each trial type (FFhc, HHhc, FFlc, HHlc, FH, HF), and parametric modulation regressors for image contrast were included with each regressor. Parametric regressors were built using values taken from subject-specific performance curves (% correct faces/houses at each contrast level) rather than values corresponding to the degree of stimulus degradation. Subject-specific parameter estimates associated with each of these six regressors were extracted from the first (within-subject) analysis. These estimates (beta coefficients) were then compared with *t*-tests at the second (group) level for analyses specific to predefined ROIs and in a voxelwise fashion across the entire brain.

Localization Face- and Place-responsive Regions

Selective face- and place-sensitive voxels were identified with a pre-experimental localizer task in which subjects passively viewed face and place stimuli in alternating blocks. Imaging data from this localizer task was modeled with two box-car functions convolved with a canonical hemodynamic response function (HRF). These regressors were contrasted with a *t*-test in each subject (faces >places, places >faces) and the resulting images were thresholded at a liberal threshold ($P < 0.001$, uncorrected) to identify face- and place-sensitive regions of the brain. Guided by an extensive previous literature, we selected peak voxels responsive to face stimuli in the inferior occiptial gyrus (the 'occipital face area' or OFA) and fusiform gyrus ('the fusifom face area' or FFA), and voxels responsive to place stimuli in the parahippocampal gyrus (the 'parahippocampal place area' or PPA). For each region in each subject, we defined a sphere of 2 mm radius (8 voxels) centered on the voxel showing the peak response to the relevant comparison. Additionally, we defined a single control region of interest, also of 2 mm radius, at the peak voxel falling in early visual cortex which responded to both faces and places in group analysis of the localizer task. The talairach coordinate locations of these voxels (OFA, FFA, PPA) for each subject is shown in Table 1.

Statistical Analyses

In order to assess predictions of interest, subject-specific parameter estimates within regions of interest (PPA, FFA, OFA) were compared using *t*-tests at the group level. In particular, we were interested in how neural responses in pre-defined face- and place-sensitive regions varied on incorrect trials (HF > FH, FH > HF). Comparisons were also undertaken for the parametric modulation regressors, to assess whether that portion of the response of each region which varied with contrast similarly differed between trial types. Estimates of the hemodynamic response for each condition were obtained by refitting the data using a finite impulse response (FIR) convolution model to provide a less constrained picture of the hemodynamic response. Note that we only remodelled the data at these ROIs for which we had already established significant responses.

Additionally, we performed a conventional whole-brain search for voxels whose activation varied as a function of stimulus and percept for high-confidence correct, low-confidence correct and incorrect trials. For these analyses, which were performed at the second (between-subject) level, only voxels which were significant at $P < 0.05$ with the correction for false discovery rate (FDR) (Genovese *et al.*, 2002) are reported.

Results

Behavioral Results

Subjects' PSE (~75% discrimination) fell within contrast values of 0.08 and 0.26. Overall mean discrimination performance on the discrimination task was 71.1 ± 5% (Fig. 2*a*). Subjects were equally likely to correctly detect houses (d' = 1.79 ± 0.14) and faces (d' = 1.76 ± 0.16) and substantial numbers of both high-confidence correct (229.3 ± 45.5), low-confidence correct (110.0 ± 43.9) and incorrect (139.9 ± 24.0) trials were obtained for each subject. High-confidence incorrect responses were rare, with an average of 8.6 ± 8.7 faces classified with high

Table 1

Voxel locations (Talairach coordinate space) of peak voxel in the fusiform gyrus (FFA) and inferior occipital gyrus (OFA) which responded to the comparisons faces > houses (FFA, OFA) and houses > faces (PPA) in the localizer task

Subject	FFA voxel	OFA voxel	PPA voxel
1	40 −63 −11	36 −79 2	19 −50 −12
2	38 −67 −5	−32 −75 −7	24 −51 −12
3	36 −52 −9	40 −81 −5	20 −44 −18
4	−32 −63 −15	−43 −83 −9	−24 −52 −14
5	36 −62 −15	−41 −79 3	−28 −46 −14
6	−40 −65 −19	−43 −81 −11	17 −44 −13
7	41 −56 −16	−43 −85 −7	−25 −62 −15
8	−38 −69 −19	−45 −83 −11	24 −44 −12

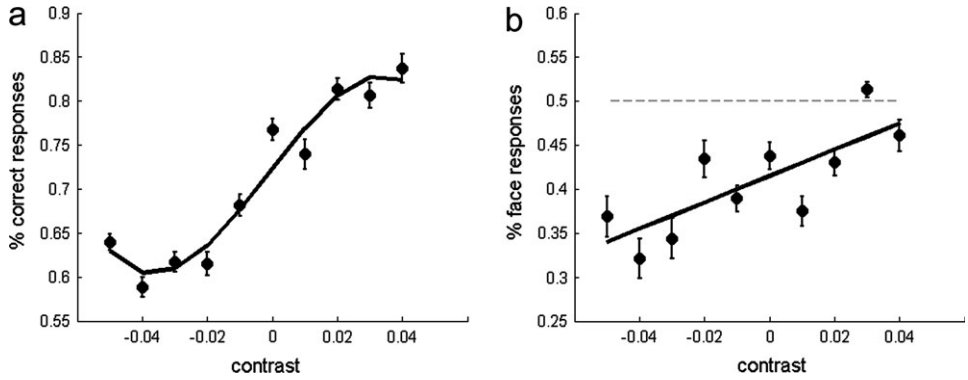

Figure 2. Behavioral data. (*a*) Psychophysical data from the experimental task (480 trials). Discrimination performance (0–1, chance = 0.5) is on the *y*-axis and level of contrast modulation (difference from PSE) is on the *x*-axis. A mean curve is fitted to the data (mean with standard error bars). (*b*) From the task, tendency to respond 'face' as a function of contrast level. The *x*-axis shows contrast level (difference from PSE); the *y*-axis shows the percentage of trials on which the subject responded 'face'. Fitted mean data (black line) is superimposed on mean data points with standard errors bars. No bias (0.5) is marked with a dashed grey line.

confidence as houses (FH), and 2.0 ± 3.0 houses classified with high confidence as faces (HF). In absolute number, more high-confidence incorrect trials were FH than HF ($t = 2.93$, $P < 0.04$), but proportionally, more high-confidence trials were HF than FH ($t = 2.56$, $P < 0.05$). Subjects displayed an overall bias to classify stimuli as houses ($t = 3.0$, $P < 0.03$). However, this tendency varied with contrast (Fig. 2*b*), with lower-contrast stimuli more likely to be classed as houses (linear trend: $F = 35.7$, $P < 0.001$).

Control Region

The location of the early visual control region is shown in Figure 3*a* (cluster thresholded at $P < 10^{-5}$). When imaging data from the discrimination task were extracted from a small, spherical ROI centered on the peak voxel from this ROI (indicated with blue cross-hairs), neural responses did not differ as a function of either stimulus or percept (FFhc > HHhc, $P = 0.47$; FFlc > HHlc, $P = 0.07$; HF > FH, $P = 0.23$). The estimated mean HRF across subjects for this ROI is shown in Figure 3*b*.

Face-responsive Regions

In the localizer task, all eight subjects showed activation in inferior occipital and fusiform regions in response to passive viewing of faces. Five subjects exhibited a peak fusiform area face-response in the right hemisphere, and three in the left hemisphere. The reverse pattern was observed in the OFA, with 6/8 subjects exhibiting maximal selectivity for faces in the left hemisphere. Table 1 shows the location of the OFA and FFA for each subject, as identified by the localizer task. All FFA peaks fell anterior to all OFA peaks (all FFA within $50 < y < 70$; all OFA within $70 < y < 90$). Additionally, the locations of selected FFA and OFA ROIs are shown rendered onto the MNI brain in Figure 4*a*,*d*.

Figure 4 shows results from the face-responsive regions (FFA, Fig. 4*a–c*; OFA, Fig. 4*d–f*). In Figure 4*b*, mean FFA responses from FFhc (blue lines) and HHhc trials (red lines) are plotted. Face-selectivity in individually defined ROIs (selected FFA ROIs shown in Fig. 4*a*) was highly preserved for these high-confidence correct trials, with robust and positive-going HRFs to FFhc trials. When compared at the group level, mean parameter estimates were significantly greater for FFhc trials than HHhc ($t = 2.7$, $P < 0.05$). However, our main comparison of interest concerned incorrect trials. Houses mistaken as faces elicited a reliably greater neural response in the FFA than faces mistaken as houses ($t = 3.1$, $P < 0.02$). HF trials (cyan line) were accompanied by a large, postitive-going hemodynamic response in the FFA (Fig. 4*c*) which peaked later (~8 s) than the response on correct trials. No differences were observed for low-confidence trials (red and blue lines, $P = 0.83$).

Figure 4*d* shows selected peak face-responsive voxels on the inferior occipital gyrus identified with the localizer task. As for the FFA, face-selectivity was preserved at these OFA voxels on the discrimination task, with a reliably greater neural response elicited on FFhc trials than HHhc trials ($t = 4.2$, $P < 0.01$). HRFs on FFhc (blue) and HHhc trials (red) can be seen in Figure 4*e*. Blood oxygen level-deficient (BOLD) responses did not differ, however, as a function of stimulus for low-confidence correct trials ($P = 0.48$) or incorrect trials ($P = 0.23$). HRFs for low-confidence and incorrect trials are shown in Figure 4*f*.

Descriptively, different patterns of results were observed in FFA and OFA, with face-selectivity preserved on incorrect trials for fusiform but not inferior occipital voxels. In order to test the statistical reliability of this result, we entered the beta coefficients from IOG and FFA ROIs into a 2 (region; FFA, OFA) × 2 (condition; FH, HF) analysis of variance. We observed a statistically significant region × condition interaction ($F = 9.4$, $P < 0.02$), indicating that face percept selectivity on incorrect trials was indeed observed for fusiform but not inferior occipital face-responsive voxels.

Place-responsive Regions

A sample of the individual PPA locations from which these data were extracted can be seen in Figure 5*a*. Comparing HHhc > FFhc trials at these individually defined ROIs located on the parahippocampal gyrus, BOLD responses were significantly greater for HHhc than for FFhc trials ($t = 2.5$, $P < 0.04$). Mean hemodynamic responses on high-confidence correct trials (HHhc, red; FFhc, blue) are shown in Figure 5*b*. In Figure 5*c*, the HRF estimates on other trial types are plotted. Within this cluster, low-confidence correct house trials (HHlc) also exhibited a reliably greater response than low-confidence correct face trials (FFlc) ($t = 5.2$, $P < 0.01$). The comparison between the two types of incorrect trial (HF > FH, FH > HF), however, failed to reach statistical significance ($t = 0.29$).

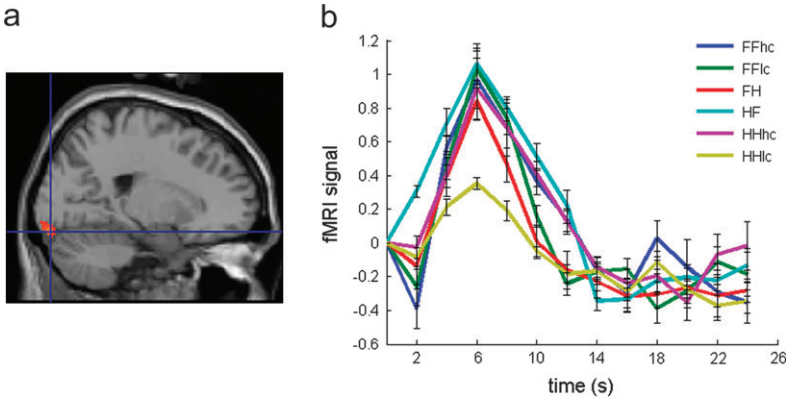

Figure 3. Imaging data: V1/V2 control region. (*a*) Voxels responding to both faces and houses in the localizer task (cluster thresholded at $P < 10^{-5}$). The peak voxel within this cluster is marked with the blue cross-hairs. (*b*) Discrimination-task hemodynamic responses (post-stimulus time histogram) from this peak voxel, for each of the six conditions. Time in seconds is on the *x*-axis. Bars represent standard errors.

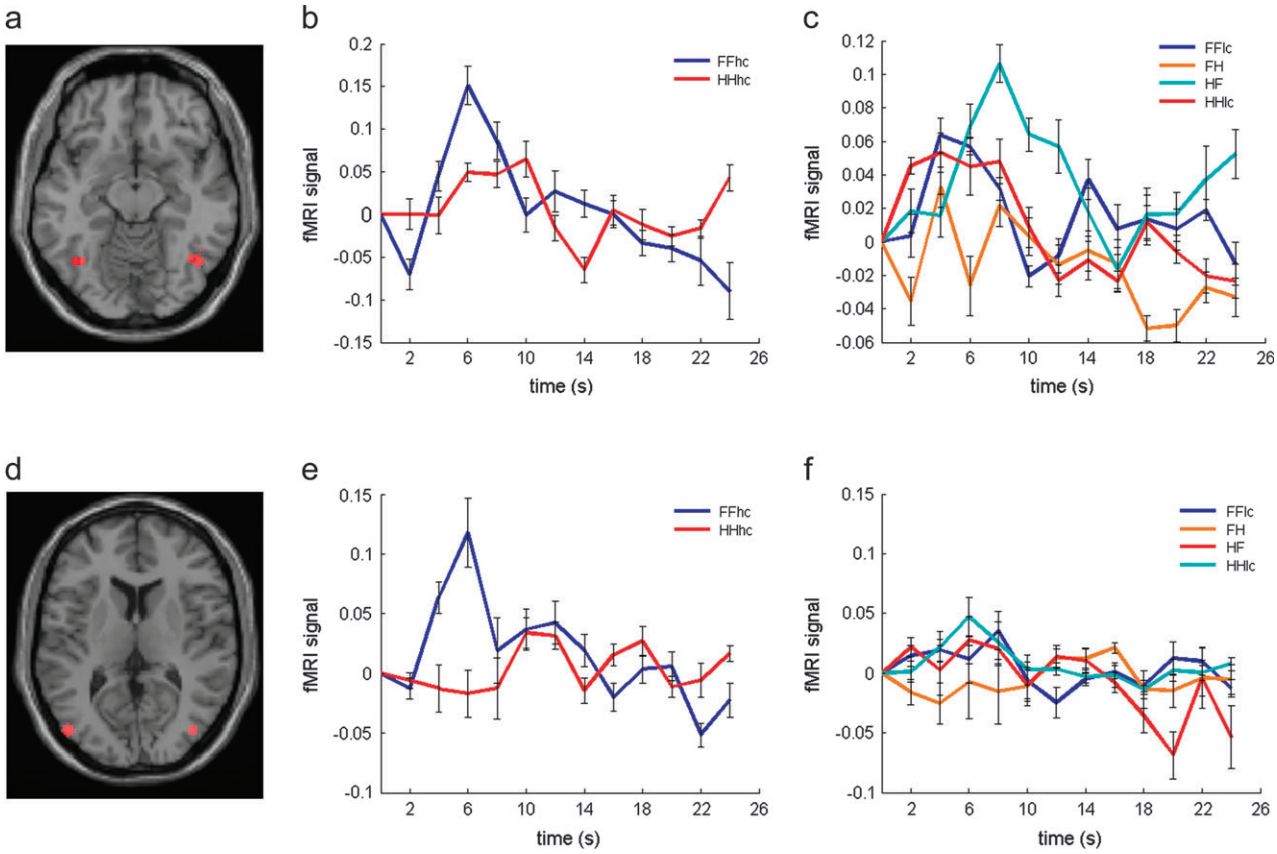

Figure 4. Imaging data: FFA and OFA. (*a*) Locations of peak fusiform gyrus (FFA) voxels responsive to faces > houses from the localizer task (selected individual subjects). (*b*) Mean discrimination-task hemodynamic responses from high-confidence correct face (blue) and high-confidence correct house (red) trials extracted from these fusiform loci. (*c*) Mean discrimination-task hemodynamic responses from low-confidence (red, blue) and incorrect (orange, cyan) trials. (*d*) Locations of peak inferior occipital gyrus (OFA) voxels responsive to faces > houses from the localizer task (selected individual subjects). (*e*) Mean discrimination-task hemodynamic responses from high-confidence correct face (blue) and high-confidence correct house (red) trials for these inferior occipital loci. (*c*) Discrimination-task hemodynamic responses from low-confidence (red, blue) and incorrect (orange, cyan) trials.

It was observed that, overall, beta values for incorrect and face-stimulus trials in the PPA were less than zero, and the HRF curves were accordingly negative-going. We reasoned that this might simply reflect parameter estimates falling below the mean, due to the capture of variance by the parametric contrast regressors, and so we re-ran the analysis without these regressors. A highly similar pattern of data was observed in the PPA, with HRF curves for face stimulus and incorrect trials dipping below zero in precisely the same fashion (data not shown).

Contrast-modulated Regressors

For each of the three regions of interest, we also compared the contrast-modulated regressors for high-confidence correct, low

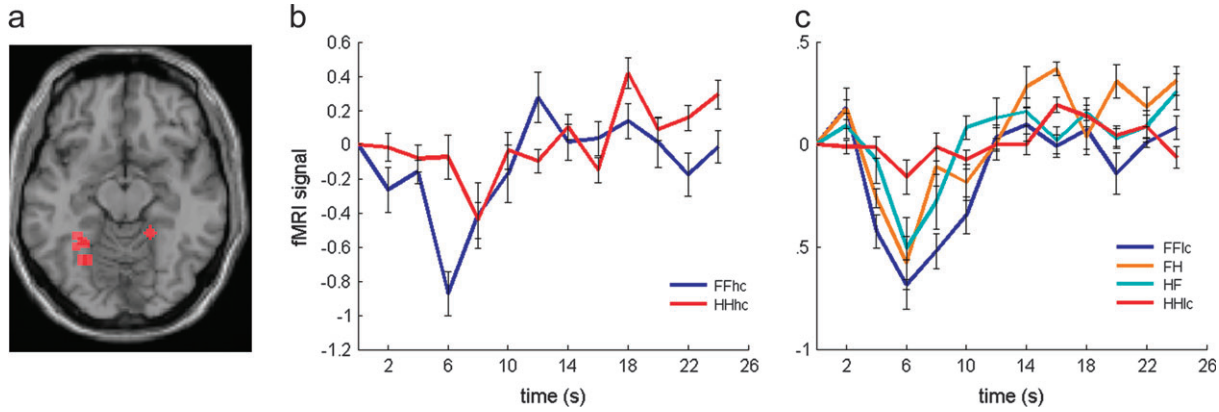

Figure 5. Imaging data: PPA. (*a*) locations of peak parahippocampal gyrus (PPA) voxels responsive to houses > faces from the localizer task (selected individual subjects). (*b*) Mean discrimination-task hemodynamic responses from high-confidence correct face (blue) and high-confidence correct house (red) trials extracted from these parahippocampal loci. (*c*) Mean discrimination-task hemodynamic responses from low-confidence (red, blue) and incorrect (orange, cyan) trials.

confidence correct, and incorrect trials. None of the comparisons reached statistical threshold (V1: all *P*-values > 0.3; PPA: all *P*-values > 0.2; FFA: all *P*-values > 0.1; OFA: all *P*-values > 0.5).

Whole-brain Analyses

In order to identify voxels associated with our perceptual decision making task, we first conducted a search for all voxels which responded to presentation of the face/house images irrrespective of trial type. In addition to expected activations in left motor cortex (subjects all responded with their right hand), we observed significant clusters in a network of brain regions previously implicated in perceptual decision making: posterior parietal cortex bilaterally, medial prefrontal cortex, right dorsolateral prefrontal cortex and right anterior insula (not shown).

Using these clusters as a mask, we then conducted a whole-brain search for voxels that differed reliably across subjects as a function of trial type. All results reported here are corrected for FDR with an alpha of $P < 0.05$. At this threshold, the comparison between high confidence correct face and house trials (FFhc > HHhc, HHhc > FFhc) revealed no significant differences. Similarly, comparing low-confidence correct face and house trials (FFlc > HHlc, HHlc > FFlc) yielded no active voxels. Houses judged to be faces, however, yielded significantly greater activation than faces judged to be houses (HF > FH) at a number of cerebral loci. Statistically significant clusters were observed in the right superior parietal lobe, Brodmann's area 7 ($t = 9.6$, FDR $P < 0.01$) extending into the precuneus ($t = 9.1$, FDR $P < 0.01$) and also in the medial frontal gyrus (Brodmann's area 32; $t = 7.3$, FDR $P < 0.01$) extending into the supplementary motor area (Brodmann's area 6; $t = 7.0$, FDR $P < 0.01$). In Figure 6, a statistical map of significant activations is rendered onto the MNI template brain. The peak voxel for the comparison HF > FH is indicated by the blue crosshairs.

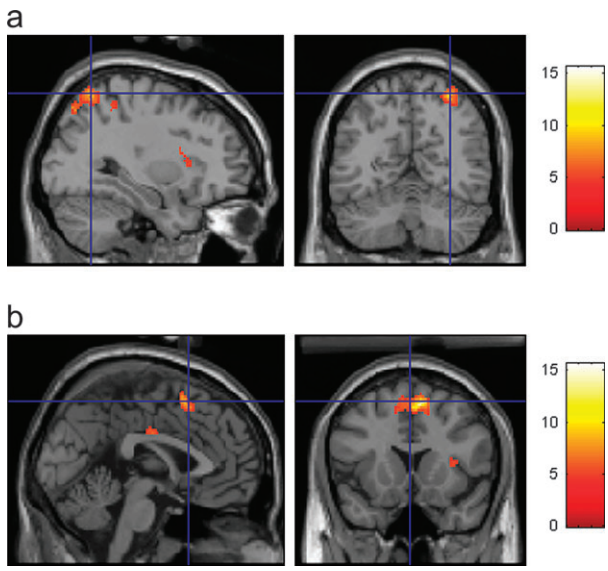

Figure 6. Imaging data: whole-brain analyses. Voxels across the brain responsive to the comparison HF > FH. The peak voxels from clusters in the right superior parietal lobe (*a*) and medial frontal cortex (*b*) achieved statistical significance at $P < 0.05$ (FDR correction for multiple comparisons). The red-yellow scale refers to the *t*-value.

Discussion

In order to simulate the experience of illusory perception in the laboratory, subjects were asked to discriminate images of houses and faces which were presented close to the threshold for perception. Image visibility was carefully controlled such that discrimination errors were made on 25% of trials. The object of the study was to determine whether selectivity of responses in face- and place-selective voxels in ventral visual cortex was preserved on these incorrect trials. The results here suggest that higher visual regions are not homogenous with respect to their responses during misperception of their preferred stimulus. Whereas the FFA responded reliably on both veridical perception and misperception trials, with robust, positive-going HRFs to both to faces judged to be faces (FF) and to houses judged to be faces (HF), other face-responsive voxels in the occipital cortex responded only during veridical face perception. A similar pattern, whereby BOLD responses indicated selectivity during veridical perception but not misperception, was observed in place-responsive voxels of parahippocampal gyrus (PPA). Thus, the FFA displayed responses to incorrect trials in line with hypotheses derived from predictive coding accounts of perception (Friston, 2003; Murray *et al.*, 2004) and top-down models of illusory perception (Collerton *et al.*, 2005; Grossberg, 2000), whereas the PPA responded in a fashion concordant with the assumption that incorrect trials simply reflected wrong guesses (Smith and Ratcliff, 2004).

Face-sensitive Regions

Our main finding may at first appear counterintuitive: when perception errs, the FFA can exhibit a stronger and statistically more robust response to images of houses than images of faces, despite the fact that, in our experiment, face-sensitive voxels were defined to be those which exhibited a greater response to faces than houses on a pre-experimental localizer task. On trials where a house is mistaken for a face (HF), the FFA is not receiving 'bottom-up' sensory input signalling the presence of a face stimulus (as no such stimulus is present). It is thus likely that the FFA is strongly modulated by a 'top-down' expectation that a face will be presented. In other words, in line with predictive coding accounts of perception, impoverished visual information is being 'explained' as corresponding to a face stimulus even where no face stimulus is present. That the hemodynamic response peaked later (~8 s post-stimulus) on these trials relative to correct trials may reflect the increased time required to resolve the ambiguous visual information into a (false) percept.

Previous studies have shown the FFA to be highly sensitive to top-down information. For example, a grey, oval-shaped stimulus was found to activate the FFA when contextual information suggested that it was a face (it was placed on top of a pair of shoulders). However, when the stimulus was viewed out of context, no face-related activity was elicited (Cox *et al.*, 2004). FFA activity has been found to track reported perception under conditions where retinal input remains constant, yet the percept varies, such as binocular rivalry (Tong *et al.*, 1998) or during the presentation of 'Mooney' faces (Andrews and Schluppeck, 2004). Anecdotal evidence suggests that humans have a strong predisposition to see faces where no face exists (in clouds, in the moon, or in landscapes) or to perceive a face from the barest of cues, perhaps resulting from the privileged place which faces are thought to hold in primate phylogeny and ontogeny (Yin, 1969; Farah *et al.*, 1998). Moreover, category-specific activation of FFA is observed when faces are imagined (Ishai *et al.*, 2000; O'Craven and Kanwisher, 2000). It is easy to see how generation of a mental image corresponding to the expected stimulus may accompany a predictive mechanism in perception. Taken together with these findings, the data reported here suggest that this propensity to perceive 'illusory' faces is likely to result from greater responsivity of FFA to top-down modulation.

Not all face-responsive voxels, however, showed preserved selectivity on incorrect trials. Voxels in inferior occipital regions (OFA) responded robustly and significantly to high-confidence correct face discriminations, but failed to dissociate HF from FH trials. This result was confirmed by a statistically significant region × condition interaction observed for these two areas. Models of face processing have proposed that faces are discriminated via a two-step mechanism, with early visual processing stages mediating 'structural encoding' of the physical properties of the face, and later stages responsible for configural processing underlying face identification (Bruce and Young, 1986). Recently, it was found that the OFA is sensitive to subtle differences in structural properties of a face, whereas the FFA tracks identity shifts across a categorical boundary (Rotshtein *et al.*, 2005). Our data provide further support for the view that there is a dissociation between these two face-processing regions, with more posterior sites sensitive to 'bottom-up' featural information, and later processing stages along the fusiform gyrus mediating configuration-based face judgements, presumably via top-down interactions with more anterior structures.

Place-sensitive Regions

By contrast, the PPA did not respond to faces 'misperceived' as houses. One possible interpretation of this finding is that mechanisms of perception in the PPA and FFA differ due to differences in the level of structural regularity of their preferred stimulus. The overall structure of a face is highly predictable (two eyes above a centrally positioned nose and mouth, etc), whereas the PPA is known to respond to a wide range of natural scenes, including both interior views, exterior scenes without obvious horizon (such as views of buildings) and views with horizon (such as views of mountains). It follows from this variability that a predictive code is less likely to be of use in the processing of natural scene stimuli, and PPA neurons may thus have to rely to a greater extent on bottom-up information. Indeed, there is evidence from previous work that PPA responses are not sensitive to stimulus familiarity or identity (Epstein *et al.*, 1999) whereas those in the FFA are (Dubois *et al.*, 1999; Rotshtein *et al.*, 2005). Unlike the FFA, the PPA does not respond in a viewpoint-invariant fashion (Epstein *et al.*, 2003), suggesting a stronger responsiveness to bottom-up input from primary visual regions. Perhaps most importantly, in patients who experience recurrent visual illusions and hallucinations, the illusory image typically occurs against a veridical background scene, and panoramic hallucinations of entire visual scenes are rare (ffytche and Howard, 1999).

However, it has also previously been reported that during rivalrous stimulation to each eye, PPA responses correlate with subjectively reported perception of buildings rather than bottom-up stimulation (Tong *et al.*, 1998). Additionally, the PPA is activated by mental imagery of places (Ishai *et al.*, 2000) and may mediate context effects in object perception (Bar and Aminoff, 2003). These results all run contrary to the intepretation that PPA responses uniquely track the 'bottom-up' veridical properties of the stimulus within minimal top-down input. Another interpretation of our data, thus, is that rather than reflecting fundamental differences in the responsivity of FFA and PPA, the failure to find parahippocampal place-selectivity on incorrect trials relates to a 'predictive' strategy employed by the subjects. Even though the face and house stimuli used in our study were well matched with regard to their level of structural regularity, in the real world houses tend to exhibit less regularity than faces, which may have prompted subjects to use a heuristic whereby they 'predicted' that the coming stimulus would be a face, using evidence against this prediction as evidence in support of the idea that the stimulus was in fact a house. Indeed, although we did not record this formally, in post-scan debriefing, subjects reported that they were more inclined to respond 'house' if they could not see the stimulus. In addition to conforming to predictive coding accounts of perception, the use of this strategic approach is described by 'random-walk' theories of decision making in two-choice decisions, which propose that subjects accumulate information along a single response dimension, such that information in favor of one response is information against the other (Link and Heath, 1975; Smith and Ratcliff, 2004). One prediction that can be made from this model is that if subjects are accumulating 'face' information during discrimination, the tendency to

respond 'face' will vary with the amount of information present in the stimulus, as at very low contrasts, most of the stimuli will be judged to be 'not-face' (i.e. house) stimuli. Our behavioral data, which revealed a linear trend for the bias to respond 'house' to increase as the stimuli were reduced in contrast (Fig. 2*b*), thus complement the fMRI data in supporting the idea that subjects were indeed using a 'face prediction' strategy.

PPA responses on all trials where the subject responded 'face' were negative-going, indicating that parameter estimates fell below the mean. This effect persisted even when contrast-modulated regressors were removed from the design matrix, indicating that it did not occur simply because contrast regressors captured much of the available variance. One conjecture is that during discrimination, use of a 'face-prediction' strategy led to active suppression of brain regions coding other non-predicted stimuli in the cognitive set. However, this interpretation must remain speculative until addressed with further research.

In an earlier report, visual responses in striate (V1) and extrastriate (V2, V3) cortex were observed to track illusory perception when subjects judged whether a simple visual stimulus was present or absent (Ress and Heeger, 2003). Here, using a forced-choice discrimination paradigm, we show a similar phenenomenon for more complex stimuli in higher visual regions. However, one of the limitations of using a forced-choice discrimination (rather than detection) is that it is difficult to draw conclusions about the subjective experiences of the subject, as below-threshold responses can drive behavior (Marcel, 1983). However, one possibility is that the failure to find PPA responses during incorrect 'house' responses reflects a failure for a conscious 'mispercept' to be formed of these stimuli, perhaps because in our study, FFA received greater top-down input from anterior control structures. This intepretation is in line with the view that neural activity in the PPA does contribute to conscious visual perception (Tong *et al.*, 1998) and, more generally, with the observation that while human observers may frequently 'misperceive' objects as faces (pareidolia), such errors are less common for natural scene stimuli.

Whole-brain Analyses

When the two classes of incorrect trial were compared with voxelwise comparisons across the brain (HF > FH), areas previously implicated in perceptual decision making were strongly activated, including the posterior parietal cortex, right dorsolateral prefrontal cortex, right anterior insula and dorsomedial prefrontal cortex. Neuroimaging (Huettel *et al.*, 2005; Pessoa and Padmala, 2005) and single-cell electrophysiology (Shadlen and Newsome, 2001) research has suggested that all of these regions play an important role in decision making under uncertainty. Considerable controversy surrounds the precise function of each part of this network, particularly with respect to whether they subserve categorical selection among alternatives or ancillary processes required to make difficult decisions (such as working memory and attention). Whilst our study was not designed to address the function of these regions in decision making, it is interesting to note that comparing the two types of incorrect trials here (HF > FH) isolated voxels in a subset of these regions: the right posterior parietal cortex and precuneus, and in mediodorsal prefrontal cortex. Our data offer a tentative new perspective on the function of these regions in decision making, by suggesting that they may be involved in testing predictions during uncertain decisions. Medial prefrontal and posterior parietal cortical regions are densely interconnected (Wise *et al.*, 1997) and extrastriate regions receive re-entrant connections particualrly from the superior parietal lobe (Van Essen *et al.*, 1992). It is plausible that these frontal and parietal sites are the 'source' of the top-down prediction about the forthcoming stimulus, and that face-selective FFA responses on incorrect trials are driven by input from these regions.

Other Considerations

It could perhaps be argued that the differences between trial types observed here simply reflect confounding differences in the basic visual properties of the images used. Indeed, one particular concern is that due to the bias exhibited by subjects to respond 'house' at low contrasts, HH and FH regressors are contaminated with larger numbers of trials where contrast levels were low. We think that it is unlikely that our results simply reflect low-contrast stimulation on FH trials for a number of reasons. Firstly, contrast regressors did not seem to capture much of the variance, presumably because the overall variation in contrast was slight, as all stimuli were presented in a narrow range around a individual determed thresholds for perception. This suggests that image contrast was not a major determinant of the response in these extrastriate regions. Most importantly, however, fMRI responses in posterior regions of the occipital cortex tend to be more sensitive to the overall contrast of the stimulus (Boynton *et al.*, 1999), and yet these regions did not show the effect for the HF > FH comparison. An a-priori defined control region which fell in or close to V1 exhibited no reliable differences between HF and FH trials, as would be predicted if this result depended on differences in contrast between trial types. Even in the extra-striate cortex the result was not ubiquitous: although HF trials exhibit stronger responses than FH trials in the FFA, in the OFA, this was not the case. We thus think it is unlikely that our results can be accounted for simply by differences in stimulus contrast.

Conclusions

In this report we describe how ventral visual regions respond during misperception of one object as another. Robust BOLD responses were observed in face-responsive regions of the fusiform gyrus (but not inferior occipital gyrus) when a house is perceived as a face, and it is argued that this activity may underlie 'illusory' face perception. Furthermore, medial frontal and parietal regions previously implicated in perceptual decision making also become active during misperception of faces. These regions may be candidates for the source of a top-down prediction about what the forthcoming stimulus is to be. These data provide support for the notion that the perceptual system makes use of a predictive code in deciding what it is that we are seeing.

Notes

We thank Suzanne Palmer for help with data collection and analysis. This work was carried out with support from the William J. Keck Foundation, the Columbia University Provosts Academic Quality Fund, and a grant from the National Insititute of Health (#R21066129) to J.A.M.

Address correspondence to C. Summerfield, Department of Psychology, Columbia University, 406 Schermerhorn Hall, 1190 Amsterdam Ave, New York, NY 10027, USA. Email: summerfd@paradox.columbia.edu.

References

Aguirre GK, Zarahn E, D'Esposito M (1998) An area within human ventral cortex sensitive to 'building' stimuli: evidence and implications. Neuron 21:373–383.

Ahissar M, Hochstein S (2004) The reverse hierarchy theory of visual perceptual learning. Trends Cogn Sci 8:457–464.

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

Bar M (2003) A cortical mechanism for triggering top-down facilitation in visual object recognition. J Cogn Neurosci 15:600–609.

Bar M (2004) Visual objects in context. Nat Rev Neurosci 5:617–629.

Bar M, Aminoff E (2003) Cortical analysis of visual context. Neuron 38:347–358.

Biederman I, Mezzanotte RJ, Rabinowitz JC (1982) Scene perception: detecting and judging objects undergoing relational violations. Cognit Psychol 14:143–177.

Boynton GM, Demb JB, Glover GH, Heeger DJ (1999) Neuronal basis of contrast discrimination. Vision Res 39:257–269.

Bruce V, Young A (1986) Understanding face recognition. Br J Psychol 77 (Pt 3):305–327.

Brunia CH, Damen EJ (1988) Distribution of slow brain potentials related to motor preparation and stimulus anticipation in a time estimation task. Electroencephalogr Clin Neurophysiol 69:234–243.

Collerton D, Perry E, McKeith I (2005) Why people see things that are not there: a novel perception and attention deficit model for recurrent complex visual hallucinations. Behav Brain Sci (in press).

Cox D, Meyers E, Sinha P (2004) Contextually evoked object-specific responses in human visual cortex. Science 304:115–117.

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

Dubois S, Rossion B, Schiltz C, Bodart JM, Michel C, Bruyer R, Crommelinck M (1999) Effect of familiarity on the processing of human faces. Neuroimage 9:278–289.

Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top-down processing. Nat Rev Neurosci 2:704–716.

Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. Nature 392:598–601.

Epstein R, Harris A, Stanley D, Kanwisher N (1999) The parahippocampal place area: recognition, navigation, or encoding? Neuron 23:115–125.

Epstein R, Graham KS, Downing PE (2003) Viewpoint-specific scene representations in human parahippocampal cortex. Neuron 37:865–876.

Farah MJ, Wilson KD, Drain M, Tanaka JN (1998) What is 'special' about face perception? Psychol Rev 105:482–498.

ffytche DH, Howard RJ (1999) The perceptual consequences of visual loss:'positive' pathologies of vision. Brain 122:1247–1260.

Friston K (2003) Learning and inference in the brain. Neural Netw 16:1325–1352.

Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15:870–878.

Grossberg S (2000) How hallucinations may arise from brain mechanisms of learning, attention, and volition. J Int Neuropsychol Soc 6:583–592.

Henderson JM, Hollingworth A (1999) High-level scene perception. Annu Rev Psychol 50:243–271.

Huettel SA, Song AW, McCarthy G (2005) Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. J Neurosci 25:3304–3311.

Ishai A, Ungerleider LG, Haxby JV (2000) Distributed neural systems for the generation of visual images. Neuron 28:979–990.

Ishai A, Pessoa L, Bikle PC, Ungerleider LG (2004) Repetition suppression of faces is modulated by emotion. Proc Natl Acad Sci USA 101:9827–9832.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311.

Kleinschmidt A, Buchel C, Hutton C, Friston KJ, Frackowiak RS (2002) The neural structures expressing perceptual hysteresis in visual letter recognition. Neuron 34:659–666.

Link S, Heath R (1975) A sequential theory of psychological discrimination. Psychometrika 40:77–105.

Marcel AJ (1983) Conscious and unconscious perception: experiments on visual masking and word recognition. Cognit Psychol 15:197–237.

Martinez M, Benavente R (1998) The AR face database. CVC Technical Report #24.

McKellar P (1957) Imagination and thinking. London: Cohen & West.

Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. Biol Cybern 66:241–251.

Murgatroyd C, Prettyman R (2001) An investigation of visual hallucinosis and visual sensory status in dementia. Int J Geriatr Psychiatry 16:709–713.

Murray SO, Schrater P, Kersten D (2004) Perceptual grouping and the interactions between visual cortical areas. Neural Netw 17:695–705.

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

Palmer S (1975) The effects of contextual scenes on the identification of objects. Mem Cognit 3:519–526.

Pascual-Leone A, Walsh V (2001) Fast backprojections from the motion to the primary visual area necessary for visual awareness. Science 292:510–512.

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

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

Rensink RA (2000) Seeing, sensing, and scrutinizing. Vision Res 40:1469–1487.

Ress D, Heeger DJ (2003) Neuronal correlates of perception in early visual cortex. Nat Neurosci 6:414–420.

Rotshtein P, Henson RN, Treves A, Driver J, Dolan RJ (2005) Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. Nat Neurosci 8:107–113.

Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol 86:1916–1936.

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

Tallon-Baudry C, Bertrand O, Henaff MA, Isnard J, Fischer C (2005) Attention modulates gamma-band oscillations differently in the human lateral occipital cortex and fusiform gyrus. Cereb Cortex 15:654–662.

Tobler PN, Fiorillo CD, Schultz W (2005) Adaptive coding of reward value by dopamine neurons. Science 307:1642–1645.

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

Van Essen DC, Anderson CH, Felleman DJ (1992) Information processing in the primate visual system:an integrated systems perspective. Science 255:419–423.

Wager TD, Nichols TE (2003) Optimization of experimental design in fMRI: a general framework using a genetic algorithm. Neuroimage 18:293–309.

Warrington EK, Shallice T (1984) Category specific semantic impairments. Brain 107:829–854.

Wise SP, Boussaoud D, Johnson PB, Caminiti R (1997) Premotor and parietal cortex: corticocortical connectivity and combinatorial computations. Annu Rev Neurosci 20:25–42.

Yin R (1969) Looking at upside down faces. J Exp Psychol 81:141–145.