

Dynamics of visual recognition revealed by fMRI

Thomas Carlson,^{a,b,*} Meike J. Grol,^{b,c} and Frans A.J. Verstraten^{b,c}

^aDepartment of Psychology, Vision Sciences Laboratory, Harvard University, Cambridge, MA 02118, USA

^bHelmholtz Institute, Psychonomics Division, Universiteit Utrecht, The Netherlands

^cF.C. Donders Centre for Cognitive Neuroimaging, Radboud Universiteit Nijmegen, The Netherlands

Received 2 November 2005; revised 14 March 2006; accepted 23 March 2006

Available online 16 June 2006

In our daily lives, recognizing a familiar object is an effortless and seemingly instantaneous process. Our knowledge of how the brain accomplished this formidable task, however, is quite limited. The present study takes a holistic approach to examining the neural processes that underlie recognition memory. A unique paradigm, in which visual information about the identity of a person or word is slowly titrated to human observers during a functional imaging session, is employed to uncover the dynamics of the visual recognition in the brain. The results of study reveal multiple unique stages in visual recognition that can be dissociated from one another based on temporal asynchronies and hemodynamic response characteristics.

© 2006 Elsevier Inc. All rights reserved.

Keywords: fMRI; Face recognition; Word recognition; Recognition memory

Introduction

Humans can effortlessly recognize an object in just a fraction of a second. During this very brief period of time, physical energy entering the eye is processed by a hierarchical network of brain regions in which representations of visual information are continually altered before it finally reaches the stage of conscious recognition. The present study employs a unique paradigm in which visual information is gradually revealed to an observer over an extended period of time. This type of paradigm has previously been employed in a variety of neuroscientific investigations of visual recognition to reveal interesting aspects of visual priming (James et al., 2000) and perceptual hysteresis (Kleinschmidt et al., 2002) in functional neuroimaging and perceptual completion in EEG (Doniger et al., 2000; 2001). In the present study, we couple

this unique paradigm with a novel analysis to examine the temporal dynamics of recognition in different brain regions.

The representation of visual information is continually evolving in the brain. Visual information that initially arrives to the visual system in form of light is first transduced to neural signals by millions of photoreceptors. This initial representation is then subsequently transformed to a representation accessible to the observer for the task at hand. While there remains considerable discussion about the exact nature of these representations and how these transformations are implemented, three representational stages are ubiquitous in modern theories of the object recognition: an image-based representation that codes low-level features like edges and color; an object-based representation which maintains a structural description of objects, and a categorical or semantic representation that allows the observer access to stored information about the objects (e.g., a cup is for drinking). The transformation of visual information to this final semantic representation constitutes recognition. Notably, despite the complexities of the aforementioned transformations, the human brain can perform these operations in just a fraction of a second.

The brain's ability to rapidly process information coupled with the slow hemodynamic response makes the study of the temporal dynamics of visual recognition challenging for functional neuroimaging studies. Significant research has been dedicated to this problem, and present day methods allow for resolution of distinct cognitive processes separated by a few hundred milliseconds (for review, see Formisano and Goebel, 2003). In some cases, there is sufficient time between cognitive processes to make this temporal distinction (e.g., Formisano et al., 2002). ERPs studies have estimated that 150 ms is sufficient time for visual information to reach frontal areas (Thorpe et al., 1996; VanRullen and Thorpe, 2001). This effectively means we can obtain only a static view of the recognition process using traditional functional neuroimaging paradigms. Still, attempts have been made to uncover the temporal dynamics of recognition by applying these methods (Pernet et al., 2004). This approach, however, rests heavily on the assumption that the hemodynamic response of different cortical areas is nearly identical. An assumption that is almost certainly invalid given the complexities of the cerebral vasculature. Given the temporal limitations and necessary assumptions, the study of the temporal

* Corresponding author. Department of Psychology, Vision Sciences Laboratory, 33 Kirkland St, Harvard University, Boston, MA 02118, USA. Fax: +1 617 495 3764.

E-mail address: tom@wjh.harvard.edu (T. Carlson).

Available online on ScienceDirect (www.sciencedirect.com).

dynamics of complex processes has in large part been forgone in functional neuroimaging studies of visual recognition.

In the present study, we employ a unique paradigm inspired by the information processing approach to the study of the mind to capture both evolving representation of visual information and the time course of recognition in the brain. Succinctly stated, the information processing approach is to study the brain in terms of different subsystems that maintain and transmit different representations of the world. Based on this principal, we employed an experimental paradigm to parse cortical areas that participate in distinct stages of recognition according to their differential responses to a steady influx of new visual information into the system (i.e., the brain). The study has two main experiments investigating the recognition of two well-studied categories of stimuli, faces and words. Stimuli in the face experiment were images of famous individuals (e.g., Jack Nicholson); the word stimuli were common nouns (e.g., ELEPHANT). Notably, exemplars from each of the categories have unique semantic interpretations; such that a match between the visual input and one's internal representation constitutes recognition.

Figs. 1A and B show graphical depictions of the face and word stimulus respectively. At the beginning of each trial, subjects are provided with no information about the identity of the stimulus, except the prior knowledge of the category (i.e., face trials and word trials were done in separate experimental runs). Initially, recognition is impossible. Visual information is then slowly titrated to the subject at a fixed rate. In the face experiment, visual information takes the form of small patches of the image. In the word experiment, the units of information come in the form of letters. As time progresses, additional information is revealed to the subject, and consequently, identification of the stimulus becomes probabilistically more likely. As the sequence progresses, a discrete point in time occurs when the subject recognizes the person or word, assuming the subject knows the identity of the stimulus.

This paradigm was used to parse cortical areas contributing to different stages of recognition based both on the hemodynamic response properties of cortical areas and the temporal dynamics of the response. As such, it represents a departure from typical functional neuroimaging paradigms. Presently, the vast majority of neuroimaging experiments employ either block and event related designs. In these paradigms, the activity evoked by a test stimulus is compared relative to some other condition. For example, greater activity while viewing faces relative to other object categories is interpreted as a face selective response (Kanwisher et al., 1997). Here, perceptual and cognitive processes contributing to unique stages of the recognition process are distinguished based on their response characteristics and temporal onset. For example, as tiles are removed in the face stimulus, there is an increase in global contrast of the image. Thus, we expect activity in areas sensitive to image contrast (e.g., early visual areas) to be correlated with the amount of visual information present at any given moment in time. Fig. 1C shows a plot of the accumulation of visual information as a function of time in the experiment. From this, we hypothesized three unique response categories (Fig. 1C). The first we refer to as an *information dependent response*. In the information-dependent response, activity correlates with the amount of visual information presented to the observer as in the example above. The second, we refer to as a *sustained response*. Here, activity is task-driven and invariant to the amount of visual information present. Together, the first two responses presumably reflect encoding processes, as activity is present prior to recognition. The final response category

is termed *recognition dependent*. Recognition-dependent processes are time locked to a cognitive event, specifically the moment that the subject achieves recognition. Notably, this event is delayed until there is sufficient visual information present for the subject to match the stimulus to memory; and the difference in the temporal onset of encoding processes and the recognition-dependent processes will be on the order of seconds (~5–30 s), which is more than sufficient time to compensate for the poor temporal resolution of fMRI. The paradigm thus allows for the uncoupling of two unique types of encoding processes, and recognition dependent processes. Furthermore, it allows us to establish a temporal sequence of events in the brain, which normally occur on a time scale of less than a few hundred milliseconds.

Materials and methods

Participants

Eight healthy English speaking volunteers (5 female, 3 male) with normal or corrected to normal vision with an average age of 26 years participated in the experiments. Informed written consent was obtained from all volunteers prior to participation in the experiment. This study was approved by the local ethics committee (Commissie Mensgebonden Onderzoek Region Arnhem-Nijmegen, The Netherlands).

Visual stimulation

Stimuli were generated in MATLAB (Natick, MA) on a Macintosh iBook computer (500Mhz) using functions provided by the Psychtoolbox (Brainard, 1997; Pelli, 1997) and projected onto a translucent screen. Subjects viewed the stimulus through a mirror mounted on the headcoil. Reaction time data were collected on separate PC computer synced to the stimulus presentation on the Macintosh iBook through the serial port.

fMRI data acquisition

Subjects were scanned on a Siemens 3 T Trio scanner at the FC Donders Centre for Cognitive Neuroimaging (Nijmegen, The Netherlands). Functional imaging was conducted using an echo-planar imaging (EPI) sequence (TR = 2000, TE = 30 ms, flip angle 90°, FOV 192 mm, matrix size 64 × 64). Scanned volumes were 30 axial slices (3 × 3 mm in plane resolution, 3.5 mm thick; interslice gap 0.35 mm) covering the whole brain. Functional runs in the main experiment consisted of 396 functional acquisitions (792 s). Localizer scans for FFA and LO used identical imaging parameters with a shorter scan duration (200 acquisitions, 400 s). The first 4 images of each functional run were discarded to allow for T1 equalization. High-resolution T1-weighted structural scans (1 × 1 × 1 mm) were acquired for each subject for cortical reconstruction and volume-based statistical analysis.

Behavioral task

Each subject participated in two runs of 24 trials for a total of 48 trials in both the face and word experiments. Individual trials were 32 s in duration. A single run was 396 functional acquisitions or 792 s (TR = 2 × 396 acquisitions) including 8 s to allow for T1 equalization, 768 s of the experimental task (24 trials × 32 s), and

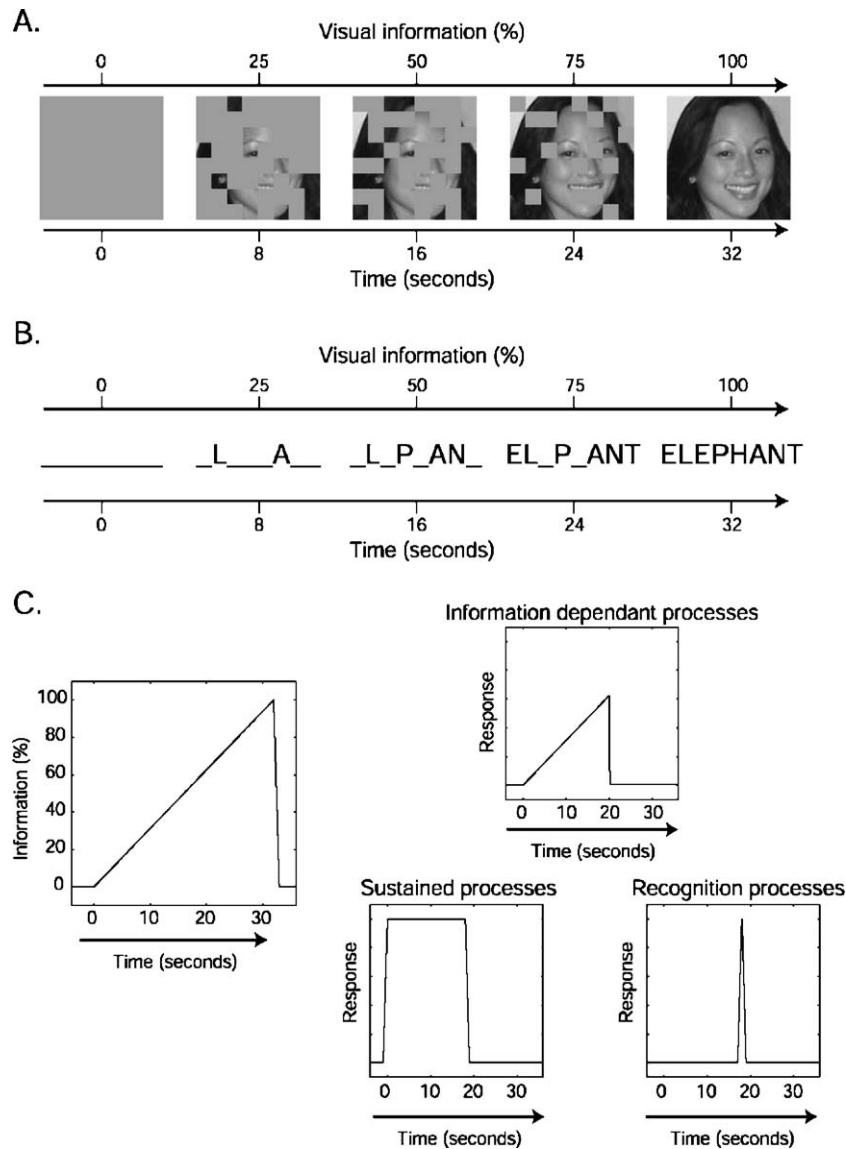


Fig. 1. Experimental paradigm. (A) Schematic illustration of the face recognition paradigm. Visual information in the form of small patches of the image was incrementally revealed to the subject over a period of 32 s. (B) Schematic illustration of the word recognition paradigm. Visual information in the form of letters was incrementally revealed over a period of 32 s. (C) Visual information plotted as function of time shown with the three hypothetical response models. Information-dependent response was modeled as linear increase in activity in response to increasing visual information. Sustained response model as box car function corresponding to the time subjects are performing the recognition task. Recognition response was modeled as a discrete event occurring at the moment of recognition.

16 s at the end of the run to allow the signal to return to baseline. Subjects were instructed to make a button press response the moment they recognized the face or word. Words and faces were randomly selected without replacement from a fixed pool of stimuli.

Face recognition experiment

Stimuli were 256×256 pixel gray scale images of 48 famous politicians, entertainers and athletes from the recent past taken from the web. Appendix A lists the names of the individuals shown in the pictures. Images subtended approximately $10 \times 10^\circ$ of visual angle. Images were hidden behind by an 8×8 matrix of uniform gray tiles (32×32 pixels) at the beginning of each trial (see Fig. 1B). At the onset of the trial, randomly selected tiles were removed uncovering the image at a rate of 2 tiles per second. The duration of

each trial was 32 s, the time required to uncover the entire image. After subjects indicated that they had recognized the stimulus, the remaining tiles changed to random noise patterns generated from a distribution of pixel intensities from the original image. This procedure was adopted to prevent a secondary recognition moment as a result of subjects changing their mind.

Word recognition experiment

Stimuli were common eight letter words subtending approximately 1 by 8° of visual angle on the screen. The words selected for use in the study were drawn from a variety of categories (e.g., object, actions, emotion etc.) of common eight-letter words Appendix B lists all of the words used in the study. The position of the 8 letters was initially marked by an array of underscores (_) that served as placeholders (see Fig. 1). Randomly selected letters

were then placed into the array at a rate of 1 letter every 4 s. The total duration of each trial was 32 s, the time required to reveal the entire word. After subjects indicated recognizing the word, the remaining letters were filled with asterisks (*) at the same rate until the termination of the trial. Again, this procedure was adopted to prevent a secondary recognition moment.

Data analysis

Acquired data were analyzed using the Brainvoyager 2000 version 4.9 (Brain Innovation, Maastricht, The Netherlands). Preprocessing of functional imaging runs consisted of motion correction, slice scan time correction, linear trend removal and spatial smoothing (4-mm FWHM). Functional data were registered to the high-resolution anatomy, corrected for intrasession head movements and transformed into Talairach space (Talairach and Tournoux, 1988) for group volume-based statistical analyses.

Three response types were modeled: information-driven processes, sustained processes, and recognition events. Information-driven processes were modeled as a triangle wave beginning at zero at stimulus onset and ending with a value of one at the end of the trial. Sustained processes were modeled as a box car function rising at stimulus onset and returning to baseline at recognition. Recognition events were modeled as a single event (delta function) occurring when subjects indicated recognizing the face or word. Modeled responses were convolved with a synthetic hemodynamic response function for statistical tests using the general linear model. Statistical maps were generated for each of the three response models for the word and face conditions. A cognitive conjunction analysis was also performed for the word and face experiments for each of the three response models (Price and Friston, 1997). An additional conjunction analysis was also performed combining the face sustained processes condition and the word information-driven processes condition (see Discussion). All statistical maps were thresholded at $P < 0.05$ for a group random effects analysis.

Regions of interests (ROIs): lateral occipital cortex (LO) and the fusiform face area (FFA)

In separate functional imaging runs, LO and FFA were localized in a block experiment. Subjects passively viewed images of faces, natural scenes, common objects, and scrambled images. Images were presented at a rate of 2 images per second in 24-s blocks. Face, natural scene, object, and scrambled image blocks were shown consecutively and repeated four times in a single functional imaging run. A model of the expected response was generated using previously reported response properties of LO and FFA to the four classes of stimuli (Grill-Spector, 2003). LO was defined as a statistically significant ($P < 0.0001$, uncorrected) cluster of voxels in lateral occipital cortex (mean volume 2952 mm³ (RH), 2494 mm³ (LH)). FFA was defined using the same criteria in the fusiform (mean volume 1433 mm³ (RH), 642 mm³ (LH)). LO was identified in both hemispheres in all 8 subjects. Bilateral FFA was identified in 7 of 8 subjects. The remaining subject had significant activation only in the right hemisphere.

In the analysis of the behavioral data with the fMRI data, reaction times were binned into intervals equal to the TR of the functional imaging acquisition (2 s). Data from the first two bins (0–2 s and 2–4 s) were excluded from the analysis as they failed to reach an arbitrary minimum criterion threshold of 5 responses across all subjects. Time series data were constructed by averaging

the data from all of the subjects for the four ROIs. The time series data were then deconvolved using a filter given by a gamma function:

$$h(t) = \frac{(t/\tau)^{n-1}}{\tau\Gamma(n-1)} \exp(-t/\tau)$$

with parameters ($n = 2$, $\tau = 1.25$) in agreement with empirical measures (Boynton et al., 1996). The purpose of the deconvolution operation was to remove the lag and blurring associated with hemodynamic response to obtain a more accurate representation of neural activity. This method has effectively been used in previous studies to make single time point predictions in fMRI time series data indicating the method provides a robust estimate of the underlying response model (Carlson et al., 2003).

Results

Behavioral data

Subjects indicated recognizing 85.2% of the faces and 99.2% of the words. A histogram of the reaction times of trials in which subjects indicated recognizing the face is shown in Fig. 2A. The mean reaction time for the face experiment was 12.67 s (on average approximately 40% of the image was opened at the time of recognition where the information density based on the window opening sequence was random) with standard deviation of 6.28 s. In the word experiment, the mean reaction time was 17.68 s (approximately 4.5 letters) with a standard deviation of 4.46 s. In a comparison of face and word trials reaction times, face were recognized significantly faster than word ($P < 0.001$; Kolmogorov–Smirnov test), but with a lower frequency ($P < 0.001$; t test).

Imaging data

The results of the experiments are divided into two sections. In the first section, a detailed analysis is performed on regions of interest (ROIs) previously identified as critical stages in perceptual processing of objects and faces. Here, we examine how visual information in face recognition is represented in the fusiform face area (FFA), and lateral occipital areas (LO) by attempting to determine which of the three underlying response models best describes the activity in these regions. In the second section, a more global approach is taken, and cortical areas are parsed based on their response characteristics (information, sustained, recognition) for both the face and word stimuli.

ROI analysis: the representation of visual information for faces in lateral occipital areas and the fusiform

We first examined activity in lateral occipital cortex (LO) and in the fusiform face area (FFA), two ROIs believed to be critical waypoints in the perceptual processing of objects and faces (Grill-Spector et al., 2001; Kanwisher et al., 1997). In separate functional runs, LO and FFA in the left (LH) and right hemisphere (RH) were independently identified to examine the response of the ROIs.

In the introduction, three hypothetical models were presented. As previously stated, the information model predicts an increasing response to increasing visual information. The sustained model predicts a sustained response while subjects are performing the task. And the recognition model predicts an event-related response

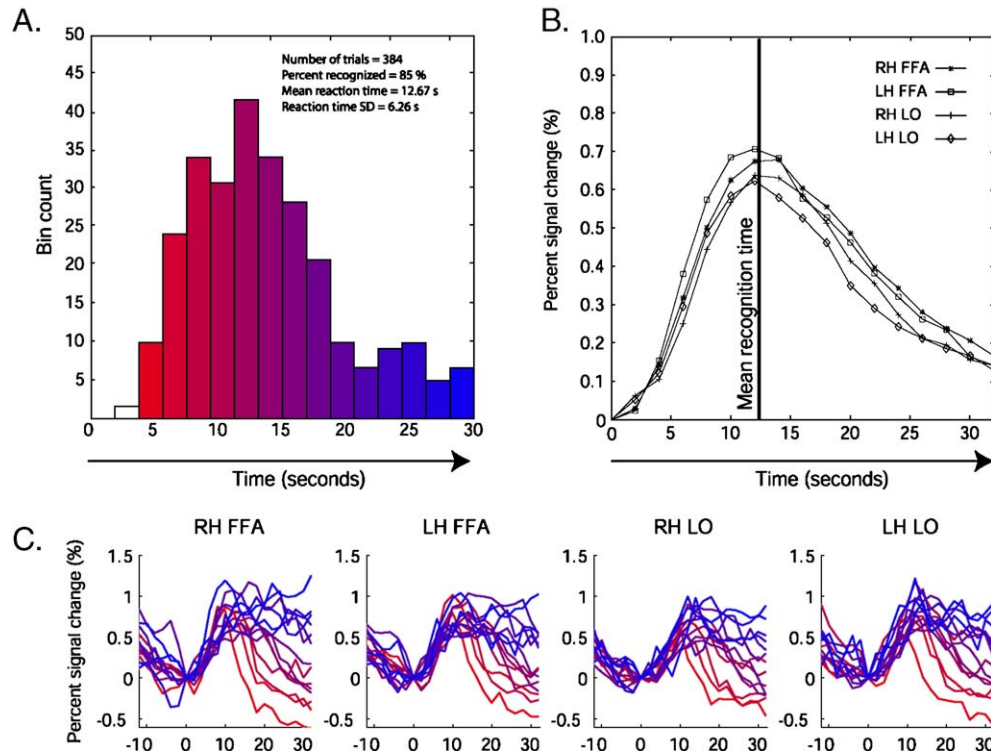


Fig. 2. ROI analysis of activity in face experiment. (A) Color-coded histogram of recognition times for face identification experiment. (B) Grand mean event-related average shown from stimulus onset constructed by averaging time courses from all the subjects across recognition times for each of the ROIs. The bisecting line marks the mean recognition time for all subjects. (C) ROI event-related averages for different recognition times. Event-related averages were constructed by binning the data in 2 s bins ($TR = 2$) averaged across subjects. Colors in the plots correspond recognition times shown in the histogram in panel A.

at the moment of recognition. The first set of analyses we performed aimed to determine which, if any, of these models most accurately describes the activity in the LO and the FFA.

Previously, James et al. (2000) using a similar paradigm with objects also examined the response of cortical areas associated with object-related activity as a function of increasing information. In their study, they examined two regions labeled peristriate cortex and the fusiform, among others areas, which can be viewed as analogs to LO and FFA for face recognition. Given the similarities between their localizer stimuli and that of the present study, peristriate cortex in their study is likely to be equivalent to the region commonly referred to as LO. FFA is argued to be a more specialized region of the fusiform for face recognition (Kanwisher et al., 1997). James et al. reported that activity in peristriate cortex and the fusiform was task-driven, increased with increasing visual information, and peaked at the moment of recognition. These observations support a slightly modified version of the information-dependent response model. The necessary modification is in regards to what happens at the moment of recognition. A strict interpretation of the information model predicts that activity would continue to rise after recognition. The authors found that activity was task-driven—that is, after recognition activity dropped. Thus, the model that best describes James et al.'s interpretation of the data would be a multiplicative hybrid of the information and sustained response models. We term this the *hybrid information model*. Their observations, however, were based on qualitative observations of normalized time courses, as the authors were primarily concerned with investigating the effects of visual priming. One of the primary motivations of the present study

was the question of how visual information is represented in different cortical areas, and as such, we performed a more detailed analysis.

Fig. 2 shows a color-coded histogram of recognition times (Fig. 2A), a grand mean time course for LO and FFA in the left and right hemispheres from the stimulus onset (Fig. 2B) with the mean recognition time indicated by a line bisecting the time course, and the time course for individual recognition times constructed by binning the data in 2-s intervals (Fig. 2C). Given the data, the recognition response model can immediately be dismissed. Activity in LO and FFA clearly rises prior to recognition. The two alternatives we considered next were the sustained model and the aforementioned hybrid information model. Based on these alternatives, we formulated two tests that make differential predictions for the two models. Individually, the two tests provide support for either of the two models; together, they are more convincing. The first test examines the relationship between visual information and the level of activity in these regions (i.e., does activity rise with increasing visual information?). The second examines the relationship between peak response and the moment of recognition (i.e., does activity peak at the moment of recognition?). Each of the tests was performed on time courses extracted from the four independently identified ROIs (RH FFA, LH FFA, RH LO, and LH FFA). Prior to the analysis, a deconvolution operation (see Materials and methods) was applied to the time courses to obtain a more accurate representation of neural activity. Effectively, this operation removes the effects of the hemodynamics from the signal (i.e., blurring and lag).

Does activity rise with increasing visual information? To answer this question, we examined the relationship between the area beneath the response curve and recognition time. As previously noted, activity in these regions was reported to rise when subjects engage in the task, and falls off when subjects recognize the stimulus (James et al., 2000). This observation coupled with the expected response of the two models makes different predictions for a regression of the area beneath the response curve as a function of recognition time. The sustained response model predicts that area will grow linearly. If activity immediately rose to an arbitrary value of 4, for example, and then maintained a sustained level of activity up to the point of recognition, then recognition times of 1, 2, 3, . . . n would result in areas of 4, 8, 12, . . . $\sum_{n=1}^i 4$. The hybrid information model makes a different prediction. Here, activity rises as a function of increasing visual information. As such, the area beneath the response curve will grow nonlinearly. If the relationship between visual information and activity were linear with a slope of 2, for example, recognition times of 1, 2, 3, . . . n would correspond to areas of 2, 6, 12, . . . $\sum_{n=1}^i 2i$. Importantly, the nonlinearity is driven by any direct relationship between activity and visual information. Only a zero correlation between the activity and visual information (i.e., the sustained response model) would result in a linear mapping of area to recognition time. The two models were tested using a polynomial regression of area as function of recognition time. In addition to a simple linear model, several nonlinear models were considered including a quadratic model, a cubic model, and a full cubic model that included a quadratic term. The nonlinear models were compared to the fit of the linear model, which served as the null hypothesis. Any evidence of nonlinear relationship between the two variables will be considered as support for the hybrid information model. Scatterplots of the data with the fits for four models are shown in Fig. 3. Visually, the data appear to be well described by the linear model in all four ROIs. This observation is confirmed by the statistical tests of linear model, which are highly significant. The table also shows the significance tests for the higher order models. In all cases, the high order models (quadratic, cubic, and full cubic) did not

significantly improve the fit of the linear model with P values ranging from 0.45 to 1. The data thus support the sustained model in favor of the hybrid information model.

Does activity peak at the moment of recognition? The analysis of area as a function of recognition time supports the sustained response model, which would indicate there is no peak at recognition—that is, the response is flat. James et al. (2000), however, found a significant correlation between mean recognition time and the normalized peak response in the fusiform. This correlation also appears to be present in our data based on visual inspection of the grand mean time courses (Fig. 2B). These observations support the hybrid information model since recognition time can be linked to the peak response. When the data are binned according to recognition time, as in Fig. 2C, the hybrid information model can make a much stronger prediction. If the hybrid information model is correct, then the peak response and recognition time will have a specific one-to-one relationship (i.e., a slope of 1). That is, if recognition occurred at time i , then the peak response should also occur at time i . This is, assuming that one has properly accounted for the lag in the hemodynamic response, which would simply shift the data and not alter the slope. Fig. 4A shows scatterplots of the peak response as a function of recognition time along with fits from a simple linear model. Indeed, there is a visible trend linking recognition time with the peak response, but the slope is shallower (~ 0.3 – 0.4) than the prediction made from the hybrid information model (slope = 1). Also evident is nonconstant variance as indicated by the cone shape distribution of the data. Fig. 4B elaborates on this by showing a plot of the residuals for all four ROIs as function of recognition time. The plot shows that the errors in the fit become larger as function of recognition time. The nonconstant variance can be explained by the sustained response model in conjunction with the observation that activity in these regions is task driven. The task-driven nature of the response means that irrespective of the underlying response model stimuli that are recognized quickly will have early peaks, since the response falls off after recognition. This property

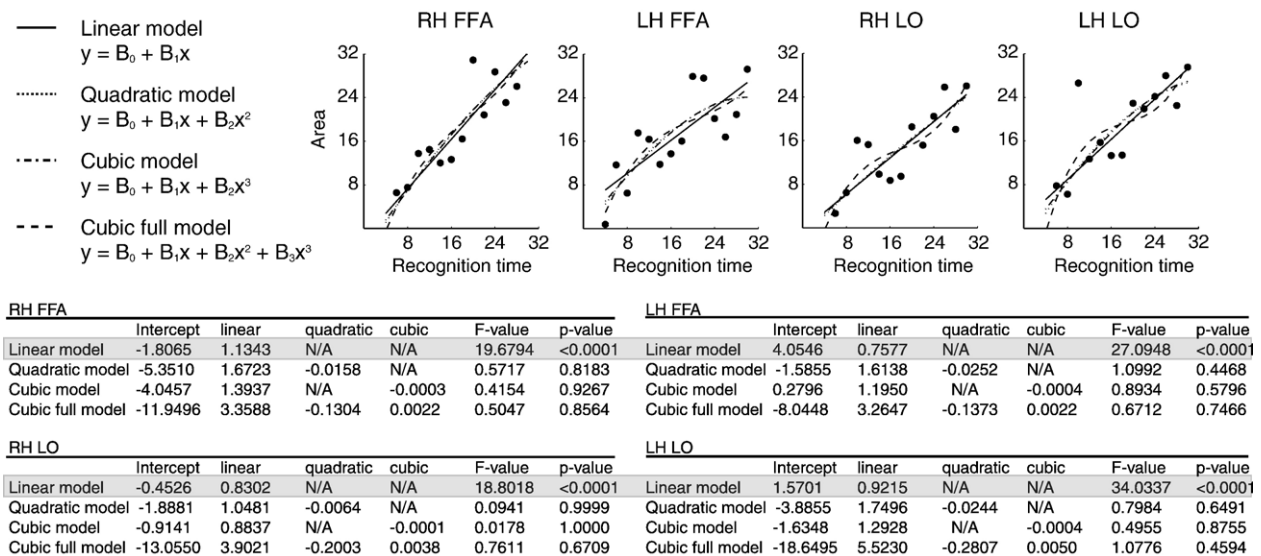


Fig. 3. Sustained vs. hybrid information model: Area beneath response curve. The sustained response model predicts a linear relationship between the area beneath the response curve and recognition time; the hybrid information predicts a nonlinear relationship (see text). Shown are scatter plots of the data of area as a function of recognition time. Fits for four models (the linear model, and three nonlinear models) are shown on each of the plots. The table below shows the fitted values from the polynomial regression. The fit from the linear model is highly significant ($P < 0.001$) in all four ROIs. All of the higher order models (Quadratic, Cubic, and Cubic full model) failed to significantly improve the fit.

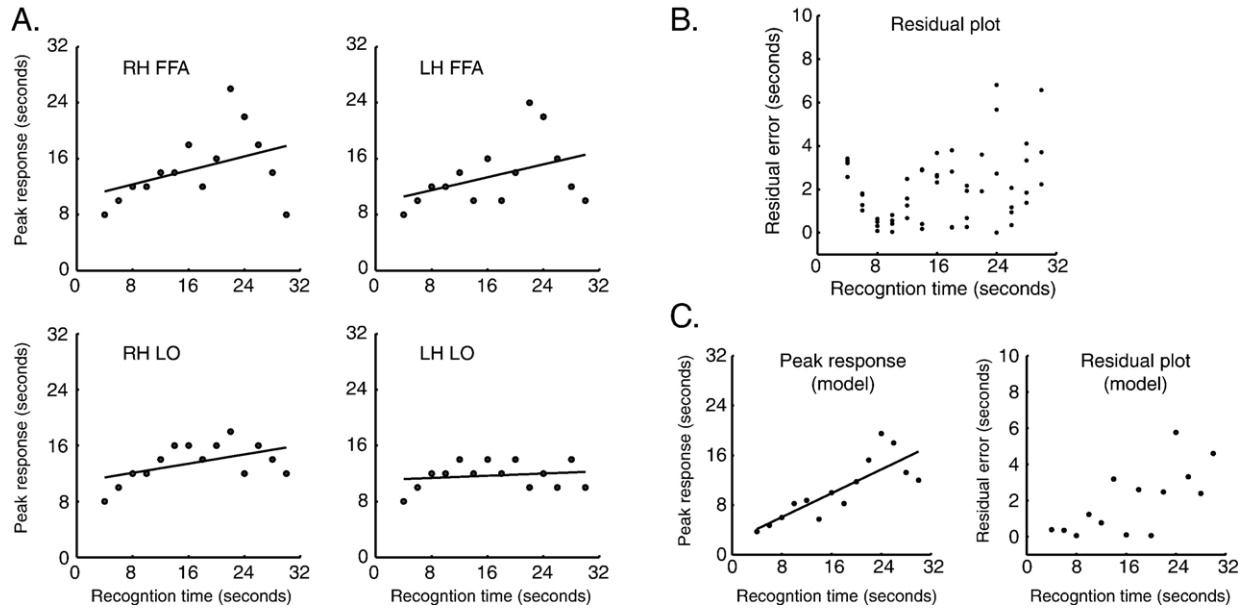


Fig. 4. Sustained vs. Hybrid information model: Peak response. (A) Scatter plots of peak response as a function of recognition time for the four ROIs with linear fits (*indicates significance $P < 0.05$). (B) Residual error from the four ROIs after fitting a simple linear function. (C) Modeled data for the sustained response model. The two plots show the peak response as function of recognition time, and residual error as function of recognition time. The sustained model predicts two prominent features of the data (1) a weak correlation between peak response and recognition time and (2) an increase in residual error as a function of recognition time.

naturally will drive a weak correlation between the peak response and recognition time, even with a sustained response. To demonstrate this, we generated hypothetical data from the sustained response model (Fig. 4C). The modeled data match well with the data from the experiment. The model exhibits a weak correlation between the peak response and recognition times (Fig. 4C Peak response). This is a result of the mean peak time growing larger as a function of time. The model also has the property of an increasing variance as a function of recognition time (Fig. 4C Residual plot). The similarities between the model and the data from the four ROIs, taken with the results of the analysis in the previous section provide strong evidence for sustained activity in these LO and FFA. Consequently, the response in LO and FFA can be interpreted and as task driven, and invariant to the amount of visual information available.

fMRI results: response models

We next performed a global analysis that effectively categorized cortical areas according to response characteristics (i.e., information driven, sustained, and recognition-dependent). Fig. 5 shows the three models with the corresponding predictors for eight consecutive trials from a representative subject. fMRI results are summarized in Fig. 6 and Tables 1–3. Statistical maps were thresholded at $P < 0.05$ for a group random effects analysis ($N = 8$, Bonferroni corrected). The three response models were tested simultaneously using the general linear model for the face and word recognition task. In addition, a conjunction analysis (Price and Friston, 1997) was performed between face and word recognition tasks for each of the three response models, with the goal identifying cortical areas that generalize across the two categories of stimuli in recognition memory. Here, word selection was a critical factor. The words used in the study were drawn from a large number of categories (e.g., objects, actions, occupations, emotions) to deter subjects from visualizing

the solution and instead focus on the structural properties of the word. Tables 1–3 summarize the results for the faces, word, and the conjunction analysis respectively.

Face recognition

Information response model. Significant activations for the information response model were found in the cingulate, frontal

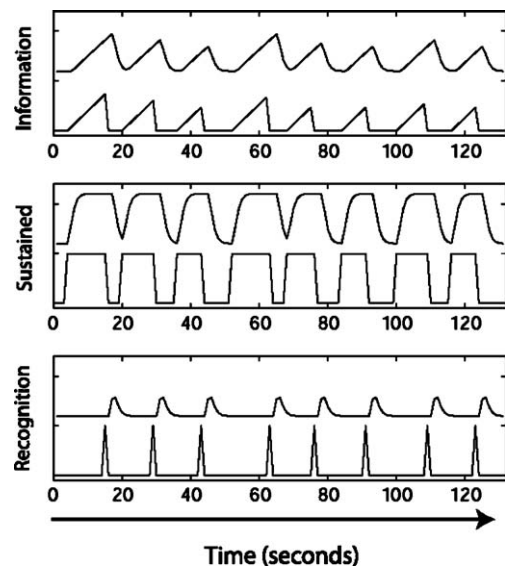


Fig. 5. Modeled time courses for the information, sustained, and recognition response. The lower time course shows the response model for eight consecutive trials from a representative subject performing the word recognition task. The upper time course is predictor for the model constructed by convolving the model with the hemodynamic response.

and occipital cortex. Frontal activations were in the inferior frontal gyrus lateralized in the right hemisphere. Occipital activations were clustered near the calcarine sulcus in early visual areas. *Sustained response model.* A network of brain regions was significantly correlated with the BOLD activation predicted by the sustained response model. Significant effects were found in cingulate cortex, bilateral central sulcus, bilateral intraparietal sulcus, the left insula, the caudate nucleus, the thalamus, bilateral fusiform and lateral occipital regions. *Recognition response model.* BOLD activity corresponding to the recognition response was found in thalamus, bilateral fusiform, supplementary motor areas, the anterior cingulate, right superior frontal sulcus, and the left precentral sulcus.

The analysis also found several cortical areas significantly correlated with multiple response models indicating these regions may be participating in multiple processes. Overlapping activity was observed in the anterior cingulate and inferior frontal gyrus between the information and recognition model and in the thalamus and bilateral fusiform between the sustained and recognition models. To confirm that these regions were indeed overlapping, a conjunction analysis was performed to test for spatial overlap between response models in these regions. The analysis confirmed significant overlap in the anterior cingulate and inferior frontal gyrus between the information and recognition model. There was also a small amount of overlap in the fusiform between the sustained and recognition models, but the regions were largely distinct. Substantial overlap was also observed in the thalamus between the sustained and recognition models.

A reanalysis of the ROIs (RH FFA, LH FFA, RH LO, and LH LO) discussed in the previous section also supported these findings. For the face task, the mean beta value for each of the ROIs was computed individually for each subject for the three response models and then tested for significance across subjects. In LO, only the sustained response model reached significance ($P < 0.01$) across observers, although the recognition response model approached significance ($P = 0.07$). The information model was not significant ($P = 0.48$) in LO. In FFA, both the sustained and recognition response models reached significance ($P < 0.01$), while the information response model remained well below significance ($P = 0.34$), thus confirming the results of both the analysis of the response models and the ROI analysis.

Word recognition

Information response model. Significant patterns of activity correlated with the information model were found in bilateral intraparietal sulcus, bilateral occipital gyri, and the left fusiform. *Sustained response model.* Regions correlated with the sustained response model were found in cingulate sulcus, and precuneus. *Recognition response model.* Areas significantly correlated with the recognition model included supplementary motor areas and the anterior cingulate, the right superior frontal sulcus, bilateral lateral parietal cortex, and bilateral inferior frontal gyrus.

Overlapping patterns of activation for faces and words

No significant overlapping activations were observed in the conjunction analysis between the word and face stimuli for either the information or the sustained response model. A common network of frontal activation was observed for the recognition response model. Specifically, overlap between the word and face stimuli were observed in supplementary motor areas, the anterior

cingulate, bilateral lateral parietal cortex, the right superior frontal sulcus, bilateral inferior frontal gyrus, and the thalamus.

Discussion

Sustained versus information response in LOC

James et al. (2000) used a paradigm very similar to our own to study the effects of priming. The effects of priming were examined in a number of foci including regions in frontal, parietal, and peristriate cortex, and LOC. The study showed that primed stimuli were recognized faster, and this increase in reaction times correlated with the response in LOC. They also observed what they interpreted to be an information-driven response in these foci, which notably included LOC. The results of the present study suggest that the conclusion of an information-driven response in LOC may require further investigation. A qualitative examination of the data from James et al. (2000) and that of our own reveals that the two data sets are remarkable similar. In addition, a quantitative link between the study by James and our own can be found in the observation that both studies found weak correlation between the peak response time and reaction times with roughly equivalent values of 0.45 and 0.3–0.4 respectively. Our model, however, revealed that this weak correlation may be an artifact that can directly be linked to the sustained response model.

Allocation of perceptual/cognitive resources

How does the brain allocate resources for a given task? Is there a minimal cost for maintaining information in a particular representation? These questions are deeply intertwined with how a cortical area represents visual information. The sustained response observed in LO for the face task might suggest there is a fixed neural expenditure for maintaining a representation of an object (a face) that is invariant to perceptually available information. That is, the smallest unit of information in a region that maintains an object-based representation is one object or one face. Alternatively, one might argue the sustained activity in LO reflects ongoing perceptual analysis of new visual information about the object while maintaining a highly efficient representation of old information. Fig. 7A shows the time course for trials where subjects recognized the word and face stimulus in the 16–18 s time range (a single bin from Fig. 2). For the face stimulus, after the expected hemodynamic lag of 6 to 8 s, activity rises in LO and then is maintained at a relatively fixed level (i.e., sustained response). What distinguishes the two aforementioned alternatives is the data from the word experiment. If the sustained activity in LO reflects ongoing perceptual analysis of new visual information, the expectation would be that LO would also exhibit sustained activity for words. LO shows an information-driven response. After the hemodynamic lag, activity rises with increasing visual information. If letters are considered to be objects, an assumption that seems reasonable given letters have unique perceptual forms and previous reports showing overlapping activation patterns between objects and letters in lateral occipital areas (Joseph et al., 2003), then the maintenance of an object-based representation in LO can account for the results of the two experiments. Fig. 7B shows a plot of neural expenditures (i.e., activity) as function of the number of objects. Since the smallest number of objects that can be represented is one, there is

a minimal cost for representing an object even if only part of the object is visible, hence, the sustained response. Above that, additional costs are incurred in whole object units (i.e., information response). Thus, the data are consistent with the maintenance of an object-based representation of visual informa-

tion in LO. Bridging data from the face and word task is less than ideal, but this pattern of results can also be observed in FFA by drawing on the data of from other studies (Fig. 7C). The data from our study show FFA exhibits sustained response even though visual information is increasing. Other studies have shown

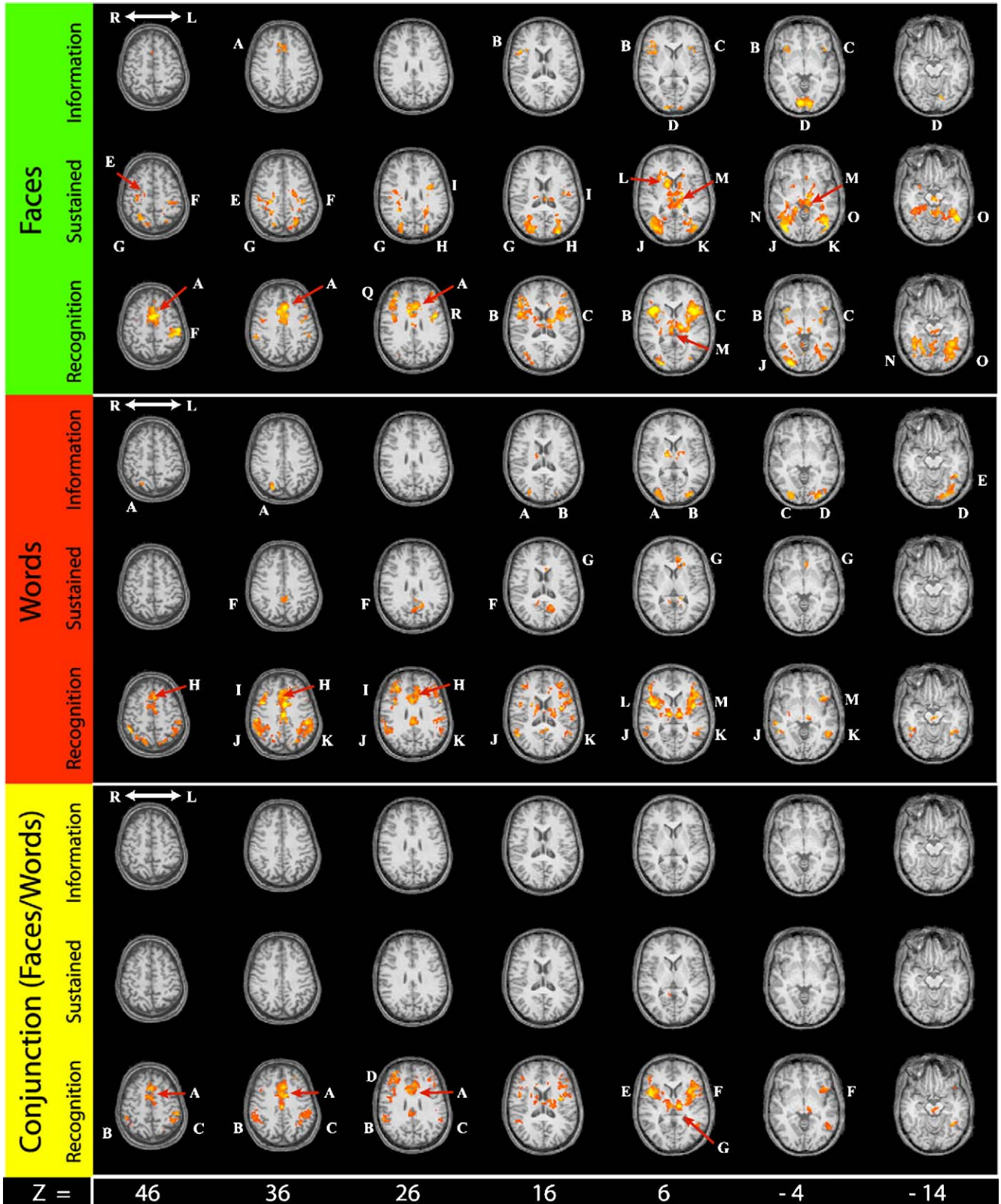


Table 1
Face recognition

Location	Coordinates			Response model (<i>t</i> value)		
	X	Y	Z	I	S	R
SMA/anterior cingulate ^a	1	4	43	4.2	0.1	6.5
R. inferior frontal gyrus ^a	42	19	12	5.2	0.4	7.1
L. inferior frontal gyrus ^a	-36	16	9	4.5	0.3	6.6
R. striate cortex	9	-85	1	4.9	0.0	0.0
L. striate cortex	-7	-88	0	4.8	0.0	0.3
R. central sulcus	27	-24	37	0.0	5.0	1.9
L. central sulcus	-38	-23	40	0.0	4.9	0.0
R. intraparietal sulcus	26	-73	40	0.0	5.0	0.0
L. intraparietal sulcus	-24	-74	43	0.0	5.0	0.0
R. occipital gyri	25	-74	18	0.0	7.4	1.5
L. occipital gyri	-27	-79	18	0.0	6.1	1.4
R. head of the caudate	9	11	9	0.0	6.4	0.0
Thalamus ^a	3	-17	11	0.0	6.1	6.0
R. fusiform ^a	41	-51	-6	0.0	6.0	5.7
L. fusiform ^a	-43	-58	-9	0.0	6.0	5.4
R. superior frontal sulcus	39	38	28	0.2	0.0	6.0
L. precentral sulcus	-39	2	29	0.0	1.1	6.1

BOLD values indicate a significance level of 0.05 or less.

^a Reported coordinates of area were significant for multiple response models.

that the active maintenance of 1, 2, 3, and 4 faces correlates with the magnitude of the response in FFA (Druzgal and D'Esposito, 2001; 2003). These data brought together show a similar pattern of results in FFA as in LO. There is a sustained response for partial information about a face and an increasing response for additional faces. Viewing the data in terms of neural expenditures thus supports the notion that FFA and LO are coding visual information in terms of whole objects.

Can the brain pre-allocate resources with prior knowledge? This is certainly an interesting question and potentially may offer explanation for the observed differences between in the present study and effect reported James et al. (2000). In our study, subjects had prior knowledge they would be performing the face or word recognition task. In the study by James, subjects were unaware of the categorical identity of the object up until the point of recognition. Is it possible then that if the category of a stimulus is known then a fixed amount of resources could be pre-allocated? And, when categorical information is unknown, the brain dynamically recruits the necessary resources. If this was true, then one might expect an increasing response to visual information for categorical unknowns and a sustained response when subjects are aware of the category of the stimulus. This is consistent with the

Table 2
Word recognition

Location	Coordinates			Response model (<i>t</i> value)		
	X	Y	Z	I	S	R
R. intraparietal sulcus	21	-66	40	5.2	0.0	1.8
L. intraparietal sulcus	-21	-64	38	4.9	0.0	2.1
R. occipital gyri	29	-82	0	5.5	0.0	0.3
L. occipital gyri	-27	-82	-1	5.6	0.0	0.7
L. fusiform	-33	-65	-7	4.9	0.0	1.2
R. cingulate sulcus	12	30	8	0.0	4.3	0.4
L. cingulate sulcus	-5	33	3	0.0	4.2	0.5
R. precuneus	6	-51	30	0.0	4.3	0.0
L. precuneus	-9	-52	28	0.1	4.4	0.0
SMA/anterior cingulate	1	3	39	0.0	0.0	4.8
R. superior frontal sulcus	36	35	29	0.0	0.0	5.5
R. lateral parietal cortex	53	-30	36	0.0	0.0	5.4
L. lateral parietal cortex	-50	-28	36	0.0	0.0	5.5
R. inferior frontal gyrus	38	14	9	0.1	0.0	6.0
L. inferior frontal gyrus	-36	12	6	0.3	0.0	5.9

BOLD values indicate a significance level of $P < 0.05$.

results of the face experiment in the present study and the interpretation of the data offered by James et al. (2000). Again, we can address this by returning to our hypothetical response models and the data from the word experiment in LO. In the word experiment, subjects were also aware of the category of the stimulus to be identified. If the visual system pre-allocates resources with prior knowledge, then we would expect a sustained response in LO. As noted earlier, LO showed an information dependent response indicating the visual system does not pre-allocate resources with prior knowledge. It should be noted, however, that a comparison between studies that involve words, faces, and objects is less than ideal for answering this question, and the notion of the pre-allocation of the resources with prior knowledge may still hold promise for future imaging studies.

The evolution of visual information in the ventral stream

The representation of visual information is continually changing, and an important question is how visual information is represented at different stages of visual processing. Image scrambling experiments have previously shown that early visual areas exhibit an increasing response as images are increasingly scrambled, while higher object areas like LO and the fusiform show a decreasing response (Grill-Spector et al., 1998; Lerner et al., 2001). The implication of these findings is that early visual areas are sensitive to local features or parts of an image, while

Fig. 6. Statistical maps of activity for three response models (information, sustained, and recognition). Colored regions exceeded a statistical threshold of the $P < 0.05$ group random effects ($N = 8$, corrected). Significance levels are coded from lowest (red) to highest (yellow). Panel 1: Face recognition. BOLD signal activations corresponding to the information model were found in supplementary motor areas and anterior cingulate (A), bilateral inferior frontal gyrus (B and C), and early visual areas (D). Areas corresponding with the sustained response model were found in bilateral central sulcus (E and F), bilateral intraparietal sulcus (G and H), the left insula (I), bilateral occipital gyri (J and K), the head of the caudate (L), the thalamus (M), and bilateral fusiform (N and O). Significant activations corresponding to the recognition model were found in supplementary motor areas and anterior cingulate (A), bilateral inferior frontal gyrus (B and C), the thalamus (M), bilateral fusiform (N and O), right superior frontal sulcus (Q), and left pre-central sulcus (R). Panel 2: Word recognition. Activity significantly corresponding to the information models were found in bilateral intraparietal sulcus (A and B), bilateral occipital gyri (C and D), and the left fusiform (E). The sustained response model produced significant activations in the precuneus (F), and anterior cingulate (G). Significant activation corresponding to the recognition model were found in supplementary motor areas and the anterior cingulate (H), right superior frontal sulcus (I), lateral parietal cortex (J and K), and bilateral inferior frontal gyrus (L and M). Panel 3: Conjunction between face and word recognition. No overlap was found between the face and word recognition tasks for either the information or sustained response models. Overlapping regions for the recognition response were found in supplementary motor areas and the anterior cingulate (A), lateral parietal cortex (B and C), right superior frontal sulcus (D), bilateral inferior frontal gyrus (E and F), and the thalamus (G).

Table 3
Face and word recognition

Location	Coordinates			Response model (<i>t</i> value)		
	X	Y	Z	I	S	R
SMA/Anterior cingulate	1	8	37	0.0	0.0	5.3
R. lateral parietal cortex	51	-29	36	0.0	0.0	3.6
L. lateral parietal cortex	-52	-27	34	0.0	0.0	3.7
R. superior frontal sulcus	37	38	28	0.0	0.0	4.1
R. inferior frontal gyrus	41	9	9	0.1	0.0	5.7
L. inferior frontal gyrus	-37	8	8	0.2	0.0	5.4
R. thalamus	8	-15	8	0.9	0.0	5.0
L. thalamus	-10	-16	9	0.1	0.0	5.6

BOLD values indicate a significance level of $P < 0.05$.

higher order object areas (e.g., LO) are sensitive to whole objects. Focusing exclusively on the data from the face experiment, we were able to visualize evolving representation(s) of visual information in the ventral stream. Three unique representations of visual information, image-based, object-based, and semantic labeled, can be captured by the three hypothetical response models: information response, sustained response, and recognition response respectively. By incrementally revealing small patches of the image, we hypothesized cortical areas that maintain an image-based representation of visual information would show an increasing response over time (i.e., information model). The results of the experiment showed that early visual areas are correlated with the information response model, which is consistent with the retinotopic organization of early visual areas (Serenio et al., 1995), generally considered an image-based representation. The prediction for an object-based representation is less clear, but it is intriguing that both LO and FFA exhibited a sustained response. If one accepts the notion that LO and FFA maintain an object-based representation of visual information, the observed sustained response suggests that the representation of an object can be instantiated with minimal information, is invariant

to new information about that object, and is released when a more compact representation of the object (e.g., a semantic or categorical label) becomes available. Finally, the recognition response reflects the semantic representation of an object as it occurs exactly at the moment of recognition. Our analysis of the recognition response model identified a bilateral activation in the fusiform medial to regions to areas exhibiting a sustained response. This indicates that cortical areas coding higher order representations are in close proximity to those maintaining object-based representation of the stimulus.

Visual information, attention, and intraparietal sulcus (IPS)

One of the motivations for the present study was to determine if a common network of cortical areas was involved when subjects were actively attempting to match visual information to memory. Our initial hypothesis was based on the idea that these areas would show a sustained response for both the face and word recognition task, reflecting the active process of recognition. Strikingly, the conjunction analysis of the word and face stimuli with the sustained response model produced no significant overlap throughout the brain. This result is interesting, in that one might expect some governing mental process exists during active recognition, the most likely candidate being attention. In retrospect, our initial prediction only took into consideration the mental state of the observer and not the information necessary to complete the task. As such, we reformulated our prediction. This reformulation required an assumption of what constitutes information. We chose to define information in terms of object units. Individual faces were considered to be objects, and based on the previous discussion letters were considered to be objects. This reformulation offers a new prediction for the data. We expect a sustained response for the face experiment, as the task only requires the active maintenance of a single object (i.e., a face). The prediction for the word stimulus changes to an information-driven response, as the task requires the active maintenance of

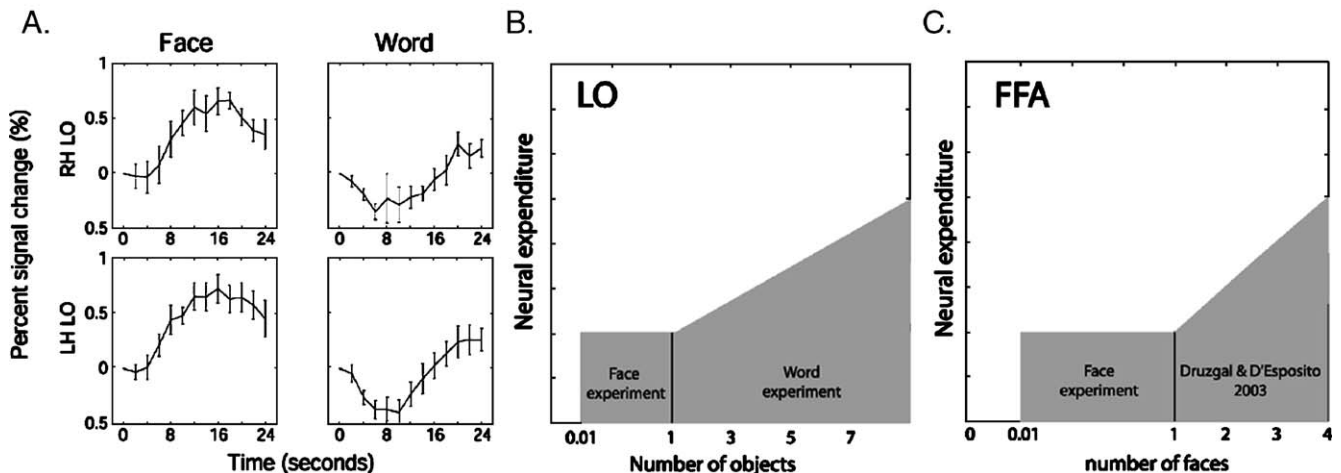


Fig. 7. Neural expenditures for an object-based representation. (A) Activity in LO for the face and word task for stimuli recognized between 16 and 18 s. After the hemodynamic lag, activity for faces is sustained at a fixed level, while activity for words rises with increasing visual information. Error bars are ± 1 SE. (B and C) Hypothetical neural expenditures for an object-based representation for LO and FFA respectively. The smallest whole unit in an object-based representation is one object. Consequently, partial information about an object will exhibit the same neural cost as a whole object. Additional costs will be expended in terms of whole objects. LO and FFA both exhibit this response property. LO exhibits a sustained response for partial information about single object (face experiment), and increasing response for actively maintaining multiple objects (word experiment). FFA shows a sustained response for varying amounts of visual information about a single face (Face experiment), and an increasing response to actively maintaining additional faces in memory (Druzgal and D'Esposito, 2003).

multiple objects (i.e., letters) that are additively introduced over time. Post hoc, we tested a conjunction between the sustained response model for faces and the information-driven response for words. Statistical maps and time courses for the face and word task are shown in Fig. 8. Significant activations were found in left the fusiform and bilateral IPS and LO consistent. An analysis of the regressors across subjects for the ROIs confirmed a highly significant interaction ($P < 0.01$) between stimulus category (face vs. words) and response model (sustained vs. information).

The results of the post hoc analysis identified three regions significantly correlated with the new prediction that takes the information necessary to perform the task into account. Activity in the fusiform and LO presumably reflects perceptual processing of the object. IPS has been shown to participate in a number of

attentionally demanding tasks (Wojciulik and Kanwisher, 1999) and has recently been identified as critical locus for visual short-term memory (vSTM) (Todd and Marois, 2004). In the study by the Todd and Marois, subjects maintained a vSTM representation of the spatial location of a varying number of colored discs. The authors manipulated vSTM load by changing the number of discs and found activity in IPS correlated with subject's behavior in the memory task. Interestingly, activity in IPS in both their study and ours was highly correlated with task relevant extrastriate regions. In their study activity in IPS closely mirrored that of area V4, an area believed to be involved in processing color information (McKeefry and Zeki, 1997). Activity in IPS in our study mirrored the activity of LO and the fusiform, areas implicated in perceptual shape and object processing (Grill-Spector, 2003; Grill-Spector et al., 2001). Furthermore, this effect appears to generalize beyond visual memory tasks. IPS has also been shown to be modulated by attentional load during multiple object tracking (Culham et al., 2001). And again, activity in IPS was echoed by task relevant extrastriate regions. In this case, areas V3 and MT were modulated by the perceptual load. The common threads across these very different studies is (1) activity in IPS is modulated by the perceptual load of the task and (2) activity in task relevant extrastriate areas (i.e., V3(motion), V4(color), LO(shape)) mirrors that of IPS. In a separate experiment, Todd and Marois showed that the link between activity in IPS and extrastriate areas breaks down when subjects maintain the visual information in memory for an extended period of time (~9 s). Activity in IPS maintained a level of activity consistent with the perceptual load but activity in extrastriate areas fell off. Our data show that the link between activity in IPS and extrastriate areas can be maintained for as long as 32 s (the duration of an individual trial). The critical difference, we believe, is whether the visual information is perceptually available to the observer. In their study, a visual memory of the stimulus was obligatory as color discs were no longer perceptually available. The stimuli in our study were always visible. Together, these studies suggest a more general role for IPS than the maintenance of vSTM information. The role of IPS may be to maintain an active link between task relevant visual information and subsequent stages of processing and, when necessary, maintain a sparse representation of visual information when it is no longer perceptually available. If one considers the involvement of IPS in a variety of attentional tasks (Wojciulik and Kanwisher, 1999), this observation is dovetails with view that the role of attention is export visual information from modular representations of for example color, motion, and form to higher level function (Cavanagh, 1999).

Temporally distinguishing events in object recognition: the role of inferior frontal gyrus in object recognition

One of the main shortcomings of fMRI is the limited temporal resolution. This lack of temporal resolution makes it difficult to draw conclusions about the temporal sequence of events in the brain, which can be critical to the theoretical implications of the data. In the domain of object recognition, for example, it has been proposed that a top down mechanism exists in frontal cortex that facilitates object recognition (Bar, 2003). Testing such theories using functional neuroimaging experiments, thus far, either requires attempting to extract a temporal sequence of events based on response latencies (e.g., Pernet et al., 2004) or basing one's conclusions on subtle aspects of the data. The

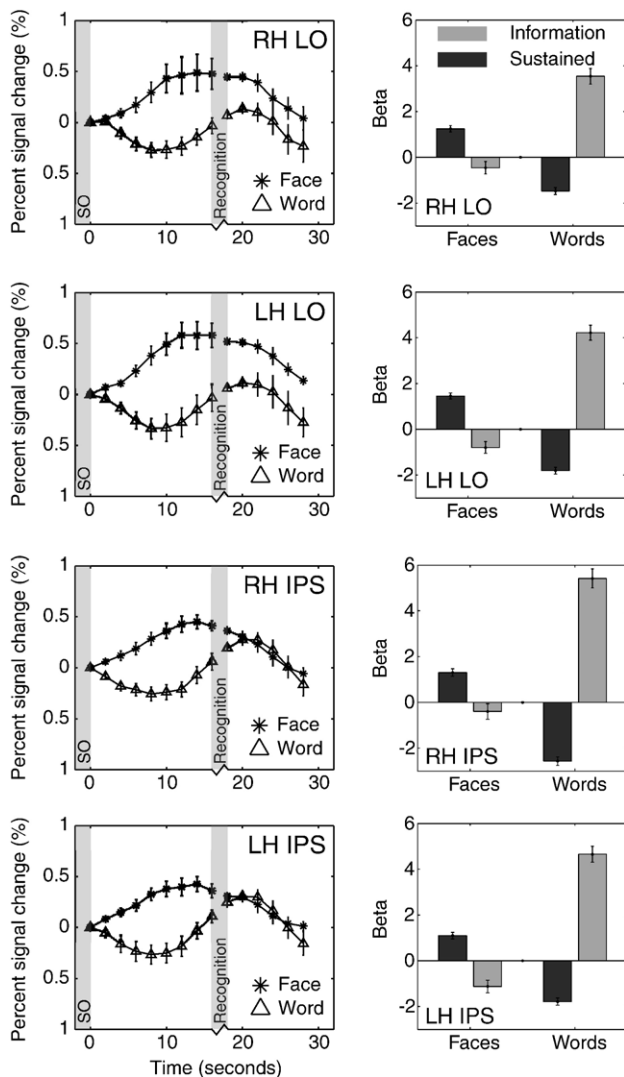


Fig. 8. Word-information/Face-sustained conjunction. Leftmost plots are split time courses for LO and IPS for face and word stimuli. The first half of the time series is an event related average starting from stimulus onset. The second half is an event related average beginning at the moment of recognition. Error bars are ± 1 SE. Rightmost plots show beta values for the response models for the word and face conditions. All four areas (RH LO, LH LO, RH IPS, LH IPS) showed a statistically significant interaction ($P < 0.05$) between response model (information vs. sustained) and stimulus category (faces vs. words).

process of recognition is quite rapid, and as such, the former requires an assumption that the hemodynamic response properties are uniform throughout the brain, which is certainly not the case. The latter requires a leap of faith in the interpretation of the data. Bar et al. (2001), for example, observed that activity in the inferior frontal gyrus (IFG) was greater when subjects were able to recognize a masked presentation of a picture than unmasked and masked presentations that subjects failed to recognize. The authors noted that “it cannot be determined whether this frontal activity started before or after recognition was accomplished”. That is, the authors were unable to distinguish between two alternative explanations of the activity in IFG. (1) Activity was related to top–down effects in object recognition (Bar, 2003; Bar et al., 2001), or (2) activity was related to semantic analysis involved in recognition (Gabrieli et al., 1998). While the present study was not explicitly designed to test these alternatives, the data can be used to make this distinction. Interestingly, activity in IFG was correlated with two of the response models (see Fig. 5). IFG exhibited an event related response time locked with the moment of recognition (i.e., recognition response), thus supporting the interpretation that activity in IFG is related to semantic analysis. Activity in IFG was also significantly correlated with an information-driven response. The data from the Bar et al. (2001) could also be interpreted as a dual response. While the greatest response in IFG was to the masked recognized stimulus, IFG also responded to both the unmasked recognized stimuli and masked stimuli that subjects failed to recognize. One might view these two responses as separate processes, a recognition response corresponding to the semantic analysis, and some other cognitive processes present when subjects perform the difficult masking task. Bar et al. (2001) cautiously offered the intriguing possibility that the latter may reflect top–down processes in object recognition. Here, we are able to show the two responses are separable in time and furthermore are able to characterize the nature of the response. Our interpretation of their proposal is that IFG would exhibit a sustained response reflecting active top–down processes. The observed information-driven response suggests an alternative hypothesis. Since the response correlates with the visual information, it may be attributed to working memory. This explanation receives support from previous functional imaging experiments (Courtney et al., 1998; Courtney et al., 1997), and offers an alternative explanation for Bar et al. (2001) data. That is, when stimuli are presented briefly and subsequently masked, subjects may obligatorily try to maintain the visual information in working memory to allow recognition processes additional time to analyze the stimulus. This would explain the response in IFG to masked, but unrecognized stimuli. The larger response to recognized masked stimuli might then reflect the additive response of the demand for working memory and semantic analysis after recognition.

Neural correlates of explicit recognition

What changes occur in the brain when a person moves from a state of unknown to a state of recognition? Subjectively, observers reported the moment of recognition had a sudden onset. This rapid transition from the unknown to the known was modeled as the recognition response. The results of the face and word experiments revealed cortical areas both unique to the two classes of stimuli and overlapping regions. Successful recognition of faces elicited a unique recognition response in the fusiform

gyrus consistent with recent findings indicating this region plays a important role in face identification (Grill-Spector et al., 2004). For words, a selective recognition response was observed in lateral parietal regions, an area observed to be correlated with successful retrieval of words (Konishi et al., 2000). In performing the conjunction analysis between the face and word task, our goal was identify cortical areas that generalize beyond the category of stimulus. The results of the analysis revealed a network of frontal regions commonly observed in studies of episodic memory retrieval (e.g., Buckner et al., 1998). While the exact contribution of these regions to explicit recognition (e.g., retrieval, error monitoring, decision making) remains elusive, the observation that these regions come online specifically at the moment of recognition irrespective of the class of stimuli is useful to constrain present and future hypothesis about the contribution of these regions to recognition memory.

Conclusions

Visual recognition to the observer is effortless and subjectively instantaneous. The process of recognition is complex, as can be measured by the considerable amount of research dedicated to the topic. In functional neuroimaging, the approach most often taken to overcome to the complexities of recognition has been that of reductionism. That is, to attempt to understand a complex system by carefully examining smaller pieces of the whole. This approach has indeed been very successful in gleaning a better understanding of the process of recognition implemented in the brain. The present study takes a more holistic approach to the complexities of recognition. By slowing the process down and characterizing the response properties in terms of the information represented, we gain a unique perspective of the recognition process that reveals interesting properties of the system and that can constrain theories of recognition memory.

Acknowledgments

We thank Sheng He and David Andresen for assistance in previous versions of this work. This research was supported by the Netherlands Organisation for Scientific Research (NOW Pionier).

Appendix A. Faces

Kofi Annan	Jennifer Aniston	Pierce Brosnan
George W. Bush	Nicholas Cage	Jimmy Carter
Jackie Chan	Prince Charles	Bill Clinton
Hillary Clinton	George Clooney	Courtney Cox
Tom Cruise	Danny Devito	Princess Diana
Minnie Driver	Albert Einstein	Morgan Freeman
Harrison Ford	Mel Gibson	Michael Gorbachev
Tom Hanks	Adolph Hitler	Elizabeth Hurley
Saddam Hussein	Michael Jackson	Samuel L. Jackson
Michael Jordan	John F. Kennedy	John Kerry
Martin Luther King Jr.	Osama Bin Laden	Madonna
Bill Murray	Jack Nicholson	Brad Pitt
Colin Powell	Prince	Ronald Regan
Adam Sandler	Jerry Seinfeld	O.J. Simpson
Britney Spears	John Travolta	Denzel Washington
Oprah Winfrey	Bruce Willis	Tiger Woods

Appendix B. Words

Accident	Activist	Aircraft	Airplane	American
Analysis	Audience	Birthday	Boundary	Building
Catholic	Ceremony	Chairman	Champion	Chemical
Children	Computer	Conflict	Congress	Creature
Criminal	Darkness	Database	Daughter	Deadline
Designer	Document	Election	Employee	Engineer
Equation	Estimate	European	Evidence	Exercise
Festival	Football	Forecast	Fragment	Graduate
Hardware	Hospital	Hydrogen	Judgment	Landlord
Magazine	Marriage	Medicine	Memorial	Minority
Molecule	Monopoly	Monument	Mountain	Musician
Newspaper	Observer	Official	Opponent	Painting
Parallel	Politics	Predator	Pressure	Prisoner
Protocol	Reaction	Recovery	Religion	Research
Resource	Sandwich	Schedule	Scrutiny	Sequence
Shopping	Software	Standard	Strategy	Surprise
Sympathy	Teenager	Traffic	Triangle	Universe
Variable	Velocity	Violence	Wildlife	Workshop

References

- Bar, M., 2003. A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609.
- Bar, M., Tootell, R.B., Schacter, D.L., Greve, D.N., Fischl, B., Mendola, J.D., Rosen, B.R., Dale, A.M., 2001. Cortical mechanisms specific to explicit visual object recognition. *Neuron* 29, 529–535.
- Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J., 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* 16, 4207–4221.
- Brainard, D.H., 1997. The psychophysics toolbox. *Spat. Vis.* 10, 433–436.
- Buckner, R.L., Koutstaal, W., Schacter, D.L., Wagner, A.D., Rosen, B.R., 1998. Functional-anatomic study of episodic retrieval using fMRI: I. Retrieval effort versus retrieval success. *NeuroImage* 7, 151–162.
- Carlson, T.A., Schrater, P., He, S., 2003. Patterns of activity in the categorical representations of objects. *J. Cogn. Neurosci.* 15, 704–717.
- Cavanagh, P., 1999. Attention: exporting vision to the mind. In: Musio, C. (Ed.), *Neuronal Basis and Psychological Aspects of Consciousness*, C. T.-F. a. World Scientific, Singapore, pp. 129–143.
- Courtney, S.M., Ungerleider, L.G., Keil, K., Haxby, J.V., 1997. Transient and sustained activity in a distributed neural system for human working memory. *Nature* 386, 608–611.
- Courtney, S.M., Petit, L., Maisog, J.M., Ungerleider, L.G., Haxby, J.V., 1998. An area specialized for spatial working memory in human frontal cortex. *Science* 279, 1347–1351.
- Culham, J.C., Cavanagh, P., Kanwisher, N.G., 2001. Attention response functions: characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron* 32, 737–745.
- Doniger, G.M., Foxe, J.J., Murray, M.M., Higgins, B.A., Snodgrass, J.G., Schroeder, C.E., Javitt, D.C., 2000. Activation timecourse of ventral visual stream object-recognition areas: high density electrical mapping of perceptual closure processes. *J. Cogn. Neurosci.* 12, 615–621.
- Doniger, G.M., Foxe, J.J., Schroeder, C.E., Murray, M.M., Higgins, B.A., Javitt, D.C., 2001. Visual perceptual learning in human object recognition areas: a repetition priming study using high-density electrical mapping. *NeuroImage* 13, 305–313.
- Druzgal, T.J., D'Esposito, M., 2001. Activity in fusiform face area modulated as a function of working memory load. *Brain Res. Cogn. Brain Res.* 10, 355–364.
- Druzgal, T.J., D'Esposito, M., 2003. Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. *J. Cogn. Neurosci.* 15, 771–784.
- Formisano, E., Goebel, R., 2003. Tracking cognitive processes with functional MRI mental chronometry. *Curr. Opin. Neurobiol.* 13, 174–181.
- Formisano, E., Linden, D.E., Di Salle, F., Trojano, L., Esposito, F., Sack, A.T., Grossi, D., Zanella, F.E., Goebel, R., 2002. Tracking the mind's image in the brain I: time-resolved fMRI during visuospatial mental imagery. *Neuron* 35, 185–194.
- Gabrieli, J.D., Poldrack, R.A., Desmond, J.E., 1998. The role of left prefrontal cortex in language and memory. *Proc. Natl. Acad. Sci. U. S. A.* 95, 906–913.
- Grill-Spector, K., 2003. The neural basis of object perception. *Curr. Opin. Neurobiol.* 13, 159–166.
- Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzhak, Y., Malach, R., 1998. A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum. Brain Mapp.* 6, 316–328.
- Grill-Spector, K., Kourtzi, Z., Kanwisher, N., 2001. The lateral occipital complex and its role in object recognition. *Vision Res.* 41, 1409–1422.
- Grill-Spector, K., Knouf, N., Kanwisher, N., 2004. The fusiform face area subserves face perception, not generic within-category identification. *Nat. Neurosci.* 7, 555–562.
- James, T.W., Humphrey, G.K., Gati, J.S., Menon, R.S., Goodale, M.A., 2000. The effects of visual object priming on brain activation before and after recognition. *Curr. Biol.* 10, 1017–1024.
- Joseph, J.E., Gathers, A.D., Piper, G.A., 2003. Shared and dissociated cortical regions for object and letter processing. *Brain Res. Cogn. Brain Res.* 17, 56–67.
- Kanwisher, N., McDermott, J., Chun, M.M., 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, K.J., Frackowiak, R.S., 2002. The neural structures expressing perceptual hysteresis in visual letter recognition. *Neuron* 34, 659–666.
- Konishi, S., Wheeler, M.E., Donaldson, D.I., Buckner, R.L., 2000. Neural correlates of episodic retrieval success. *NeuroImage* 12, 276–286.
- Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., Malach, R., 2001. A hierarchical axis of object processing stages in the human visual cortex. *Cereb. Cortex* 11, 287–297.
- McKeefry, D.J., Zeki, S., 1997. The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. *Brain* 120 (Pwt. 12), 2229–2242.
- Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442.
- Pernet, C., Franceries, X., Basan, S., Cassol, E., Demonet, J.F., Celsis, P., 2004. Anatomy and time course of discrimination and categorization processes in vision: an fMRI study. *NeuroImage* 22, 1563–1577.
- Price, C.J., Friston, K.J., 1997. Cognitive conjunction: a new approach to brain activation experiments. *NeuroImage* 5, 261–270.
- Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268, 889–893.
- Talairach, J., Tournoux, P., 1988. *Co-planar Stereotactic Atlas of the Human Brain*. Thieme, New York.
- Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. *Nature* 381, 520–522.
- Todd, J.J., Marois, R., 2004. Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 428, 751–754.
- VanRullen, R., Thorpe, S.J., 2001. The time course of visual processing: from early perception to decision-making. *J. Cogn. Neurosci.* 13, 454–461.
- Wojciulik, E., Kanwisher, N., 1999. The generality of parietal involvement in visual attention. *Neuron* 23, 747–764.