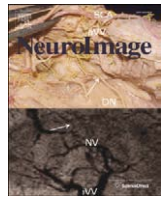




Contents lists available at ScienceDirect

NeuroImage

journal homepage: [www.elsevier.com/locate/ynimg](http://www.elsevier.com/locate/ynimg)

## Specialization for written words over objects in the visual cortex

Marcin Szwed<sup>a,c,f,\*</sup>, Stanislas Dehaene<sup>a,c,d,g</sup>, Andreas Kleinschmidt<sup>a,c,g</sup>, Evelyn Eger<sup>a,c</sup>, Romain Valabrègue<sup>h</sup>, Alexis Amadon<sup>c</sup>, Laurent Cohen<sup>b,e,f</sup>

<sup>a</sup> INSERM U992, Cognitive Neuroimaging Unit, IFR 49, Gif sur Yvette, France

<sup>b</sup> Université Pierre et Marie Curie-Paris 6, Faculté de Médecine Pitié-Salpêtrière, IFR 70, Paris, France

<sup>c</sup> CEA, NeuroSpin Center, IFR 49, Gif sur Yvette, France

<sup>d</sup> Collège de France, Paris, France

<sup>e</sup> AP-HP, Groupe hospitalier Pitié-Salpêtrière, Department of Neurology, Paris, France

<sup>f</sup> INSERM, ICM Research Center, UMRS 975, Paris, France

<sup>g</sup> Université Paris XI, Orsay, France

<sup>h</sup> CENIR, Université Paris 6, Faculté de Médecine Pitié-Salpêtrière, Paris, France

### ARTICLE INFO

#### Article history:

Received 27 September 2010

Revised 26 December 2010

Accepted 28 January 2011

Available online xxxx

#### Keywords:

High-level vision

Letter recognition

Object recognition

Perceptual learning

Visual word form recognition

### ABSTRACT

The Visual Word Form Area (VWFA) is part of the left ventral visual stream that underlies the invariant identification of visual words. It remains debated whether this region is truly selective for words relative to common objects; why this particular part of the visual system is reproducibly engaged in reading; and whether reading expertise also relies on perceptual learning within earlier visual areas. In this fMRI study we matched written words and line-drawings of objects in luminance, contour length and number of features. We then compared them to control images made by scrambling procedures that kept local features intact. Greater responses to written words than to objects were found not only in the VWFA, but also in areas V1/V2 and V3v/V4. Furthermore, by contrasting stimuli reduced either to line junctions (vertices) or to line midsegments, we showed that the VWFA partially overlaps with regions of ventral visual cortex particularly sensitive to the presence of line junctions that are useful for object recognition. Our results indicate that preferential processing of written words can be observed at multiple levels of the visual system. It is possible that responses in early visual areas might be due to some remaining differences between words and controls not eliminated in the present stimuli. However, our results concur with recent comparisons of literates and illiterates and suggest that these early visual activations reflect the effects of perceptual learning under pressure for fast, parallel processing that is more prominent in reading than other visual cognitive processes.

© 2011 Elsevier Inc. All rights reserved.

### Introduction

Humans are virtually alone in the animal world in their capacity to radically alter their natural and cognitive environment by means of culture. The invention of writing is one of the most important such cultural changes, and the question of how literacy modifies existing brain mechanisms raises many intriguing questions (Price and Devlin, 2003; Dehaene, 2005; Dehaene, 2009). Notably, because reading was invented only ~5400 years ago, there was no sufficient time or evolutionary pressure to evolve a brain system devoted to visual word recognition. Nevertheless, in fluent adult readers, reading engages a well-defined brain region in the left occipito-temporal cortex, the Visual Word Form Area (VWFA), which is reproducibly located across different individuals, scripts and cultures (Cohen et al., 2000; Bolger et al., 2005; Baker et al., 2007).

Several basic questions on the nature of visual word form recognition remain to be answered. First, in expert readers, are there really subparts of the ventral visual system that are specialized enough to respond more strongly to words than to other kinds of visual shapes? We have defined functional specialization as the possibility that the visual system has become, at least in part, attuned to the requirements of reading in a given script (Cohen and Dehaene, 2004). Some researchers have argued that line-drawings of objects activate the VWFA just as much as letter strings, and that letter strings are actually processed by the same general-purpose system for recognition and naming that also processes other common visual objects (Price and Devlin, 2003; Wright et al., 2008; Kherif et al., 2010). This issue bears similarity with the debate as to whether face-specific mechanisms or general-purpose expertise suffices to account for face-related activations in the Fusiform Face Area (Tarr and Gauthier, 2000; McKone et al., 2007). A difficult but important requirement for comparing words vs. objects is to avoid, in as much as possible, all low-level physical differences between the two classes of stimuli. While some studies matched words and objects in luminance (Baker et al., 2007; Ben-Shachar et al., 2007) and overall shape (Ben-

\* Corresponding author at: Inserm-CEA Cognitive Neuroimaging Unit (U992, former U562), CEA/NeuroSpin, Bat 145, Point Courrier 156, F-91191 Gif/Yvette, France. Fax: +33 169087973.

E-mail address: [mfszwed@gmail.com](mailto:mfszwed@gmail.com) (M. Szwed).

Shachar et al., 2007), none of them matched these stimuli for other low-level attributes such as contour length and number of features, nor used appropriately matched low-level control stimuli. There are obviously considerable physical differences between words and objects, and these features alone might have introduced confounds in the search for specialized mechanisms for the two classes of stimuli. Conversely, if voxels more responsive to words than to objects emerged after controlling for low-level features, it would be a strong argument in favor of a partial specialization for reading.

A second open question concerns which levels of the visual system are actually modified through the acquisition of literacy. The VWFA in the fusiform region displays several high-level features specifically associated with reading, such as invariance for case or font (Dehaene et al., 2001; Binder et al., 2006; Vinckier et al., 2007; Levy et al., 2008; Glezer et al., 2009; Qiao et al., 2010). However, words traverse all stages of the visual system. Given what is known about the impact of perceptual learning and expertise (Fahle and Poggio, 2004; Gilbert et al., 2009; Harel et al., 2010), Nazir and colleagues have proposed that even the early visual cortex might develop preferential tuning for letters (Nazir, 2000; Nazir et al., 2004; Sigman et al., 2005; Grainger et al., 2010; Tygdat and Grainger, 2009). Indeed, learning to identify T-like shapes during brief visual presentation induces plasticity in the primary visual cortex (Sigman et al., 2005). The impact of literacy might thus be more widespread than is usually proposed.

A third major issue concerns the reasons why the acquisition of literacy affects a reproducible sub-region within the ventral visual system. The object recognition system encompasses extensive sectors of both fusiform and lateral occipital regions (Kanwisher et al., 1997; Grill-Spector and Malach, 2004). However, the VWFA always develops at a fixed location within the lateral occipito-temporal sulcus bordering the fusiform gyrus, as further demonstrated for example by the anatomy of pure alexia (Cohen et al., 2000; Gaillard et al., 2006). We have proposed the hypothesis that this consistent localization relates to prior properties of the corresponding tissue, which make it particularly suitable to the specific problems posed by the invariant visual recognition of written words (Dehaene, 2005; Dehaene, 2009). These properties include a bias for foveal as opposed to peripheral stimuli (Hasson et al., 2002), a posterior-to-anterior increase in perceptual invariance (Grill-Spector et al., 1998; Lerner et al., 2001), and possibly more direct projection fibers to language areas (Cohen et al., 2000; Epelbaum et al., 2008).

Here, we test a further possibility: the reproducible localization of the VWFA might be due to the fact that this region has an initial bias or preference for geometrical features that are close to those used in letter shapes (Dehaene 2005; Dehaene, 2009). One candidate for such geometrical features is the viewpoint-invariant line junctions that occur at the points where the edges formed by the contours of the objects meet (the “vertices” of the graph formed by the contours). For instance, whenever a part of an object occludes another, their contours tend to join and form a local “T” junction (the horizontal contour occluding the vertical one). Such T, Y or L-shaped junctions provide important cues to three-dimensional shape. Indeed, recently it was found that word recognition (Lanthier et al., 2009; Szwed et al., 2009), similar to object recognition (Biederman, 1987) relies heavily on such line junctions: when they are deleted, recognition of both words and objects is more impeded than when an equivalent amount of contour is deleted from the edges only. Furthermore, it has been shown that all writing systems make use of a similar array of such line junctions, with a reproducible statistical distribution that matches that of natural scenes (Changizi et al., 2006). During the evolution of writing, the shapes of written symbols may have been selected to match the shapes that were already encoded in the visual system because of their usefulness in recognizing objects and scenes (Dehaene, 2005; Changizi et al., 2006; Dehaene, 2009). Under this hypothesis, even prior to learning to read, the VWFA may already

have a bias for recognizing line junctions. (Dehaene, 2005; Dehaene, 2009), and this bias would make it particularly suitable for recognition of written words.

Here, we implemented an original stimulus design (Fig. 1) that allowed us to address those three questions simultaneously. First, we designed tightly matched sets of words and of line drawings. This was achieved by fragmenting their contours and equating several low-level attributes: luminance, contour length and number of vertex features (see Methods). Still, some systematic differences could not be avoided. In particular, objects and words differed in their overall shape, which was horizontal and rectangular for words and more variable for objects. The number of individual fragments was also larger for words, and the individual fragments were on average shorter and less curved (Fig. 1 and Fig. S1). To factor out the impact of these differences on neural activations, we devised scrambled control stimuli (Fig. 1, see Methods), which contained the same individual features and occupied the same overall retinal location as the original stimuli. Subtracting the activations evoked by these scrambled controls allowed us to factor out several remaining low-level differences across categories. Finally, to control for the specific type of grouping by proximity that define individual letters within a word, we devised an additional control: “gestalts”, which were made by re-arranging individual letters' features into pseudo-objects. The gestalts had the same type of grouping into several distinct elements as words (Fig. 1D), and the same amount of contour. Overall, the multiple controls used in our study allowed us to detect regions responding preferentially to words with a higher degree of confidence than in previous studies.

In particular, the use of carefully matched controls allowed us to determine whether early visual cortex processes words differently from objects. In previous comparisons (e.g. Cohen et al., 2002), the control stimuli for words often “over-controlled” for low-level visual differences, for instance by presenting flashing checkerboards that exceeded the words in both spatial extent and luminance, thus subtracting out any early visual activation. With our better matched stimuli, we are in a position to detect putative differences in the responses of early visual cortices to words and objects. In particular, if low-level cortex develops specific letter processing abilities as a result of extensive perceptual learning (Sigman et al., 2005; Gilbert et al., 2009), we should observe larger activations for words compared to scrambled words in occipital areas, but not for objects vs. scrambled objects.

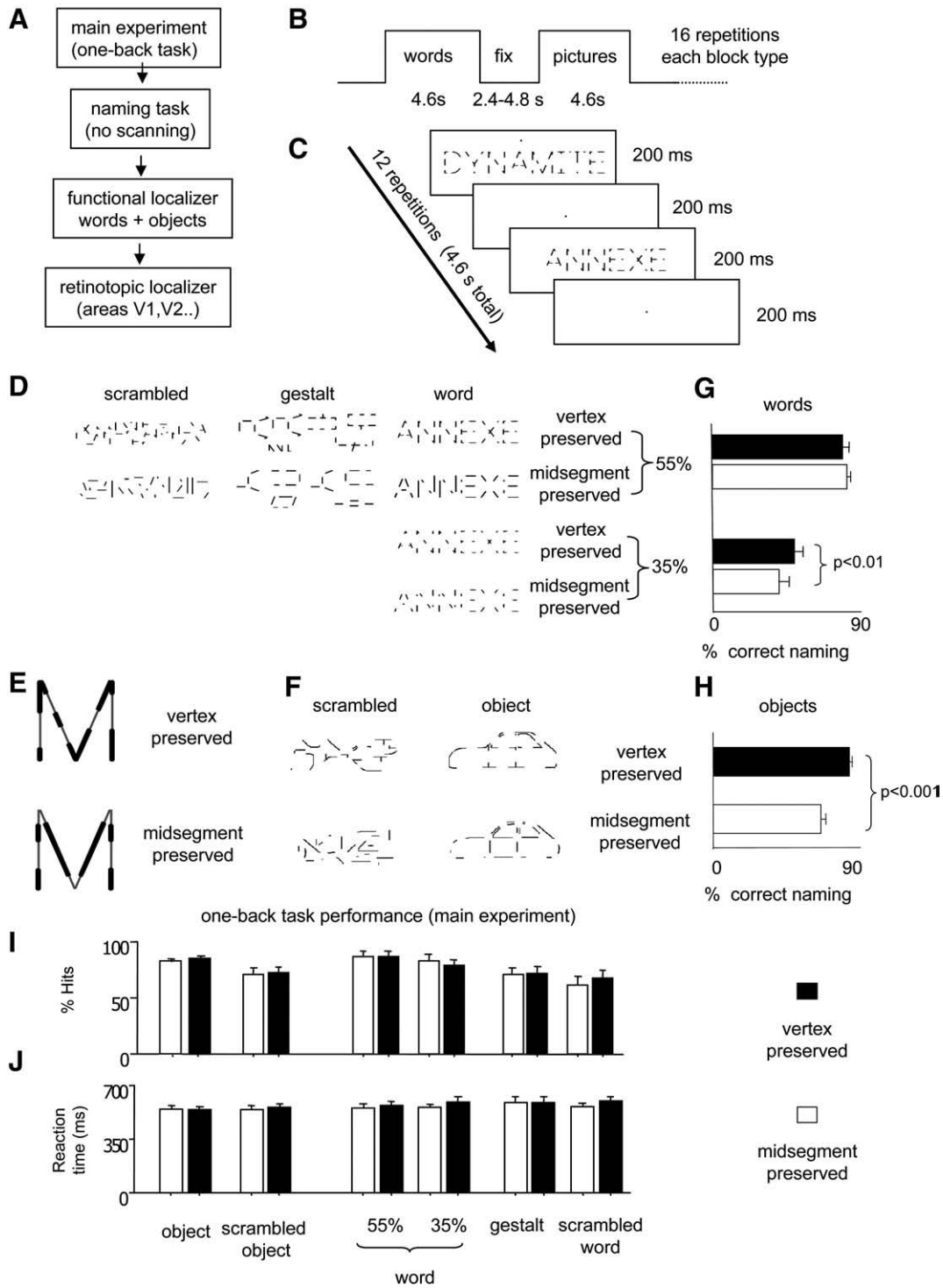
Finally, to test the hypothesis that line junctions play a special role in the origins of the VWFA, we designed two versions of both the words and the line-drawing stimuli. In one version, parts of the lines were deleted so that the stimuli consisted only of the characteristic line junctions occurring at vertices, including a variety of T, Y and L vertices. In the other version, conversely, all line junctions were deleted, so that the words and line drawings were now exclusively made of line segments. Contrasting these two stimulus variants allowed us to test our hypothetical account of the reproducible localization of the VWFA as arising from perceptual mechanisms sensitive to the presence of viewpoint-invariant line junctions (Biederman, 1987; Lanthier et al., 2009; Szwed et al., 2009) might be localized in the fusiform, but not in the lateral occipital sector of the object recognition system.

## Methods

### Outline of the experimental session

The outline of the experimental session is depicted in Fig. 1A. It started with the main fMRI experiment (Figs. 1B–C, see details below), in which subjects saw short blocks with the various versions of the words, objects, and scrambled stimuli. Next, while they were still in the magnet, subjects performed an overt naming task on the same

Design and behavioral results



**Fig. 1.** Design and behavioral results. (A) Overall flowchart of the experimental session. Following the main experiment, subjects performed an overt naming task on the same objects and words. Then, they underwent a functional localizer scan which allowed us to define the object recognition system and the VWFA independently of the main session. Finally, visual areas were localized with the meridian mapping method. (B) There were 16 blocks for each type of stimuli. Blocks were separated by fixation periods that lasted either 2.4, 3.6 or 4.8 s. (C) In each block, 12 stimuli of one given type were displayed for 200 ms separated by a 200 ms fixation screen (4.6 s total block duration). (D–F) Words and objects were degraded by partial deletion of some of their component lines, leaving either the vertex features or the midsegment features. In control stimuli, objects and words were scrambled in a way that kept the individual features intact. (D) Examples of words, scrambled words and ‘gestalts’. For words, either 55% (top) or 35% (bottom) of original line length was preserved. In scrambled words and objects, fragments were randomly shuffled, while in word ‘gestalts’, they were recomposed into pseudo-objects that had the same amount of collinearity and grouping as words. (E) A sample letter in both variants. (F) Example object stimuli (see Fig. S1 for more examples). (G–H) Naming performance for words (G) and objects (H) with midsegment preserved (white bars) and vertex preserved (black bars) features. Accuracy did not differ between equally degraded words and objects, and was better for the vertex preserved than for the midsegment preserved version of the highly degraded stimuli, both for words and for objects. (I–J) Performance in the one-back repetition detection task performed during the main experiment. (I) Percentage of correct responses did not differ between words and objects, and was somewhat lower for meaningless control stimuli. Response times did not differ across all stimuli.

words and objects they just saw. No fMRI images were acquired during this task. The procedure was similar to the one described previously (Szwed et al., 2009). Stimuli were presented for 200 ms, and subjects were instructed to name them as quickly as possible while minimizing errors. This allowed us to assess the level of recognition of each type of stimuli, and to correlate the BOLD signal with the subjects' behavior. Then, subjects underwent a short functional localizer scan which allowed us to define the object recognition system (intact objects vs. objects scrambled with a  $20 \times 20$  pixel grid) and the VWF system (intact words vs. checkerboards), independently of the main experiment. Finally, the subject's retinotopic visual areas were mapped using the meridians method (Claeys et al., 2004).

#### *Stimuli for the main fMRI experiment*

Stimuli were derived from printed words and line drawings of objects. The original words and drawings were degraded by removal of line fragments (Figs. 1D–F; Supplementary Fig. S1 contains further examples of object stimuli). Two modes of degradation were used, depending on the type of visual features which were preserved: in the 'vertex preserved' variant, the line junctions were preserved, while in the 'midsegment preserved' variant they were suppressed (Figs. 1D–F).

The selection of objects and words and the definition of vertices and line midsegments are described in detail elsewhere (Szwed et al., 2009). Briefly, word stimuli ( $n = 48$ ) consisted of 6–8 letter French nouns with a frequency higher than one per million (median = 8.7). We chose a line width and font size allowing us to match satisfactorily luminance and contour length across words and objects. Line-drawings of objects ( $n = 72$ ), including images from the Snodgrass and Vanderwart set (1980), were simplified from their original versions by removing textures and redundant details to better match them to words; line width was also adjusted when necessary. We checked that the resulting images were still recognized at near 100% by running a pilot naming task.

We defined vertices and line midsegments following the principles used by Biederman (1987) and Changizi et al. (2006). Vertices were defined as any junction of 2 or more lines. The transitions of straight lines into curves such as in the letter "J" were also treated as vertices. We defined midsegments as line fragments at least 4 pixels away from any vertices. In the curvy parts of some letters, when distinct vertex and midsegment deletions could not be defined (e.g. anywhere in the letter "S"), identical deletions were made in the vertex preserved and midsegments versions.

Word and object sets were matched in average number of vertices (5% difference in mean vertex count between words and objects), in luminance, and in total contour length. Naturally, objects and words differed in their overall shape, which was rectangular for words and more variable for objects. Moreover, because words consisted of separate letters and thus were more fragmented to begin with, their component fragments were on average somewhat more numerous and shorter than in objects. To control for these differences, for both words and objects, we devised scrambled control stimuli (Figs. 1D,F). They were made by randomly scrambling the fragments while keeping constant the horizontal and vertical dimensions of the image (custom-written Matlab code). The amount of collinearity and grouping among fragments was necessarily reduced in scrambled relative to original stimuli (see Discussion), but an additional control condition, "gestalts", controlled for this factor (see next paragraph). We checked that scrambled objects were impossible to recognize after scrambling by running a pilot naming task.

For words, additional conditions were used (Fig. 1D). As opposed to objects, in which 55% of the contour was always preserved, words were shown with either 55% (top) or 35% (bottom) of their original contour. We used these two levels of degradation following Szwed et al. (2009) who found that for the 35% words, subjects were better at

recognizing words in the vertex preserved form than in the midsegment preserved form, while for the 55% words the recognition levels did not differ. For words, we also devised an additional control condition: 'Gestalt' stimuli were made by re-arranging the individual fragments of letters into pseudo-objects, equal in number to the original number of letters, (Fig. 1D). Just like words, which consisted of 6, 7 or 8 letters, the gestalts had a distinct grouping into 6, 7 or 8 elements. The gestalts therefore preserved an equivalent amount of collinearity and gestalt grouping as in the letters in the original word stimuli. Their low level visual features were as close as possible to words, while at the same time, through arranging them in two rows of horizontally oriented objects, they differed visually from the false fonts used in other studies (e.g. Vinckier et al., 2007; Levy et al., 2008).

The reordering of letter fragments into pseudo-objects was done in CorelDraw and Adobe Photoshop. For 5 letters of the alphabet, small adjustments of straight line length (i.e. lengthening one line by an amount of X pixels, and shortening another one by the same amount) were necessary to achieve good completeness of the resulting pseudo-objects. Measurement tools included in Adobe Photoshop were then used to assess the total line length after filling in the deleted parts of the contours (i.e. the length of physically present contour + the length that can be interpolated on the basis of collinearity). The total line length for gestalts was very similar to the one for letters (average 115 and 116 pixels per pseudoobject/letter, respectively; median difference of line length between corresponding pairs = 5%). However, in order to maintain good separation between individual pseudo-objects inside the gestalts, it was necessary to space them by a greater amount than the typical spacing of letters in words. This manipulation increased the mean height of the entire Gestalt stimuli by 30% relative to words and scrambled words (mean total width was kept identical).

To verify that perceptual grouping operated similarly in gestalts and in words, we ran a pilot behavioral experiment in which subjects saw words, letter strings and gestalts for 200 ms and were asked to report the number of elements/letters (which could be 6, 7 or 8). Subjects were very accurate at reporting the number of letters and the number of pseudo-objects (mean % correct 92%, 90% and 89% for words, strings and gestalts, respectively, difference n.s.  $p = 0.65$ ). Thus, we can conclude that the individual elements in gestalts had the same perceptual grouping as letters in words.

In summary, our study included two complementary controls for words: Gestalts were matched to words in collinearity and grouping, but differed slightly in height, while scrambled words were matched to words in height and width, but exhibited less internal structure and collinearity. We reasoned that any effect common to words > gestalts and words > scrambled would not be attributable to any of the above-mentioned factors, and would likely reflect a partial cortical specialization for alphabetic stimuli.

Overall, a total of  $6 \times 2 = 12$  types of stimuli were used: objects, scrambled objects, words, scrambled words, gestalts, and words with 35% contour, each of those stimuli in both the "Vertex preserved" and the "Midsegment preserved" variants. Stimulus variants were counter-balanced between subjects. Each subject saw any given word in only one out of its four variants, and the object in one of its two possible 'meaningful' versions (e.g., a subject who saw the car in midsegment preserved-55% variant, would not see it in a vertex preserved-55% variant).

#### *Structure of the main fMRI experiment*

The experiment was composed of a series of short blocks, each containing stimuli of one of the 12 possible types, in order to yield maximal activation in the occipitotemporal cortex while minimizing top-down effects (Vinckier et al., 2007) (Fig. 1B). In each block, 12 stimuli were displayed at a fast rhythm: 200 ms presentation duration with 200-ms blank ISI (4.6 s total block duration; Fig. 1C). Each block contained the same 12 stimuli, presented in random order. Blocks

were randomly separated by 2.4, 3.6 or 4.8 s blanks. There were 8 blocks for each type of stimuli in each of the two scanning sessions. The control blocks used scrambled stimuli derived from the same words and objects as shown in the non-scrambled blocks. Subjects performed a one-back repetition detection task; in half the blocks, one stimulus was repeated twice in a row. Subjects were instructed to respond by pressing a button with their right hand.

#### Functional localizer

The functional localizer used to map object-related and reading-related activations independently from the main experiment included 4 types of stimuli: intact objects and intact words (the same stimuli as in the main experiment, without line fragment removal), scrambled objects (scrambled on the basis of a  $20 \times 20$  grid, while keeping the horizontal and the vertical dimensions constant), and alternating checkerboards (a commonly used control condition for words, see for example ref. [Cohen et al., 2002](#)). The same trial structure and block design were used as in the main experiment. The localizer scan lasted 6 min. Each stimulus block was repeated 12 times.

#### Mapping of retinotopic areas

Retinotopic visual areas were mapped using the meridians method ([Claeys et al., 2004](#)). Stimuli consisted of flashing checkerboard wedges covering either the lower vertical meridian, the upper vertical meridian, or both horizontal meridians, as well as flashing checkerboard rings covering either central (2 degrees) or peripheral (beyond 5 degrees) regions of the visual field. The borders of the retinotopic areas (e.g. between V1 and V2) were localized along the line of highest response to the meridian wedge stimulus. The ROIs were defined, in each individual subject, as 50 most significant voxels responding to the central ring stimulus in areas V1/V2 and V3v/V4.

#### Stimulation and acquisition parameters

Stimuli were presented using the E-prime software (PST, Pittsburgh, PA) in the center of the visual field. Objects subtended a visual angle of up to  $3.9 \times 4.6^\circ$ . Words subtended a more elongated field of  $0.8 \times 5^\circ$ . Images were acquired on a 3-Tesla MRI scanner (Siemens Trio TIM) with a 12-channel head coil and a gradient-echo planar high resolution imaging sequence sensitive to brain oxygen-level dependant (BOLD) contrast (32 contiguous axial slices, 1.8 mm thickness, TR = 3000 ms; angle =  $84^\circ$ , TE = 30 ms, in-plane resolution =  $1.5 \times 1.5$  mm, matrix =  $128 \times 128$ , no iPAT acceleration; 6 cm slab covering the ventral and middle parts of the brain, including, notably, ventral parts of the intra-parietal sulcus, the middle and superior temporal gyri and the inferior frontal gyrus). High resolution fMRI ( $1.5 \times 1.5 \times 1.8$  mm voxels) was used to optimize detection of small cortical patches which could be selectively responsive to words. The main experiment was divided into 2 equivalent acquisition sessions, each comprising 8 repetitions of each type of block. T1-weighted images were also acquired for anatomical localization.

#### Subjects

16 right-handed, native French speakers, 18 to 32 year-old (6 men) gave written informed consent to participate in the present fMRI study. They had no history of neurological or psychiatric disease. Their vision was normal or corrected to normal. The project was approved by the regional ethical committee.

#### Analysis

Individual imaging data processing was performed with SPM5 software and included corrections for EPI distortion, slice acquisition

time, and motion; normalization to the MNI anatomical template; Gaussian smoothing (3 mm FWHM); fitting with a linear combination of functions derived by convolving a standard hemodynamic response function with the time series of the stimulus categories. Individual contrast images were computed for each stimulus type minus baseline, then smoothed (3 mm FWHM), and eventually entered in an ANOVA for random effect group analysis. Functional maps were created using the xjview toolbox (<http://people.hnl.bcm.tmc.edu/cuixu/xjView/>). Flattened maps were created using Caret (<http://brainmap.wustl.edu/caret>) and the PALS atlas ([Van Essen, 2005](#)).

We used a voxel-wise threshold of  $P < 0.001$ , with a threshold for cluster extent of  $P < 0.05$  corrected for multiple comparisons across the whole brain, unless stated otherwise. Activation values reported in ROI plots are in arbitrary units proportional to BOLD activation percentage (beta). For the ROI analyses, we sampled the activity within a  $3 \times 3 \times 3$  voxel cube ( $4.5 \text{ mm} \times 4.5 \text{ mm} \times 5.4 \text{ mm}$ ).

For the analysis depicted in [Fig. 4](#), the location of individual ROIs was optimized to take full advantage of the high-resolution fMRI data. We first defined a cylindrical region with 10 mm diameter running along the anteroposterior axis of the ventral occipitotemporal cortex, centered on the peaks from the group analysis of the functional localizer. We then divided this cylindrical region into slices centered on each peak. For example, for the two neighboring average peaks located at  $y = -60$  and  $y = -51$ , the boundary lies at  $y = -55.5$ . Within these slices we selected, for each subject individually, an ROI centered on the voxel that showed maximum activation in the functional localizer scan for that particular subject.

For the analysis of activation asymmetry (Supplementary Fig. S4), individual normalized anatomical images were flipped; and then normalized back to the original anatomy; the corresponding normalization matrices were applied to the flipped contrast images, allowing for an accurate match of the left and right hemispheres; flipped contrast images were then subtracted from the original contrast images. The resulting difference images were smoothed (3 mm FWHM), and were entered in the same ANOVA as before, allowing us to test the interaction of any given contrast with the left/right hemisphere factor.

Negative activation values are commonly observed in fMRI and might introduce strong bias into selectivity analyses ([Simmons et al., 2007](#)). Thus, for our selectivity analysis ([Fig. 4C](#)) we corrected the activation results following the procedure described by [Simmons et al., \(2007\)](#). We identified all ROIs in which the response to words or objects category was less than 0 and added, to the responses to each of the categories and their controls whatever value was necessary to make the smallest response across the categories equal to 0. As an additional precaution, we also rejected regions that even after correction had very low activations for both stimuli (both  $\beta < 0.2$ ). In each of the analyses, fewer than 3% of data points were rejected for such reason.

## Results

#### One-back detection task

During the main scanning session, subjects performed a simple one-back repetition detection task. On the whole subjects were less accurate with the meaningless scrambled control stimuli (69% hits) than with the word and picture targets (84% hits) ([Fig. 11](#);  $p < 0.001$ ). Hit rate did not differ between words and objects, between vertex preserved and midsegment preserved stimuli, and between 35% and 55% word variants. There was also no difference in the rate of false alarms between words and objects (5% and 4%, respectively,  $p = 0.2$ ). Reaction times in the one back task were on average 567 ms, and did not differ significantly between words and objects, between vertex preserved and midsegment preserved variants, between 35% and 55%

word variants, between meaningless scrambled control stimuli and meaningful objects and words. We conclude there was no difference in task set across critical experimental conditions.

#### Overt naming task

Immediately after the main fMRI experiment, subjects performed a naming task in the scanner with the same stimuli. Figs. 1G–H show the effect of stimulus category, degradation level and of feature type (vertex preserved vs. midsegment preserved) on naming performance. In agreement with (Szwed et al., 2009), naming accuracy did not differ between equally degraded words and objects (55% words vs. 55% objects,  $p=0.26$ ). Subjects were better at naming the vertex preserved version than the midsegment preserved version of the highly degraded stimuli (35% variant) both for words ( $p=0.002$ ) and for objects ( $p<0.001$ ). Thus, our behavioral results support the hypothesis that line junctions play a particular role in both object recognition and reading. Note that, because subjects were naming objects that they had already seen in the main fMRI experiment, their performance may have contained some degree of priming. However, the behavioral results obtained are very consistent with those of Szwed et al., (2009), who used the same stimuli with subjects who saw them for the first time. This argues against a significant effect of priming on our conclusions. In the same study, Szwed et al., (2009) also reported that mean response times in the naming task were 716 ms for 55% words and 923 ms for 55% objects.

#### Imaging results: Brain areas for reading and object recognition

We first describe the basic set of regions activated by words and objects relative to their scrambled controls in the main experiment (Figs. 2A–D). Unless stated otherwise, only the 55% variant of word stimuli, which had equal amount of remaining contour as the objects, was included in the analyses.

A listing of areas activated by the main conditions is provided in Table 1. Words activated strongly left-predominant areas (Fig. 2A). This included a left occipitotemporal cluster extending from early retinotopic areas V1 and V2 (MNI  $-10 -96 0$ ,  $Z=4.60$ , left hemisphere; MNI  $15 -92 -3$ ,  $Z=4.50$ , right hemisphere) to the anterior part of the fusiform gyrus (between MNI  $y=-25$ ), through areas V3V/V4 and the mid-fusiform cortex of the Visual Word Form Area (MNI coordinates:  $-45 -41 -18$ ;  $Z>8$ ). Other activated areas included the right fusiform region (MNI  $41 -41 -25$ ,  $Z=5.55$ ), the left superior temporal sulcus/gyrus, and the middle temporal gyrus (STS/STG/MTG; MNI  $-56 -51 5$ ,  $Z>8$ ) and the inferior frontal gyrus (MNI  $-47 23 11$ ,  $Z=4.56$ ). Including hemisphere as a factor in the SPM ANOVA revealed that this increase of activation for words over controls was left-lateralized in the Visual Word Form Area (MNI  $-48 -53 -18$ ;  $Z>8$ ) and backward to V3V/V4 (MNI  $-33 -80 -11$ ,  $Z=4.66$ ,  $p<0.001$ ; Fig. 2A, arrow and Fig. S4), while it was symmetrical in V1/V2. In all those occipital areas, each of the 16 subjects had higher activations for words than for scrambled words. The differences in responses to the 55% and the 35% variants of words were small and restricted to area V1/V2 (see Supplementary result R1).

Besides scrambled words, our study included a second control condition, the gestalts, which were made by re-arranging individual letters into pseudo-objects that had the same amount of contour, collinearity and grouping as in the original words. As shown in Supplementary Fig. S2A, we indeed observed stronger activations for gestalts than for scrambled controls all across the ventral visual system starting approximately from around  $y=-75$ , suggestive of an effect of grouping (Altmann et al., 2003; Kourtzi et al., 2003; Dumoulin and Hess, 2006; Ostwald et al., 2008). However, we found that the responses to words in the left hemisphere visual system were substantially stronger than responses to gestalts (Supplementary Fig. S2B). In fact, activations to words relative to gestalts revealed peaks in

the VWFA (MNI  $-45 -41 -18$ ,  $Z>8$ ) and the occipital areas (MNI  $-18 -96 -11$ ,  $Z=7.25$ ) at virtually the same coordinates as the words–scrambled words contrast. The similarity of the activations to words relative to the two control conditions was confirmed when the same pattern of activation was observed again after masking the words–scrambled controls contrast with the words–gestalts contrast (Supplementary Fig. S2C).

Finally, objects activated a more extensive bilateral set of regions (Fig. 2C), described previously as the ‘object system’ (Grill-Spector and Malach, 2004). It stretched from the Lateral Occipital (LO) region/inferior temporal sulcus (right: MNI  $46 -69 12$ ,  $Z=6.01$ ; left: MNI  $-40 -64 9$ ,  $Z=5.85$ ) downwards to the fusiform gyrus (right: MNI  $38 -42 -23$ ,  $Z>8$ ; left: MNI  $-39, -51, -20$   $Z>8$ ) and to more anterior parts of the inferior temporal lobe (around MNI  $y=-10$ ). Other areas activated included the parahippocampal gyrus (right: MNI  $20 -38 4$ ,  $Z=4.99$ ; left:  $-17 -36 4$ ,  $Z=5.15$ ), and the right cuneus (MNI  $15 -84 14$ ,  $Z=4.25$ ). The pattern of areas activated by intact words and objects during the localizer experiment was very similar (Figs. S3A and B).

#### Selective word-related activation

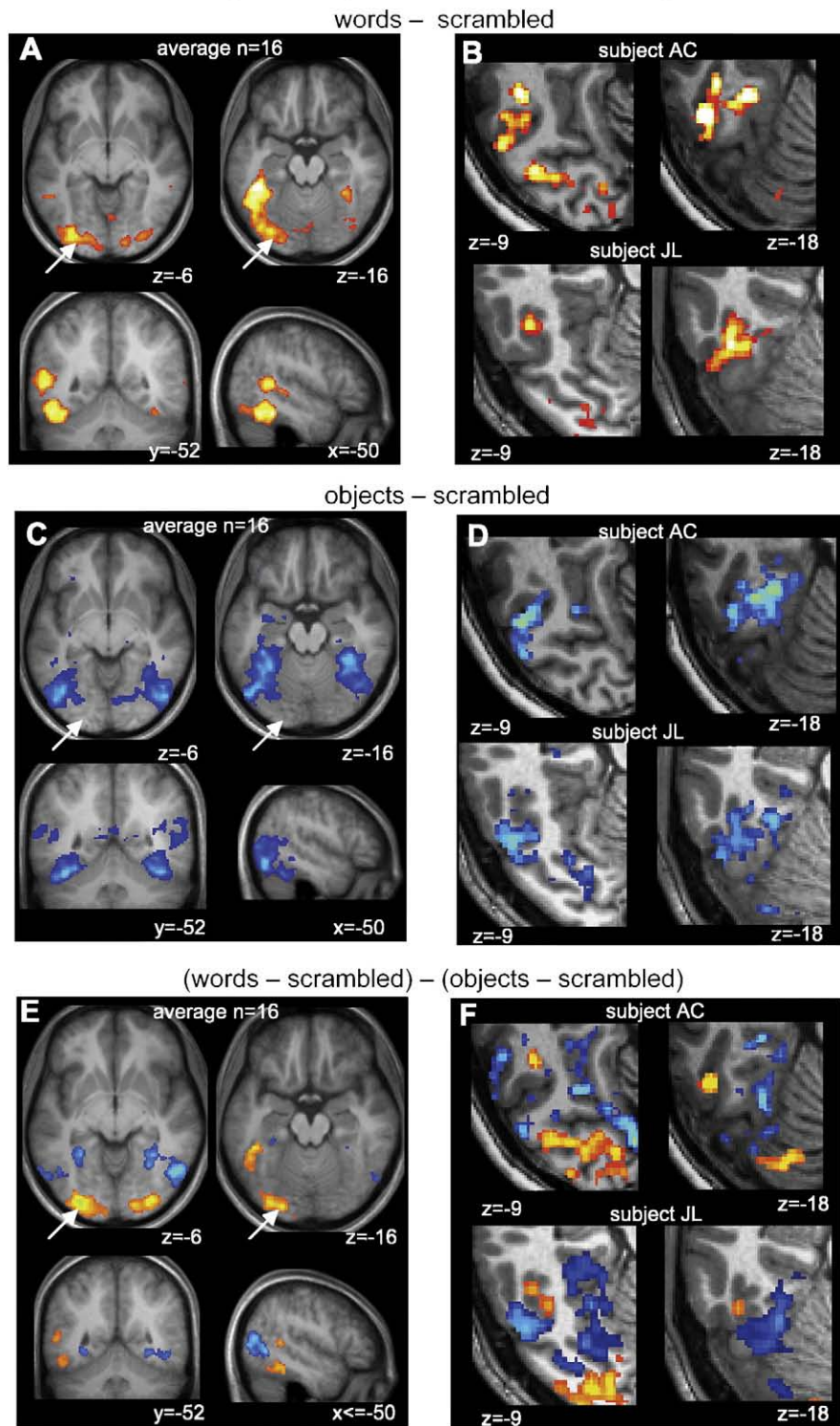
We then looked for preferential activations to words relative to objects. Because our study contained control stimuli that used exactly the same low-level components as words and objects, we were able to contrast the responses to words and objects with their respective controls subtracted (i.e. [words–scrambled words] – [objects–scrambled objects]), thus removing the effects of low-level differences between categories (Figs. 2E–F). This comparison revealed a cluster in the VWFA, peaking at MNI  $-47 -41 -18$  ( $Z=6.87$ ) and extending posteriorly to MNI  $y=-60$ . Other activations included bilateral occipital areas (right: MNI  $30 -96 -4$ ,  $Z=6.6$ ; left: MNI  $-28 -95 -4$ ,  $Z=7.8$ ) and the left STS/STG/MTG (MNI  $-68 -33 5$ ,  $Z=5.69$ ). Figs. S3C,E show the direct comparisons between words and objects, made without subtracting the controls, and the corresponding contrasts for the localizer experiment. All those analyses showed essentially the same results, particularly the existence of a cluster in the VWFA activated more for words than for objects. These activations could also be seen in individual subjects’ maps (Fig. 2F; subject AC: MNI  $-51 -53 -18$ , 144 voxels,  $Z>8$ ; subject JL: MNI:  $-47 -45 -22$ , 249 voxels,  $Z>8$ ). A comparison of panels B, D, and F of Fig. 2 shows that in those subjects some, but not all, of the areas highly activated by words were activated more for words than objects.

#### Further analyses of the expertise for alphabetic stimuli in early visual cortex

We then explored in more detail the bilateral occipital regions more responsive to words than to objects (Figs. 2A,C,E arrows). These activations, also visible in individual subjects (Fig. 2F, left panels) were of particular interest as they may reveal a ‘tuning’ of early visual cortex for the particular shapes used in reading (Nazir, 2000; Nazir et al., 2004; Nazir and Huckauf, 2006). To explore this issue, we examined regions of interest (ROI) located in areas V1/V2 and V3V/V4 (Fig. 3) defined on the basis of the retinotopic localizer (see Methods).

Consistent with previous reports (see for example Grill-Spector and Malach, 2004, Fig. 11A), the early retinotopic visual areas were either equally or more activated by scrambled objects than by intact objects (Fig. 3, blue bars; V1/V2: both  $p<0.006$ ; V3V/V4  $p=0.3$  left and  $p=0.03$  right hemisphere). Remarkably, for words we observed the reverse pattern: words caused more activation than scrambled words (Fig. 3, orange bars, left V1/V2  $p=0.005$ , right V1/V2  $p=0.052$ ; left V3V/V4  $p=0.001$ ; right V3V/V4  $p=0.01$ ). The difference in activation profile between words and objects, as measured by the subtraction of words–scrambled words minus objects–scrambled objects, was significantly left-predominant in both V1/V2 and V3V/V4 (both interactions with hemisphere:  $p<0.01$ ).

## Brain regions responsive to words or to objects



**Fig. 2.** Brain regions responsive to words or to objects. Activations induced by words minus scrambled words (top row), by objects minus scrambled objects (middle row), and by words vs. objects with their respective scrambled controls subtracted (bottom row; hot: words minus objects; cold: objects minus words), in the group of 16 subjects (left column) and in two illustrative subjects (right column). Words induced stronger activations than objects in the left fusiform Visual Word Form region, as well in bilateral occipital areas (arrows). Group results are overlaid on the groups' average normalized T1 anatomy, and individual results are overlaid on individual T1 anatomies. Thresholds:  $p < 0.001$  voxel-wise, and  $p < 0.05$  cluster-wise corrected for multiple comparisons across the whole brain.

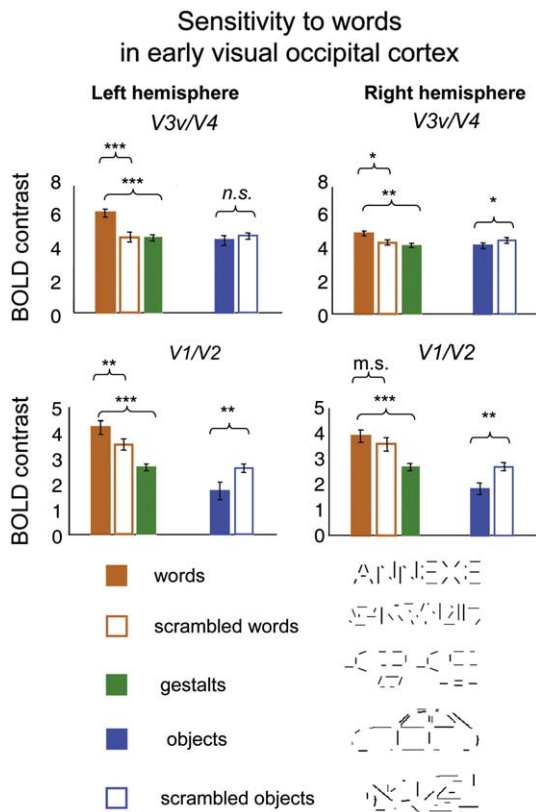
One could argue that this striking difference in activation pattern between words and objects was a consequence of some low-level geometrical properties specific to words as compared to objects (e.g.

shape of stimulus envelope or distribution of spatial frequencies). If this were the case, the word-derived 'gestalts', which share most low-level features with words, should yield activations similar to words,

**Table 1**  
A listing of areas activated by the main conditions.

Contrast	Region	Hemisphere	Z score at $p < 0.001$	MNI coordinates		
Words–scrambled words	Fusiform gyrus, BA37	L	>8	–45	–41	–18
		R	5.5	41	–41	–25
	Middle and superior temporal gyri, BA21/22	L	>8	–56	–51	5
		R	4.9	66	–42	7
	Inferior occipital gyrus, BA17/18	L	5.68	–27	–88	–4
		R	4.6	–10	–96	0
Objects–scrambled objects	Precentral gyrus, BA4	L	3.9	–63	–10	32
		R	3.8	–47	23	11
	Fusiform gyrus, BA37	L	>8	–39	–51	–20
		R	>8	38	–42	–23
	Inferior temporal sulcus, BA39	L	6.01	–40	–64	9
		R	5.85	46	–69	12
(Words–scrambled words) – (objects–scrambled objects)	Parahippocampal gyrus, BA30	L	5.15	–17	–36	4
		R	4.99	20	–38	4
	Cuneus, BA18	L	4.25	15	–84	14
		R	4.25	15	–84	14
	Fusiform gyrus, BA37	L	6.87	–47	–41	–18
		R	7.8	–28	–95	–4
Inferior occipital gyrus BA17/18	L	6.6	30	–96	–4	
	R	6.6	30	–96	–4	
Middle and superior temporal gyri, BA21/22	L	5.69	–68	–33	5	
	R	5.69	–68	–33	5	
Precentral gyrus, BA4	L	4.3	–60	–12	32	
	R	4.3	–60	–12	32	

i.e. higher activations than for scrambled words. However, word-derived 'gestalts' (Fig. 3, green bars) behaved just like regular objects (i.e. equal or smaller activations relative to scrambled controls), and



**Fig. 3.** Sensitivity to words in early visual occipital cortex. Plot of activations by words (orange, solid), scrambled words (orange, outline), objects (blue, solid) and scrambled objects (blue, outline) in early occipital regions of interest. Heightened activity relative to scrambled controls is seen for words only. Gestalt stimuli, which share grouping and low-level features with words, nevertheless show a profile similar to pictures. This activation pattern may be a consequence of perceptual learning driven by the pressure for fast, high spatial frequency, parallel processing of words. Error bars represent the SEM across subjects after subtraction of the individual subjects' mean. (\*\*\*)  $p < 0.001$ ; (\*\*)  $p < 0.01$ ; (\*)  $p < 0.05$ ; m.s. – marginally significant,  $p = 0.052$ ; n.s. – not significant,  $p > 0.1$ .

they again differed from words. This demonstrates that heightened responses to letters in early visual cortex are not due to low-level visual properties of words and their controls.

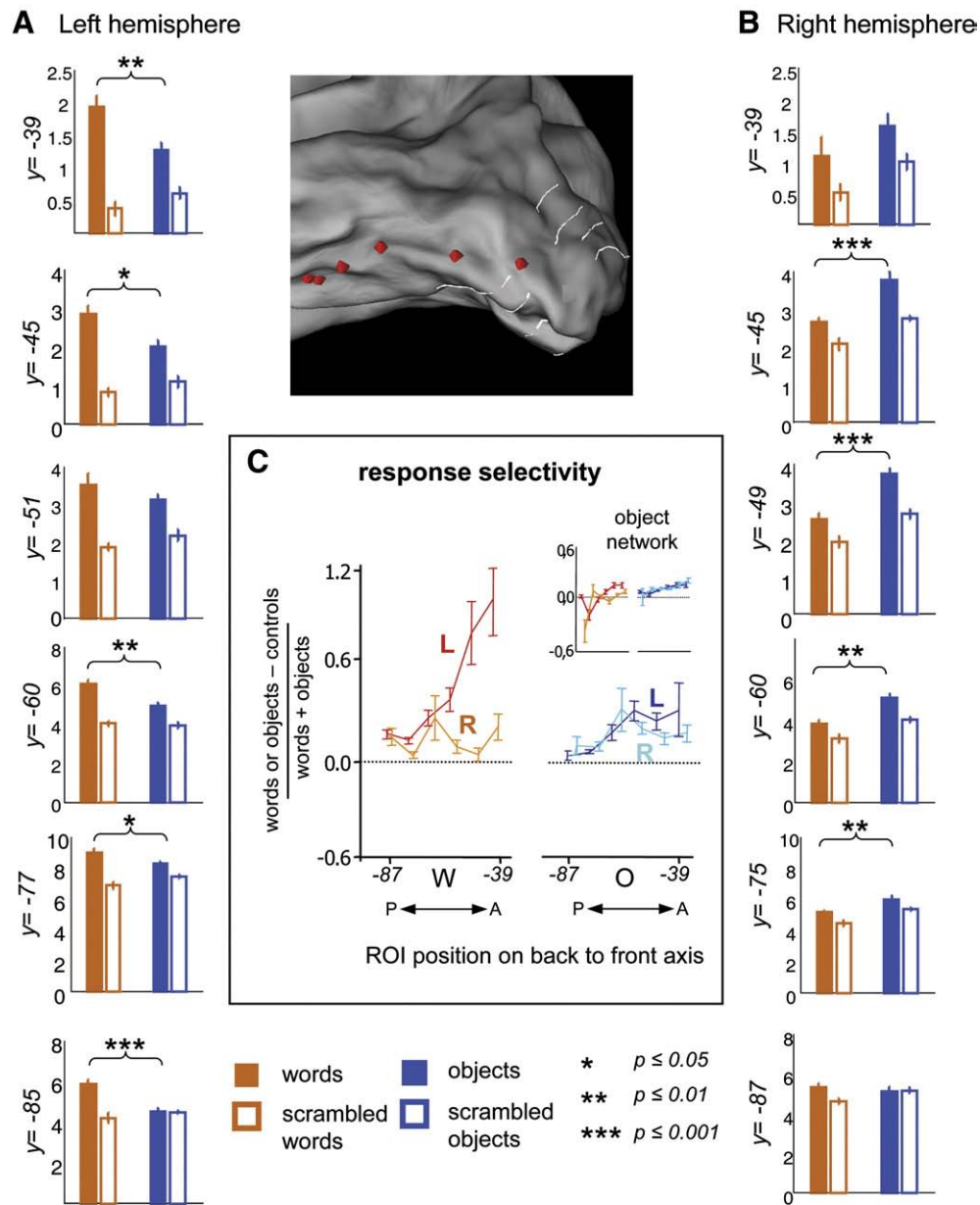
It may be argued that, although subjects performed the same task in word and object blocks, the early visual differences could be due to a greater attention paid to words, driven by top-down signals from the frontoparietal attentional network. Note however that behavioral performance was identical for words and for objects during the fMRI acquisition task (Fig. 1H). The same was true for the additional naming task performed in the scanner after the fMRI acquisition (Figs. 1G–H). Those results do not support the hypothesis of a differential deployment of attention to word and object blocks. Furthermore, we also tested directly whether words were associated with greater activation than objects in the frontoparietal network driving attentional control. No such difference was found, even at very low statistical thresholds (Fig. S5,  $p < 0.05$  voxel-wise uncorrected). Note that in the literature, whenever attentional amplification effects are present, they are typically much larger in frontoparietal regions than in visual areas which are the targets of such influence (Kastner et al., 1999). It is therefore unlikely that the early visual activations for words observed here are due to top-down attentional amplification from the dorsal attentional network. However, this does not rule out the possibility that early occipital activations to words are due to feedback from other areas, such as the VWFA, or the possibility that both early occipital and VWFA responses are due to feedback from high-level phonological/lexical areas (see Discussion).

#### Individual ROI analysis of word- and object-related activations

We then studied the functional properties of the bilateral ventral pathway in greater detail, using ROIs sampling the whole length of the ventral visual pathway. ROIs were defined in two steps. We first selected 6 major peak voxels located along the ventral stream based on the group level words minus checkerboards contrast in the functional localizer (Fig. S3A). Those peaks ranged from MNI  $y = -86$  to  $y = -40$  and are shown in Fig. 4 (center). The location of ROIs was optimized on an individual basis. Around the above mentioned peaks we defined, for each subject individually, an ROI centered on the voxel that showed maximum activation in the same contrasts for this particular subject (see Methods). This allowed us to take full advantage of the high-resolution fMRI data.



### Activation of occipito-temporal ROIs by words, objects, and control stimuli



**Fig. 4.** Activation of occipito-temporal ROIs by words, objects, and control stimuli. (A–B) Activations along the ventral reading network, in  $3 \times 3 \times 3$  voxel regions of interest determined on the basis of the individual subjects' functional localizer. The average peaks around which we searched for individual peak activations are shown in the top center render (red dots). ROIs are arranged from the most posterior (bottom) to the most anterior one (top). In the left hemisphere (A), areas more active for words than objects were found at most levels of the reading system. Error bars represent the SEM across subjects after subtraction of the individual subjects' mean. (C) We calculated an index of response selectivity defined as the difference between the response to words (respectively objects) and their corresponding controls, normalized for response strength. The selectivity of the response to words increased steadily from posterior to anterior regions in the left but not the right ventral occipitotemporal cortex. We repeated our selectivity analysis for the object-activated network (inset in panel C, ROI location shown in Fig. S6), and did not find patches of cortex with selectivity for objects comparable to that seen for words.

An ANOVA on the left-hemisphere ROIs (Fig. 4A) revealed higher activation relative to baseline for words than for objects in all ROIs (all  $p < 0.05$ ) except for the ROI at MNI  $y = -51$  ( $p = 0.39$ ). A difference was also observed using the better-controlled contrast for [words–scrambled words]–[objects–scrambled objects] (all  $p \leq 0.017$ , except for MNI  $y = -51$ ,  $p = 0.1$ ). It should be noted that this difference was always relative, and voxels showing maximum activation for words were also significantly activated by objects. Activation patterns were markedly different across hemispheres in all ROIs (Fig. 4B): Contrary to the left hemisphere, right-hemispheric

responses relative to baseline were either stronger to objects than to words (MNI  $y = -75$  to  $y = -45$ , all  $p \leq 0.008$ ) or comparable with both types of stimuli ( $y = -87$  and  $y = -39$ ,  $p > 0.31$ ).

We then plotted an index of response selectivity for words or objects over their respective controls, normalized for overall activation strength (Fig. 4C, see Methods for details) and corrected for baseline activity (Simmons et al., 2007). As the normalizing factor is the sum activation for words and objects, the selectivity index can be directly compared across the two stimulus classes. The selectivity index of the response to words increased steadily from the back to the

front of the left VOT ( $p < 0.001$  for the comparison across the 5 anterior ROI's). In contrast, in the right hemisphere the selectivity index was roughly constant across ROIs ( $p = 0.39$  for the comparison across the 5 anterior ROI's; significant interaction between ROI and hemispheres,  $p < 0.001$ ). For objects, the pattern was identical in both hemispheres.

ROIs described so far were selected on the basis of their response to words. We also defined a second set of ROIs, this time looking for areas activated by objects, on the basis of the objects–controls contrast in the localizer (see Supplemental result R2). The selectivity plots for those ROIs are shown in the inset within Fig. 4C. Briefly, we could not find patches of cortex with selectivity for objects comparable to the selectivity we found for words.

#### Role of vertex invariants in reading

When subjects see partially deleted objects and printed words, in which either vertices or line midsegments are preserved (Figs. 1D–F), they make fewer errors and are faster to respond to the vertex preserved version (Biederman, 1987; Szwed et al., 2009 and Figs. 1G–H). Using stimuli developed by Szwed et al., (2009) we were able to probe the neural mechanisms of this classical effect. We examined whether part of the visual cortex preferentially responds to vertices than to line midsegments and, if so, whether this preference could be related to the reproducible cortical localization of the VWFA.

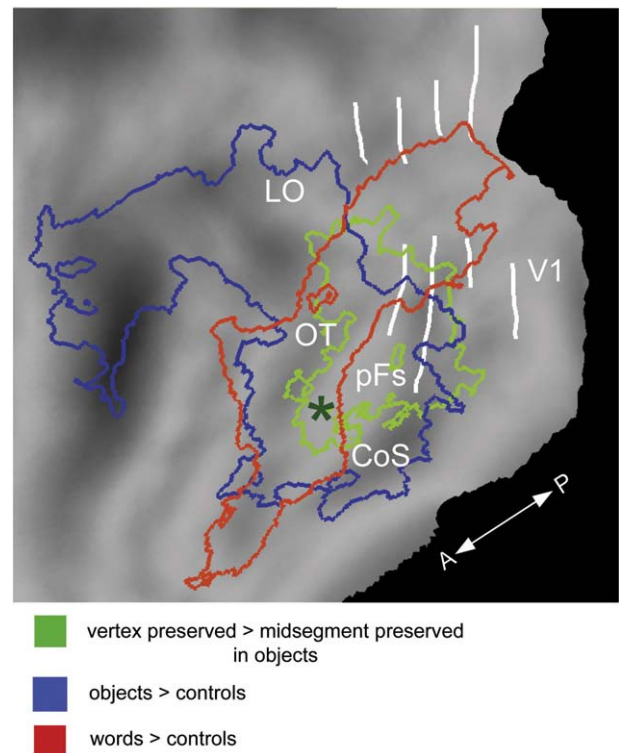
First, we investigated the effect of feature type (vertices vs. midsegments) in objects. We searched for brain regions that would follow the subjects' behavior (better recognition for objects with vertex preserved than with midsegment preserved) and would therefore respond more to the vertex-preserved than to the midsegment-preserved variants of objects. We found such a profile only in the bilateral fusiform regions (Fig. 5; left: MNI  $-26 -68 -4$ ,  $Z = 4.34$ ; right: MNI  $35 -71 -4$ ,  $Z = 4.26$ ). No regions were found for the opposite contrast (midsegment preserved minus vertex preserved). Thus the preference for vertices over midsegments (Fig. 5; green outline) was confined to a subpart of the object recognition system corresponding to the fusiform gyrus (Fig. 5; blue outline). Notably, the effect of feature type was absent from the object-responsive lateral occipital (LO) cortex, even at very low thresholds ( $p < 0.05$  voxelwise).

We then studied the overlap of this feature-sensitive region with word-responsive regions as defined by the words–scrambled words contrast of the main experiment (Fig. 5; red outline). In the mid part of the occipitotemporal cortex along the anteroposterior axis, there was a large overlap of the word-responsive regions with the vertex-sensitive sector of the object system: on the flattened cortex, 65% of vertex sensitive pixels also belonged to the reading system. This suggests that recognition mechanisms that rely on vertex invariants are mostly located in the parts of the object recognition system that are involved in reading (see Discussion).

We then examined the effect of feature type in words, including both the 55% and the 35% variants of words (the latter showed a behavioral advantage for the vertex preserved variant, see Fig. 1). The contrast of vertex preserved minus midsegment preserved variants of words showed activations slightly below the extent threshold within the Visual Word Form system (MNI  $-33 -45 -22$ ,  $Z = 3.79$ , 105 voxels, peak marked by a star in Fig. 5, area completely overlapping with the vertex-sensitive cortex observed with objects), and in the left STS (MNI  $-56 -50 2$ ,  $Z = 3.88$ , 120 voxels). This finding is consistent with consistent with greater activation in the context of more efficient recognition at the whole word level (Binder et al., 2006; Vinckier et al., 2007; Levy et al., 2008; Glezer et al., 2009).

In the set of ventral ROIs described in the previous section, the effect of feature type was in agreement with those SPM analyses (Supplemental Fig. S6): we found larger activations for vertex preserved than for midsegment preserved stimuli in left mid-fusiform

#### Brain regions sensitive to line junctions (vertices)



**Fig. 5.** Brain regions sensitive to line junctions (vertices). Outline of left occipitotemporal group activations, overlaid on the inflated cortical sheet: activations by objects minus scrambled objects (blue); by vertex preserved variants minus midsegment preserved variants of objects (green); and by words minus scrambled words (red). There was a large overlap of the word-responsive regions with the vertex sensitive sector of the object system, suggesting that recognition mechanisms that rely on vertex invariants are mostly located in the parts of the object recognition system that are involved in reading. The peak of the preference for vertex preserved variants of words over midsegment preserved variants is marked by a star. The average boundaries of areas V2, V3, V3v, and V4 are marked as white lines. LO: lateral occipital cortex; OT: occipito-temporal sulcus; pFs: posterior fusiform gyrus; CoS: colateral sulcus; A: anterior; P: posterior. Contrasts of words and objects minus controls were thresholded at  $p < 0.001$  voxel-wise, and  $p < 0.05$  cluster-wise corrected for multiple comparisons across the whole brain; the contrast of vertex preserved minus midsegment preserved variants was thresholded at  $p < 0.01$  voxel-wise,  $p < 0.05$  cluster-wise corrected for multiple comparisons across the whole brain. Supplementary Fig. S6 contains a corresponding ROI analysis.

ROIs for words and objects, and in symmetrical right-hemispheric ROIs for objects only.

Finally, we compared vertex preserved vs. midsegment preserved versions of scrambled objects and scrambled words, and found no activations, even at low statistical thresholds ( $p = 0.01$  voxel-wise). This last finding indicates that the putative representations sensitive to the presence of vertices can be observed at a relatively high level of integration which is not attained by scrambled stimuli.

#### Discussion

The main findings of this study can be summarized as follows. First, once stimuli are controlled for low-level visual features, and a design minimizing top-down influences is used, areas more active for words than objects can be found at several levels of the ventral visual system. Second, early visual areas V1/V2 and V3V/V4 show more activation to words than to scrambled words, but not to objects relative to scrambled objects. Third, a restricted part of the object perception system, limited to the fusiform gyrus, is sensitive to the

presence of viewpoint-invariant line junctions. This preference partially overlaps with the reading system. While this result might be a coincidence, it could also indicate a reason behind the highly reproducible cerebral localization of the Visual Word Form Area. We now discuss these points in turn.

#### *Cortical specialization for words and for objects and task effects*

Our results bear directly on the issue of whether a genuine cortical specialization for reading exists in the ventral visual system. Several imaging studies found fusiform areas activated more for words than objects (Hasson, 2002; Ben-Shachar, 2007; Baker et al., 2007). Other reports, however, contested this result (Price and Devlin, 2003; Wright et al., 2008; Kherif et al., 2010), while yet others suggested that these areas are activated more by written words than by objects only under certain circumstances (Starrfelt and Gerlach, 2007). The discrepancies between the above-cited studies may stem from the fact that while some of them matched words and objects in luminance (Baker et al., 2007), or shape (Ben-Shachar, 2007), in most studies no attempt was made to match these categories, and the objects always had more features than the words. In particular, no prior study controlled for the density, type and number of vertices and line junctions, which we suggest are essential features shared by words and line drawings of objects.

In this study we attempted to reduce as much as possible the low-level visual differences between words and objects, and then to factor out the remaining category-specific low-level confounds by subtracting the activations evoked by category-specific scrambled controls. Naturally, the scrambling manipulation itself changes both the amount of collinear contour a factor which is known to influence early visual responses (Altmann et al., 2003; Kourtzi et al., 2003; Dumoulin and Hess, 2006; Ostwald et al., 2008). However, for this subtraction method to be a valid control for comparing word vs. object activations, it is sufficient that the effect of scrambling be comparable for words and objects.

An important low-level visual difference between words and objects is that letters in words form separate entities, while objects appear as one global shape. While our object and word stimuli had the same amount of contour length, they differed in this parameter of internal grouping. Is it therefore possible that the partially specialized responses that we observed for words are not due to expertise for letter shapes per se, but to this particular grouping factor? Several observations argue against this possibility. First, our study included “gestalt” stimuli which had the same type of grouping by proximity and collinearity into separate distinct parts as words (Fig. 1D). Yet the responses to gestalts were substantially lower than responses to words, both in occipital and in fusiform areas (Supplementary Fig. S2; Fig. 3). Indeed, both the occipital and the fusiform regions showed very similar increases in response to words relative to the scrambled and to the gestalt control stimuli (Supplementary Fig. S2). Second, the fMRI responses to grouping by proximity are known to be widespread, particular intense in the lateral occipital cortex (LO), and, most importantly, identical in both hemispheres (Altmann et al., 2003; Ostwald et al., 2008). Yet, the responses to words that we observed were left-lateralized already at the level of V1/V2, and focused on the left lateral occipito-temporal sulcus, the classical location of the VWFA, rather than the LO. Thus, at this level at least, the responses to words we observed cannot be explained by the particular type of grouping into letters that is specific to words.

Completely equating all visual attributes is hardly possible considering the intrinsically different structure of words and common objects. Thus is still possible that the stronger responses to words than to objects might be due to some remaining differences between these stimuli and their respective controls, not eliminated by the present design. In the final analysis, conclusive evidence for reading expertise in early visual areas can be obtained only through using identical

stimuli while varying the expertise of the participants. Our team recently conducted such a study, by comparing fMRI activations to a variety of visual stimuli in literate vs. illiterate adults (Dehaene et al., 2010). In broad agreement with the present results, the acquisition of literacy led to increased activations to letter strings, first and foremost at the VWFA site, while the activations to other categories of stimuli was unchanged or even decreased. The end result was, therefore, a significantly increased difference between letter strings and other categories at the VWFA in literates compared to illiterates, congruent with the present findings. Furthermore, as in the present study, a positive effect of literacy was also found at the level of the occipital cortex (peak at MNI:  $-12, -88, 2$ , consistent with the occipital peak at MNI:  $-10 -96 0$  reported in the present study, see Table 1). Given these convergent findings, and because we also believe that in the present study low-level differences between words and objects were reasonably controlled for, we conclude that the stronger activations we found for words than objects most likely reflect partial specialization for reading at multiple levels in the visual cortex.

This conclusion does not exclude the possibility that these stronger activations for words than objects require top-down interactions with lexical/phonological areas, or that they are partially dependent on the task (see Table 2 in Cohen et al., 2004a,b; Starrfelt and Gerlach, 2007). In an experiment using an overt naming task with masked priming, Kherif et al. (2010) found that top-down influences generated by naming objects aloud influence activations to words, and vice-versa. Thus, even if words and objects stimuli activate different neuronal populations in the bottom-up processing phase, feedback responses that converge on both these neuronal populations from higher level areas can “mask” differences in activations to these two classes of stimuli. Our study, was designed so that the subjects did not have to name the objects aloud. We also used a rapid presentation mode which minimized top-down influences. This might have made it easier to reveal neuronal activations partially specialized for written words.

#### *How selective are the voxels responsive to words?*

Our results show that selectivity of fMRI responses was never absolute. Clusters maximally activated by words still showed significant responses to object stimuli (Fig. 4; see also similar result by Baker et al. (2007)). This finding is congruent with several neuroimaging studies (e.g. Ishai et al., 2000; Haxby et al., 2001). Note however, that Allison et al., (1999) using subdural electrodes, observed left occipitotemporal P150 or N200 waves that were occasionally exclusive for letter strings as compared to a variety of objects. Single-neuron studies will be needed to unequivocally determine whether there could be an absolute selectivity for words at a microscopic columnar or single-neuron level, presumably beyond the current resolution of fMRI (Dehaene and Cohen, 2007).

At the macroscopic resolution accessible to our fMRI method (1.5 mm voxels), our results (Fig. 4) suggest that cortical patches that respond maximally to letters and words still participate in the encoding of other stimulus classes. Congruent with this finding, a recent single-unit electrophysiological study of large numbers of neurons from monkey inferotemporal (IT) cortex, the homologue of human fusiform area, found a distributed representation of object identity, with weaker responses to non-preferred stimuli contributing significantly to encoding (Kiani et al., 2007). We propose therefore that when neurons develop expertise for letter shapes in the process of learning to read, they may still continue to participate in encoding of objects. Whether and to what extent the process of learning to read reduces or displaces object-related activations in the left inferotemporal area remains to be clarified, particularly by studying reading acquisition in children and illiterate adults. Some results suggest that perceptual expertise for one category of object (e.g. cars, written words) may cause small reductions in the cortical representation of

other categories (e.g. faces) (Gauthier et al., 2003; Dehaene et al., 2010). It is likely, however, that neural circuits for reading and for object recognition remain tightly intermingled even in expert readers, explaining that fMRI contrasts of words vs. pictures rarely reveal a strong form of functional specialization for words, unless the stimuli are very tightly matched.

#### *Early visual effects of perceptual expertise in reading*

By using better-controlled subtractions than previous studies, our research evidenced a greater response to written words than to line drawing not only in the VWFA, but also at much earlier levels of the visual system. We found that areas V1/V2 and V3V/V4 show more activity for words than for scrambled words (Figs. 2 and 3). Objects, on the other hand, evoke less activation than scrambled objects, which is consistent with several previous studies and the notion that object form is encoded in higher areas of the visual system, (e.g. Grill-Spector and Malach, 2004). Following Nazir (Nazir, 2000; Nazir et al., 2004; Nazir and Huckauf, 2006) we suggest that this enhanced responsivity of occipital cortex to words might be a neuronal correlate of the rapid and massively parallel processing of letters, and may be accounted for by perceptual learning mechanisms. Perceptual learning is a form of implicit learning that involves improvement in sensory discrimination by repeated exposure to sensory stimuli (Fahle and Poggio, 2004). It is possible that the early visual activations we observe may be a consequence of perceptual learning driven by the unique pressure for fast, high spatial frequency, and parallel processing of words, which placed a particularly strong constraint on early visual processing especially when reading fine print. Indeed, studies in nonhuman primates show that perceptual learning can lead to plastic changes in the visual cortex going as far back as V1 (Schoups et al., 2001; Li et al., 2008). In humans, following extensive training in detecting T shapes, Sigman et al. (2005) observed an increased level of activity in early visual cortex, together with a decrease of activation in higher visual cortex and the dorsal attention network. Arguably, this letter-detection experiment constitutes an analog of the perceptual training provided by reading experience, and further strengthens our conclusion that reading can lead to specific enhancements to words compared to scrambled words.

Why would early visual processing be more prominent for reading than for visual recognition of other categories such as objects? We believe that objects are much more complex and varied in their appearance than letters. They are typically seen at various locations and sizes in the visual field, and they are seen from many viewpoints, without the kind of intensive training with a small set of shapes that we experience when reading printed words, and which was found necessary for retinotopic expertise effects to arise (Sigman et al., 2005). Moreover, real objects generally do not present themselves solely as line combinations, but afford many other clues to recognition such as texture, color and 3D cues. In contrast, letters are simpler, come in a very limited set, are usually read at the same retinotopic location, always appear in the same orientation, and are exclusively defined on the basis of line patterns. Such factors are known to facilitate perceptual learning (Fahle and Poggio, 2004). Furthermore, in the highly educated subjects that we scanned, the necessity for speeded reading probably places a high emphasis on the optimization of a fast, highly parallel recognition of letter strings — and this is what skilled readers actually do, showing no effect of word length when words are approximately 3–8 letters long (New et al., 2006).

Our findings in occipital regions contrast in part with the studies by Vinckier et al. (2007) and Levy et al. (2008), which did not find significant differences between activations to false fonts and words in the most posterior sectors of occipital cortex, but only starting around  $y = -80$  (corresponding approximately to area V8). Further research will be necessary to determine whether the choice of task (one-back in the present study, vs. an easy oddball detection task in Vinckier et

al. (2007)) might explain this discrepancy. Note that the results were obtained with distinct stimuli and are thus not necessarily contradictory. On the contrary, they might simply suggest that the occipital effect of expertise that we describe here is not limited to alphabetic stimuli but, at this early visual stage, generalizes to any stimuli that are visually close enough to letter strings, including the false font stimuli used by Vinckier et al. (2007).

While the possibility of reading expertise in early visual areas is attractive, there always remains an alternative interpretation of our results: the occipital differences we observed could still be due to some remaining difference between words and their controls which was not eliminated in the present stimuli, but was better controlled by using false fonts in the Vinckier et al. (2007) study. Candidates include features such as combinations of parallel lines, quasi-periodic recurrence of individual elements, and global alignment of letter tops and bottoms, which were not preserved with our method of generating the gestalt stimuli. Given that such factors have not been studied extensively in other studies, we do not have precise knowledge of how they can affect the overall responsiveness of retinotopic areas. Further, such differences are extremely difficult to eliminate. Ultimately, as already noted, the proof of an effect of reading expertise in the low-level visual cortex must come from studies using identical stimuli while varying the expertise of the participants, for instance by comparing literate vs. illiterate adults (Carreiras et al., 2009; Dehaene et al., 2010), children at different stages of reading acquisition, or adults mastering different writing systems (Bolger et al. 2005; Baker et al., 2007).

#### *Relation to existing results and models of word recognition and interplay of bottom-up and recurrent mechanisms*

We believe that our hypothesis of early visual cortex involvement in reading can be reconciled with existing models of word recognition. The initial VWFA hypothesis proposed that the VWFA is responsible both for the speed of word form encoding, notably due to parallel letter recognition, and for its invariance for position or font (Cohen et al., 2002). Later propositions, notably the Local Combination Detector model, expanded this view by proposing that word recognition occurs at several levels of the visual system, in a hierarchical manner (for similar models see also Whitney, 2001; Dehaene et al., 2005; Grainger et al., 2008). Our current findings bring new evidence for this multi-level view of visual word form recognition by suggesting that the remarkable efficacy of reading partly originates earlier than the VWFA, in the occipital cortex. Under this view, areas V1/V2 and V3V/V4 would be recruited to detect and amplify visual features relevant for reading, providing highly parallel input to the VWFA, and greatly accelerating recognition. We argue that this recruitment of early visual areas is more prominent in reading than other visual cognitive processes such as object recognition. The VWFA region would then carry out orthographic analysis of whole word forms (Binder et al., 2006; Vinckier et al., 2007; Levy et al., 2008; Glezer et al., 2009).

In the present study, we showed that it is unlikely that the heightened responses to words in early visual areas are due to differential attentional amplification by the frontoparietal attentional network, because those regions are not more activated by words than by objects (Fig. S5). Nevertheless, our findings do not rule out the possibility that if the enhanced response to words in early visual regions are really due to expertise for reading, they could arise from feedback from other areas such as high-level visual areas (e.g. VWFA) or other downstream areas involved in linguistic and phonological processing. Indeed, other experiments on perceptual expertise for complex shapes show that the effects of perceptual learning disappear under anesthesia despite preserved bottom-up responsiveness (Li et al., 2008). It is thus possible that, in our case as well, early visual responses to letters could arise through a combination of bottom-up mechanisms and recurrent processing in loops with higher visual (e.g.

Williams et al., 2008) lexical and phonological (Cornelissen et al., 2009; Wheat et al., 2010) areas, and that they need a particular behavioral context to emerge (Harel et al., 2010). In the same manner, the responses in the VWFA most likely also result from a combination of bottom-up processing and recurrent interactions with lexical/phonological areas. Indeed, Szwed et al., (2009) in a study using the same stimuli reported that naming times were shorter for words than objects. This result, together with stronger activations for words than for objects in lexical/phonological areas (Fig. 2, Table 1) suggests that words were more automatically associated with language processing. Consequently, stronger recurrent interactions between the VWFA and lexical/phonological areas could have contributed to stronger responses to words than objects we observed in this experiment. In fact, recent reports of very early (~150 ms post-stimulus) activations of areas as “downstream” in language processing as the Middle Frontal Gyrus (Cornelissen et al., 2009; Wheat et al., 2010) demonstrate that the time window for such recurrent interactions is larger than previously thought. The exact contributions of bottom-up and recurrent processes in visual word recognition will have to be addressed by techniques with greater temporal and spatial precision than fMRI, such as intracranial recordings.

#### *Putative origin of the VWFA's localization*

The previous sections of the discussion dealt with visual word form recognition in mature, expert readers. In this last section, we discuss the putative developmental origin of the VWFA, addressing the puzzling issue of its consistent localization at the same cortical site across scripts, subjects and cultures (Cohen et al., 2000; Bolger et al., 2005; Baker et al., 2007; Qiao et al., 2010). The object recognition system encompasses wide sectors of the fusiform and lateral occipital areas (Kanwisher et al., 1997; Grill-Spector and Malach, 2004). It is therefore surprising that the reading system develops at a precise location within the former of these two regions, in the lateral occipitotemporal sulcus bordering the fusiform gyrus. We previously hypothesized that, even prior to reading acquisition, this region may already have an initial preference for intermediate features consisting of junctions of intersecting lines that play a key role in object recognition (Dehaene, 2005; Dehaene, 2009). Indeed, previous behavioral results (Biederman, 1987; Lanthier et al., 2009; Szwed et al., 2009), which were replicated here, showed that reading, similar to object recognition, relies heavily on such line junctions (vertices). Note that the importance of vertices as stimuli for the ventral visual system has been well established with electrophysiological methods in primates (Tsunoda et al., 2001; Kayaert et al., 2003; Brincat and Connor, 2004).

To our knowledge, we provide here the first demonstration of a neural correlate for the behavioral preference for vertices first described by Biederman (1987), and we derive from this finding a novel hypothesis concerning the localization of the VWFA. Still, some caution is advised concerning the activation difference which we observed between objects reduced to vertices or to midsegments. Indeed, at a behavioral level, objects with vertices present are better recognized than objects where these features were deleted (see Fig. 1B and Biederman, 1987; Szwed et al., 2009). Hence, instead of reflecting bottom-up selectivity for vertices vs. midsegment features, the activity difference in the fusiform region might be related to recognition performance. This would also be consistent with the fact that we did not observe any activity differences between vertices and midsegments within the scrambled stimuli. While real, this possibility is however mitigated by prior research demonstrating that both fusiform and lateral occipital areas showed correlations of activation level with object recognition performance (Grill-Spector et al., 2000). By contrast, in the present study only the fusiform area, but not the lateral occipital area, activated more for objects with preserved vertices than for objects where these features were deleted. One may thus argue that the effect we observed cannot be reduced to a simple

difference in performance level, since it displays a greater regional specificity than the correlation with performance previously reported by Grill-Spector et al. (2000).

Most interestingly, the area that responds more vigorously to vertex-preserved versions of the objects partially overlaps with the reading system (Fig. 5). While this result might be a coincidence, this response bias towards line junctions could also indicate one reason behind the remarkably reproducible localization of the VWFA. This circumstantial evidence might also provide a clue as to why the VWFA is activated not only by the subjects' native alphabet but also by unknown alphabet characters (e.g. Xue and Poldrack, 2007). Under this view, the activation of the VWFA by unknown scripts would be explained by the fact that all the world's writing systems contain the same set of basic line-junction features (Changizi et al., 2006). Changizi (2006) found that vertices such as L or T shapes obey a universal distribution across writing systems, a distribution which is shared with the patterns of contour intersections that arise in natural scenes. Thus, the shape of written characters may have been culturally selected to match the pre-existing organization of our visual system (Dehaene, 2005; Changizi et al., 2006; Dehaene, 2009).

Neural network models have described the emergence of specialized neural structures for letters and numbers (e.g. Polk and Farah, 1998; Price and Devlin, 2003), but these simulations rely on “spontaneous symmetry breaking”, i.e. the emergence of cortical specialization at random sites from an amplification of initial noise. Therefore, they cannot explain, without additional hypotheses, why reading expertise systematically draws upon the same cortical location in all subjects and cultures. Our hypothesis of a pre-existing bias for line junctions supplements these empiricist models and, together with other constraints, may begin to account for the remarkably reproducible localization of the VWFA. These other constraints may include a bias of the lateral fusiform cortex for foveal rather than peripheral stimuli (Hasson et al. 2002); a posterior-to-anterior increase in perceptual invariance (Grill-Spector et al., 1998; Rolls 2000; Lerner et al. 2001); and more direct projection fibers of the left lateral fusiform gyrus to left-lateralized language areas (Cohen et al., 2000; Epelbaum et al., 2008; Pinel and Dehaene, 2009). Together, these biases might explain why neural structures related to reading reproducibly begin to develop at this particular location. After several years of schooling, this development would finally lead to the neural hallmarks of reading expertise: a posterior-to-anterior gradient of sensitivity to whole word form (Binder et al., 2006; Vinckier et al., 2007; Glezer et al., 2009) and increased interactions with higher-order lexical/phonological areas (for a recent review see: Price, 2010).

Our hypothesis might be extrapolated to other domains. For example, the fact that visual expertise for categories of stimuli such as faces, body parts, places, or even cars and birds, engages distinct areas of the visual brain (Tarr and Gauthier, 2000; Bukach et al., 2006; Wong et al., 2009) could also be explained by the congruence between the stimuli, the kind of experience we have when learning about them (Gauthier et al., 2010; Op de Beeck and Baker, 2010), and the pre-existing local biases in the micro-architecture and connections of the visual cortex.

Finally, it is important to note that the above-mentioned influences are only biases. Efficient reading can still develop when the VWFA per se is missing. For instance, following early removal of the left occipitotemporal cortex, normal reading can develop on the basis of the right-hemispheric symmetrical region (Cohen et al., 2004a,b).

Supplementary materials related to this article can be found online at doi:10.1016/j.neuroimage.2011.01.073.

#### **Authors' contributions**

M.S., S.D. and L.C. designed the experiments. M.S. programmed and performed the experiments, analyzed the results and wrote the paper

with contributions and supervision from L.C. and S.D. In addition, A.K. and E.E. contributed to experimental design and analysis, R.V. contributed to analysis and A.A. optimized the neuroimaging acquisition sequences.

## Acknowledgments

We thank Pierre Lebranchu, Guy Orban and Alain Berthoz for sharing their retinotopic area localizer, Michel Vidal-Naquet, Barak Blumefeld, Yulia Lerner, Fabien Vinckier, Christophe Pallier, Jean Lorenceau and Stephane Lehericy for helpful comments, and the entire CENIR team for technical assistance. This fMRI experiment was part of a general research program on functional neuroimaging of the human brain which was sponsored by the Atomic Energy Commission (Denis Le Bihan). The study was supported by INSERM, a grant (CORELEX) from the French Agence Nationale de la Recherche (ANR) and a grant from the Federative Institute of Research on Functional Neuroimaging (IFR 49). M.S. was supported by an International Human Frontier Science Program Organisation Long-Term fellowship.

## References

- Allison, T., Puce, A., Spencer, D.D., McCarthy, G., 1999. Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb Cortex* 9 (5), 415–430.
- Altmann, C.F., Bulthoff, H.H., Kourtzi, Z., 2003. Perceptual organization of local elements into global shapes in the human visual cortex. *Curr Biol* 13 (4), 342–349.
- Baker, C.I., Liu, J., Wald, L.L., Kwong, K.K., Benner, T., Kanwisher, N., 2007. Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proc Natl Acad Sci USA* 104 (21), 9087–9092.
- Ben-Shachar, M., Dougherty, R.F., Deutsch, G.K., Wandell, B.A., 2007. Differential sensitivity to words and shapes in ventral occipito-temporal cortex. *Cereb Cortex* 17, 1604–1611.
- Biederman, I., 1987. Recognition-by-components: a theory of human image understanding. *Psychol Rev* 94, 115–147.
- Binder, J.R., Medler, D.A., Westbury, C.F., Liebenthal, E., Buchanan, L., 2006. Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage* 33 (2), 739–748.
- Bolger, D.J., Perfetti, C.A., Schneider, W., 2005. Cross-cultural effect on the brain revisited: universal structures plus writing system variation. *Hum Brain Mapp* 25, 92–104.
- Brincat, S.L., Connor, C.E., 2004. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat Neurosci* 7 (8), 880–886.
- Bukach, C.M., Gauthier, I., Tarr, M.J., 2006. Beyond faces and modularity: the power of an expertise framework. *Trends Cogn Sci* 10 (4), 159–166.
- Carreiras, M., Seghier, M.L., Baquero, S., Estevez, A., Lozano, A., Devlin, J.T., Price, C.J., 2009. An anatomical signature for literacy. *Nature* 461, 983–985.
- Changizi, M.A., Zhang, Q., Ye, H., Shimojo, S., 2006. The structures of letters and symbols throughout human history are selected to match those found in objects in natural scenes. *Am Nat* 167, e117–e139.
- Claeys, K.G., Dupont, P., Cornette, L., Snaert, S., Van Hecke, P., De Schutter, E., Orban, G.A., 2004. Color discrimination involves ventral and dorsal stream visual areas. *Cereb Cortex* 14 (7), 803–822.
- Cohen, L., Dehaene, S., 2004. Specialization within the ventral stream: the case for the Visual Word Form Area. *Neuroimage* 22, 466–476.
- Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.A., Michel, F., 2000. The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* 123, 291–307.
- Cohen, L., Jobert, A., Le Bihan, D., Dehaene, S., 2004a. Distinct unimodal and multimodal regions for word processing in the left temporal cortex. *Neuroimage* 23 (4), 1256–1270.
- Cohen, L., Lehericy, S., Chochon, F., Lemer, C., Rivaud, S., Dehaene, S., 2002. Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain* 125 (Pt 5), 1054–1069.
- Cohen, L., Lehericy, S., Henry, C., Bourgeois, M., Larroque, C., Sainte-Rose, C., Dehaene, S., Hertz-Pannier, L., 2004b. Learning to read without a left occipital lobe: a single case of right-hemispheric shift of the Visual Word Form Area. *Ann Neurol* 56, 890–894.
- Cornelissen, P.L., Kringelbach, M.L., Ellis, A.W., Whitney, C., Holliday, I.E., Hansen, P.C., 2009. Activation of the left inferior frontal gyrus in the first 200 ms of reading: evidence from magnetoencephalography (MEG). *PLoS ONE* 4 (4), e5359.
- Dehaene, S., 2005. Evolution of human cortical circuits for reading and arithmetic: the “neuronal recycling” hypothesis. In: Dehaene, S., Duhamel, J.R., Hauser, M., Rizzolatti, G. (Eds.), *From Monkey Brain to Human Brain*. MIT Press, Cambridge, MA, pp. 133–157.
- Dehaene, S., 2009. *Reading in the Brain*. Penguin Viking, New York.
- Dehaene, S., Cohen, L., 2007. Cultural recycling of cortical maps. *Neuron* 56 (2), 384–398.
- Dehaene, S., Cohen, L., Sigman, M., Vinckier, F., 2005. The neural code for written words: a proposal. *Trends Cogn Sci* 9, 335–341.
- Dehaene, S., Naccache, L., Cohen, L., Bihan, D.L., Mangin, J.F., Poline, J.B., Riviere, D., 2001. Cerebral mechanisms of word masking and unconscious repetition priming. *Nat Neurosci* 4 (7), 752–758.
- Dehaene, S., Pegado, F., Braga, L.W., Ventura, P., Filho, G.N., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., Cohen, L., 2010. Impact of literacy on the cortical networks for vision and language. *Science* 3, 1359–1364.
- Dumoulin, S.O., Hess, R.F., 2006. Modulation of V1 activity by shape: image-statistics or shape-based perception? *J Neurophysiol* 95 (6), 3654–3664.
- Epelbaum, S., Pinel, P., Gaillard, R., Delmaire, C., Perrin, M., Dupont, S., Dehaene, S., Cohen, L., 2008. Pure alexia as a disconnection syndrome: new diffusion imaging evidence for an old concept. *Cortex* 44 (8), 962–974.
- Fahle, M., Poggio, T., 2004. *Perceptual Learning*. MIT Press, Cambridge.
- Gaillard, R., Naccache, L., Pinel, P., Clemenceau, S., Volle, E., Hasboun, D., Dupont, S., Baulac, M., Dehaene, S., Adam, C., Cohen, L., 2006. Direct intracranial fMRI and lesion evidence for the causal role of left inferotemporal cortex in reading. *Neuron* 50, 191–204.
- Gauthier, I., Curran, T., Curby, K.M., Collins, D., 2003. Perceptual interference supports a non-modular account of face processing. *Nat Neurosci* 6 (4), 428–432.
- Gauthier, I., Wong, A.C., Palmeri, T.J., 2010. Manipulating visual experience: comment on Op de Beek and Baker. *Trends Cogn Sci* 14, 235–236.
- Gilbert, C.D., Li, W., Piech, V., 2009. Perceptual learning and adult cortical plasticity. *J Physiol* 587 (Pt 12), 2743–2751.
- Glezer, L.S., Jiang, X., Riesenhuber, M., 2009. Evidence for highly selective neuronal tuning to whole words in the “Visual Word Form Area”. *Neuron* 62, 192–204.
- Grainger, J., Rey, A., Dufau, S., 2008. Letter perception: from pixels to pandemonium. *Trends Cogn Sci* 12, 381–387.
- Grainger, J., Tygdat, I., Iselle, J., 2010. Crowding affects letters and symbols differently. *J Exp Psychol Hum Percept Perform* 36 (3), 673–688.
- Grill-Spector, K., Kushnir, T., Hendler, T., Malach, R., 2000. The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci* 3 (8), 837–843.
- 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 (4), 316–328.
- Grill-Spector, K., Malach, R., 2004. The human visual cortex. *Annu Rev Neurosci* 27, 649–677.
- Harel, A., Gilaie-Dotan, S., Malach, R., Bentin, S., 2010. Top-down engagement modulates the neural expressions of visual expertise. *Cereb Cortex* 20, 2304–2318.
- Hasson, U., Levy, I., Behrmann, M., Hendler, T., Malach, R., 2002. Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* 34 (3), 479–490.
- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293 (5539), 2425–2430.
- Ishai, A., Ungerleider, L.G., Martin, A., Haxby, J.V., 2000. The representation of objects in the human occipital and temporal cortex. *J Cogn Neurosci* 12 (Suppl 2), 35–51.
- 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.
- Kastner, S., Pinsk, M.A., De Weerd, P., Desimone, R., Ungerleider, L.G., 1999. Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron* 22 (4), 751–761.
- Kayaert, G., Biederman, I., Vogels, R., 2003. Shape tuning in macaque inferior temporal cortex. *J Neurosci* 23 (7), 3016–3027.
- Kherif, F., Josse, G., Price, C.J., 2010. Automatic top-down processing explains common left occipito-temporal responses to visual words and objects. *Cereb Cortex*.
- Kiani, R., Esteky, H., Mirpour, K., Tanaka, K., 2007. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J Neurophysiol* 97 (6), 4296–4309.
- Kourtzi, Z., Tolias, A.S., Altmann, C.F., Augath, M., Logothetis, N.K., 2003. Integration of local features into global shapes: monkey and human fMRI studies. *Neuron* 37 (2), 333–346.
- Lanthier, S., Risko, E.F., Stolz, J.A., Besner, D., 2009. Not all visual features are created equal: early processing in letter and word recognition. *Psychon Bull Rev* 16, 67–73.
- 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 (4), 287–297.
- Levy, J., Pernet, C., Treserras, S., Boulouar, K., Berry, I., Aubry, F., Demoullins, T., Celsis, P., 2008. Piecemeal recruitment of left-lateralized brain areas during reading: a spatio-functional account. *Neuroimage* 43, 581–591.
- Li, W., Piech, V., Gilbert, C.D., 2008. Learning to link visual contours. *Neuron* 57 (3), 442–451.
- McKone, E., Kanwisher, N., Duchaine, B.C., 2007. Can generic expertise explain special processing for faces? *Trends Cogn Sci* 11 (1), 8–15.
- Nazir, T.A., 2000. Traces of print along the visual pathway. In: Kennedy, A., Radach, R., Heller, D., Pynte, J. (Eds.), *Reading as a Perceptual Process*. Elsevier, Amsterdam, pp. 3–22.
- Nazir, T.A., Ben-Boutayab, N., Decoppet, N., Deutsch, A., Frost, R., 2004. Reading habits, perceptual learning, and recognition of printed words. *Brain Lang* 88 (3), 294–311.

- Nazir, T.A., Huckauf, A., 2006. The visual skill reading. In: Grigorenko, Naples (Eds.), *Single Word Reading: Cognitive, Behavioral and Biological Perspectives*. Lawrence Erlbaum.
- New, B., Ferrand, L., Pallier, C., Brysbaert, M., 2006. Reexamining the word length effect in visual word recognition: new evidence from the English Lexicon Project. *Psychon Bull Rev* 13 (1), 45–52.
- Op de Beeck, H.P., Baker, C.I., 2010. Informativeness and learning: response to Gauthier and colleagues. *Trends Cogn Sci* 14 (236–237).
- Ostwald, D., Lam, J.M., Li, S., Kourtzi, Z., 2008. Neural coding of global form in the human visual cortex. *J Neurophysiol* 99 (5), 2456–2469.
- Pinel, P., Dehaene, S., 2009. Beyond hemispheric dominance: brain regions underlying the joint lateralization of language and arithmetic to the left hemisphere. *J Cogn Neurosci*.
- Polk, T.A., Farah, M.J., 1998. The neural development and organization of letter recognition: evidence from functional neuroimaging, computational modeling, and behavioral studies. *Proc Natl Acad Sci USA* 95 (3), 847–852.
- Price, C.J., 2010. The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann NY Acad Sci* 1191 (1), 62–88.
- Price, C.J., Devlin, J.T., 2003. The myth of the visual word form area. *Neuroimage* 19 (3), 473–481.
- Qiao, E., Vinckier, F., Szwed, M., Naccache, L., Valabregue, R., Dehaene, S., Cohen, L., 2010. Unconsciously deciphering handwriting: subliminal invariance for handwritten words in the visual word form area. *Neuroimage* 49 (2), 1786–1799.
- Rolls, E.T., 2000. Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron* 27 (2), 205–218.
- Schoups, A., Vogels, R., Qian, N., Orban, G., 2001. Practising orientation identification improves orientation coding in V1 neurons. *Nature* 412 (6846), 549–553.
- Sigman, M., Pan, H., Yang, Y., Stern, E., Silbersweig, D., Gilbert, C.D., 2005. Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron* 46 (5), 823–835.
- Simmons, W.K., Bellgowan, P.S., Martin, A., 2007. Measuring selectivity in fMRI data. *Nat Neurosci* 10 (1), 4–5.
- Snodgrass, J.G., Vanderwart, M., 1980. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J Exp Psychol Hum Learn Mem* 6, 174–215.
- Starrfelt, R., Gerlach, C., 2007. The Visual What For Area: words and pictures in the left fusiform gyrus. *Neuroimage* 35 (1), 334–342.
- Szwed, M., Cohen, L., Qiao, E., Dehaene, S., 2009. The role of invariant features in object and visual word recognition. *Vis Res* 49, 718–725.
- Tarr, M.J., Gauthier, I., 2000. FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat Neurosci* 3 (8), 764–769.
- Tsunoda, K., Yamane, Y., Nishizaki, M., Tanifuji, M., 2001. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat Neurosci* 4, 832–838.
- Tygdad, I., Grainger, J., 2009. Serial position effects in the identification of letters, digits, and symbols. *J Exp Psychol Hum Percept Perform* 35, 480–498.
- Van Essen, D.C., 2005. A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *Neuroimage* 28 (3), 635–662.
- Vinckier, F., Dehaene, S., Jobert, A., Dubus, J.P., Sigman, M., Cohen, L., 2007. Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word form system. *Neuron* 55, 143–156.
- Wheat, K.L., Cornelissen, P.L., Frost, S.J., Hansen, P.C., 2010. During visual word recognition, phonology is accessed within 100 ms and may be mediated by a speech production code: evidence from magnetoencephalography. *J Neurosci* 30 (15), 5229–5233.
- Whitney, C., 2001. How the brain encodes the order of letters in a printed word: the SERIOL model and selective literature review. *Psychon Bull Rev* 8 (2), 221–243.
- Williams, M.A., Baker, C.I., Op de Beeck, H.P., Shim, W.M., Dang, S., Triantafyllou, C., Kanwisher, N., 2008. Feedback of visual object information to foveal retinotopic cortex. *Nat Neurosci* 11 (12), 1439–1445.
- Wong, A.C., Palmeri, T.J., Rogers, B.P., Gore, J.C., Gauthier, I., 2009. Beyond shape: how you learn about objects affects how they are represented in visual cortex. *PLoS ONE* 4 (12), e8405.
- Wright, N.D., Mechelli, A., Noppeney, U., Veltman, D.J., Rombouts, S.A., Glensman, J., Haynes, J.D., Price, C.J., 2008. Selective activation around the left occipito-temporal sulcus for words relative to pictures: individual variability or false positives? *Hum Brain Mapp* 29 (8), 986–1000.
- Xue, G., Poldrack, R.A., 2007. The neural substrates of visual perceptual learning of words: implications for the visual word form area hypothesis. *J Cogn Neurosci* 19 (10), 1643–1655.