Engaging in word recognition elicits highly specific modulations in visual cortex

Alex L. White[1*], Kendrick Kay[2], Kenny Tang[3], and Jason D. Yeatman[3];

[1] Department of Neuroscience & Behavior, Barnard College, Columbia University

[2] Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota

[3] Graduate School of Education, Division of Developmental-Behavioral Pediatrics, Stanford University

*Corresponding author: Alex L. White, alwhite@barnard.edu

Competing Interest Statement: The authors have no competing interests to declare.

Keywords: word recognition; visual word form area; object recognition; visual attention;

ABSTRACT

The ability to read depends on a region in ventral occipito-temporal cortex known as the "visual word form area" (VWFA). The VWFA, which has several sub-regions, lies amidst a collection of areas involved in visual recognition. Although it responds best to written words, its selectivity is not absolute, and it exhibits top-down modulations that are not well understood. Here, we used fMRI to investigate the interaction of bottom-up visual factors and top-down cognitive factors in the VWFA and neighboring regions. We presented participants with strings of letters and non-letter shapes at a range of visual field locations. For each stimulus type, participants performed a task in which the stimuli were task-relevant (lexical decision and gap localization, respectively), and a task in which the stimuli were irrelevant (detecting fixation dot color changes). Standard models of attention predict that all stimuli would evoke larger responses when task-relevant than irrelevant, throughout visual cortex. To the contrary, the data showed surprising patterns specific to the VWFA. Letter strings did evoke much larger responses when they were task-relevant than irrelevant, even when presented too far in the periphery to be recognized. In contrast, non-letter shapes evoked *smaller* responses when they were task-relevant. Connectivity analyses suggest that these task effects are due to flexible communication between the VWFA and Broca's area. We conclude that top-down modulations in visual cortex do not merely enhance representations of attended stimuli, but can boost processing in specific brain regions, contingent on engagement in specific tasks.

SIGNIFICANCE

A person's cognitive state determines how their brain responds to visual stimuli. The most common such effect is a response enhancement when stimuli are attended rather than ignored. We report a surprising twist on such attention effects, focusing on a brain region that selectively processes written words. There, the enhancement of responses to attended stimuli only occurred for letter strings, whereas responses to visually similar shapes were *suppressed* when attended. This selective task effect was accompanied by correlated activity in higher-level language regions, which appear to send excitatory feedback into particular visual regions only when the observer is trying to read. That feedback enables the discrimination of real and nonsense words, and is distinct from generic effects of visual attention.

INTRODUCTION

Visual cortex is capable of processing a wide variety of stimuli for any number of behavioral tasks. This raises a question: how exactly does the specific information required at any given moment get selected and used to execute a task? One key finding is that visual cortex does not perform a static stimulus-response mapping. Rather, the organism's goals influence ongoing activity (1, 2). The most prominent forms of top-down influence relate to attention: stimuli that are relevant to the current task evoke stronger responses than stimuli that are irrelevant, due to selection on the basis of visual field location or non-spatial features (3–5). However, we hypothesize that the brain performs more than simple amplification of relevant stimuli, because different tasks require particular visual information to be routed to different brain networks. For example, reading a word, judging the expression of a face, and catching a frisbee are three tasks that rely on visual input but engage different cortical networks.

The focus of this study is word recognition, an important visual task that engages a specific network beyond visual cortex (6, 7). The "visual word form area" (VWFA) in left ventral occipito-temporal cortex is key to the transformation of retinal input into the orthographic and lexical information that is conveyed to language regions during reading (8–11). Neighboring regions are specialized for recognizing faces, bodies, objects and scenes (12).

There are two ways of conceptualizing the VWFA, which are not mutually exclusive: first, it could be essentially a visual area, like its neighbors, with intrinsic selectivity for stimulus features in certain parts of visual space. Second, the VWFA could be unique due to its connection to the brain's language network, which regulates its activity and infuses it with lexical information via top-down signals.

Some evidence favors the first view: like other visual areas, the VWFA is sensitive to visual stimulus properties and modulated by spatial attention. It responds most strongly to strings of letters, although it does respond in a graded fashion to other categories of images (13–15). Importantly, the VWFA has several subregions that increase in selectivity along the posterior-to-anterior axis (11, 16–19). It also is sensitive to the visual field position (20), responding most strongly to words in the fovea and right parafovea (21, 22). Similar to effects of attention observed elsewhere in visual cortex (3), words evoke stronger responses in the VWFA when they are attended than ignored (14, 18).

The VWFA's function extends beyond the purely visual, however, as demonstrated by activity related to linguistic processing. There is evidence that it represents whole words as distinct identities (23–25), and that it responds differentially to frequent words, infrequent words, and novel pseudowords (e.g., 11, 27). Importantly however, the VWFA's sensitivity to higher-level linguistic features is stronger when the participant's task requires judging those features (11, 15, 27). Even more dramatically, the activation of the VWFA in some tasks does not require visual stimulation at all, for instance during Braille reading (28–30), or certain auditory judgments (36–40).

Altogether, the variety of extra-visual cognitive effects have led some researchers to the second view: that the VWFA plays a special role in word recognition only as a result of feedback from the language network, which uses the VWFA to link visual and language information (31). Indeed, the VWFA has white matter connections to spoken language regions, as well as regions associated with the control of attention (32–35). There is also strong resting state functional connectivity between the VWFA and those regions (33, 37–40; see also 41). The precise effects of such connectivity on the VWFA's activity, however, remain unknown.

The goal of this study is to clarify the role of the VWFA by measuring its stimulus selectivity and functional connectivity under varying task demands. We asked: what is the precise nature of the interaction between bottom-up stimulus features and top-down modulation? Does "visual attention" boost any task-relevant sensory signal in the VWFA, or are top-down enhancements contingent on engagement in a linguistic task? Are such effects present in other visual areas? To answer these questions, we recorded fMRI activity while participants viewed words and non-letter shapes and performed three different tasks. We recorded responses when each stimulus type was task-relevant and attended, and when it was task-irrelevant and ignored.

On each trial, one stimulus appeared at a random position along the horizontal meridian (**Figure 1A**). Each stimulus was either a string of four letters (forming either a real word or a pronounceable pseudoword), or a string of four squares and circles matched in size to the letter strings. In half the scans, participants attended to those stimuli. When they saw letter strings, they performed the *lexical decision task*: to report whether the stimulus was a real word or a pseudoword. When they saw shape strings, they performed the *gap localization task*: to report whether a gap in one of the inner two shapes was on the top side or bottom side. In the other half of scans, participants

ignored the stimuli and performed the *fixation color detection task*: to report whether or not the fixation dot turned slightly red. The visibility of the red color was controlled by a staircase to keep the fixation task sufficiently challenging.

We designed these tasks to strictly control attention. The fixation dot changed color simultaneously with the appearance of the letters or shapes, which lasted for only 150 ms. The participant did not have time to switch attention from one to the other, or to make a saccade. We also localized several subregions of word- and face-selective cortex (as well as earlier retinotopic areas), in each participant's cortical surface, rather than relying on average templates that obscure fine-grained patterns.

RESULTS

Task performance

We first assessed how stimulus discrimination accuracy (in units of *d')* varied as a function of stimulus position (**Figure 1B),** and across the three tasks. In both the gap task and the lexical task, accuracy was near perfect (>95% correct) for stimuli at fixation (0º), but dropped off quickly with increasing eccentricity, approaching chance by ±9º (<55% correct). Averaged across stimulus positions, the lexical, gap and fixation tasks were similarly difficult: mean (± SEM) d' = 1.66 ± 0.08, 1.38 ± 0.11, and 1.72 ± 0.15, respectively. d' in the gap task was slightly but significantly lower than in the lexical task (t(14)=2.52, P=0.024, BF=2.71), and than in the fixation task (t(14)=2.51, P=0.025; BF=2.64). The lexical and fixation tasks did not differ (t(14)=0.33, P=0.75; BF=0.28).

A surprising interaction of stimulus and task effects in text-selective areas

The primary goal of this study was to evaluate how the BOLD response to each type of stimulus depended on the participant's task. We used independent localizer scans (**Figure 1C**) to define regions of interest (ROIs) on each participant's cortical surfaces (**Figure 1D**). We focus on two text-selective regions in left occipito-temporal sulcus, VWFA-1 and VWFA-2 (18). Also of interest are two face-selective regions, FFA-1 and FFA-2, medial to the text-selective regions in both hemispheres, as well as posterior retinotopic visual areas (V1-hV4, VO1/2 and LO1/2). See Methods for details.

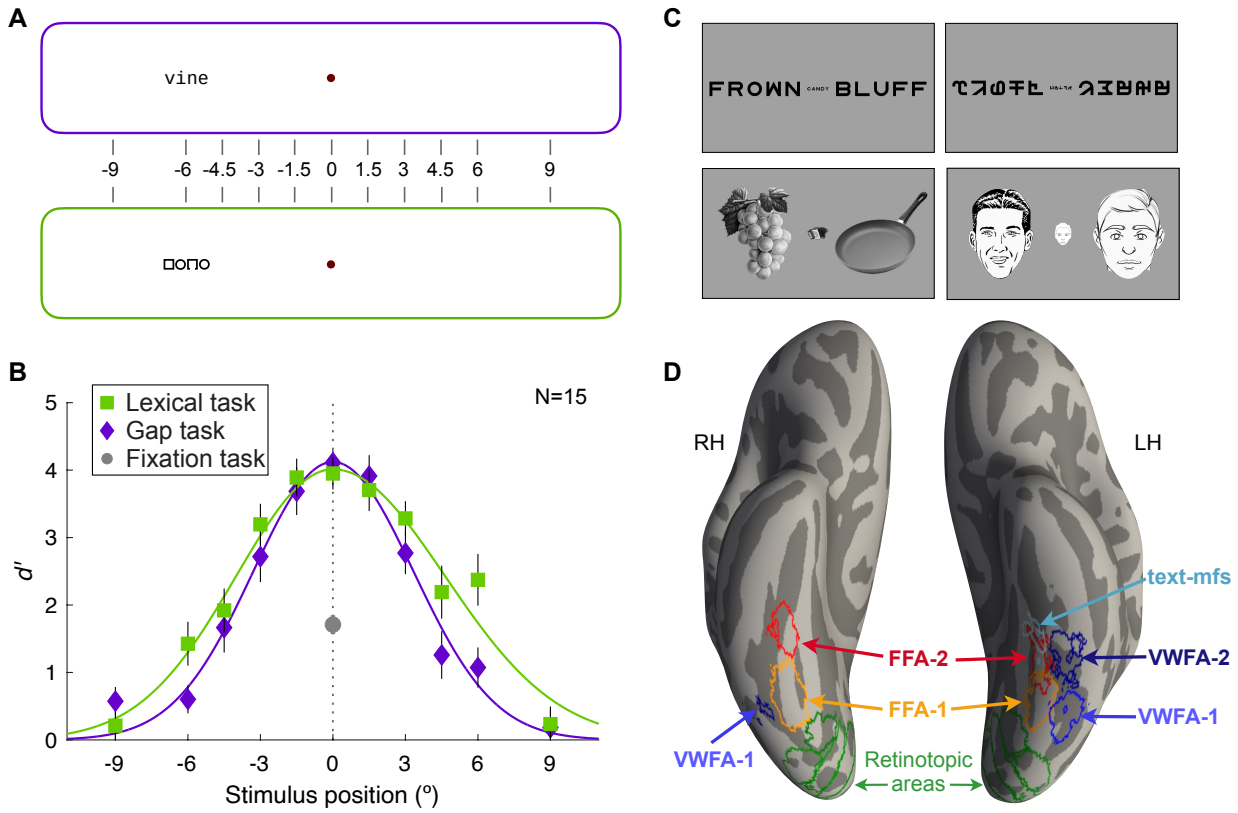

Figure 1. Stimuli, behavioral performance, and ROIs. (A) Stimuli on two example trials, with the 11 possible stimulus locations marked in units of degrees of visual angle. **(B)** Mean behavioral task accuracy (*d'*) for the three tasks, as a function of stimulus position. **(C)** Examples of the four stimulus categories in the localizer scan. *Note: the actual experiment used photographs of real faces, not the cartoon faces shown here.* **(D)** Most likely locations of the ROIs, visualized on a ventral view of the fsaverage surface. This image was created by first projecting individual participant ROIs into the fsaverage surface, then counting the number of participants who had each ROI at each vertex. The outlines show half-max contours of the participant count for each ROI.

The mean response in left VWFA-1 and VWFA-2 (beta weights from a GLM in units of percent signal change, p.s.c.) are shown in **Figure 2**. Both areas showed the expected stimulus selectivity: larger responses to letter strings than shape strings (both F>70, p<10-7). They also showed the expected spatial tuning: larger responses to stimuli in the central visual field than periphery (21). Linear mixed effect models detected significant negative effects of absolute eccentricity (both $P<10^{-4}$). VWFA-1 responded more strongly to stimuli in the right visual field than the left (t(56)=5.46, $p=10^{-6}$). That hemifield asymmetry was marginally stronger for letter stimuli than shapes (F(1,56)=3.77, p=0.057), and did not interact with the task (F(1,56)<0.1). VWFA-2 had no significant preference for stimulus hemifield (t(48)=0.98, p=0.33), for any stimulus type.

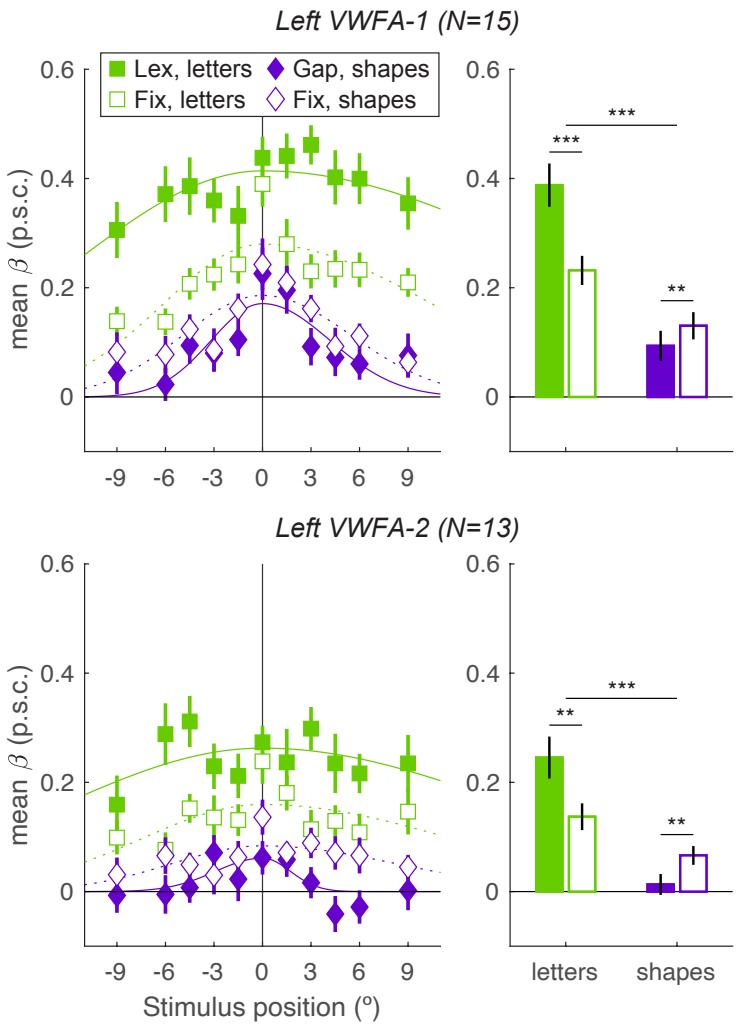

Figure 2: Mean BOLD responses in two sub-regions of the VWFA. *Left column:* Responses as a function of stimulus position, stimulus type, and task (Lex = lexical task; Fix = fixation task). The smooth lines are asymmetric Gaussian functions fit to the data for each stimulus/task condition. *Right column:* The same responses collapsed across stimulus positions. Asterisks indicate the *P*-value for the task effects (short bars over each stimulus type), and task-by-stimulus interactions (long bars): *** *P*<0.001; ** *P*<0.01.

We also found large effects of the participant's task. Responses to letter strings were much larger in the lexical than fixation task, in all text-selective areas. The mean lexical task response was 1.67 times larger than the mean fixation task response in VWFA-1, and 1.79 times larger in VWFA-2. These task effects for letter strings were strong in both areas (both $p<0.01$; BF>14; see **Table 1**). In a secondary analysis that included stimulus position as a predictor, we found that the task effect on responses to letters did not significantly vary with stimulus position, nor did the overall preference for letters over

shapes (all p>0.09). In short, written words evoke much larger responses in the VWFA when task-relevant than ignored.

However, such an enhancement did not occur for the non-letter shape strings. Indeed, responses to shapes were *smaller* during the gap task (when task-relevant) than during the fixation task (when ignored). Overall, the gap task response was 0.72 times the fixation task response in VWFA-1, and 0.18 in VWFA-2. This task-related *suppression* was statistically reliable in both areas (both P<0.01, BF>15). Importantly, in both areas there were also significant interactions between stimulus type (letters, shapes) and task ("attend-stimuli", "attend-fixation"). In VWFA-1, t(14)=9.49, $P=2\times10^{-7}$; BF=83,775. In VWFA-2, t(12)=4.07, P=0.002; BF=27.3.

An additional analysis demonstrated that the VWFA's suppression of shapes in the gap task was absent on the 1st trial of each block. Because the stimulus type varied randomly across blocks of trials, the participant did not know which task they would do (lexical or gap) until the first stimulus in each block appeared. Therefore, the top-down suppressive effect in the VWFAs requires that the subject is actively looking for the gaps in non-letter shapes. See **Supplementary Figure S1** for details.

These curious task effects were largely absent in the other visual ROIs we analyzed. **Figure 3** plots mean responses in other areas, collapsed over stimulus position, and **Table 1** lists the statistics for task effects and the task-by-stimulus interaction in each area. **Supplementary Figure S2** visualizes the strength of that interaction on the average cortical surface, showing that it is restricted to the VWFA. For detailed plots of data from a third text-selective area, text-mfs, see **Supplementary Figure S3.**

Responses in retinotopic visual areas (V1-LO) were weak overall, perhaps because of the brief duration and small size of the stimuli, despite selecting subsets of voxels for each retinotopic position. Nonetheless, significant main effects of task (stimuli attended>ignored) emerged in V3, hV4, VO and LO. The only area with an overall preference for shapes over letters was LO (P=0.005, BF=26). Reassuringly, two visual areas, VO and LO, responded more strongly to shapes during the gap task than during the fixation task (the opposite of what we observed in the VWFAs).

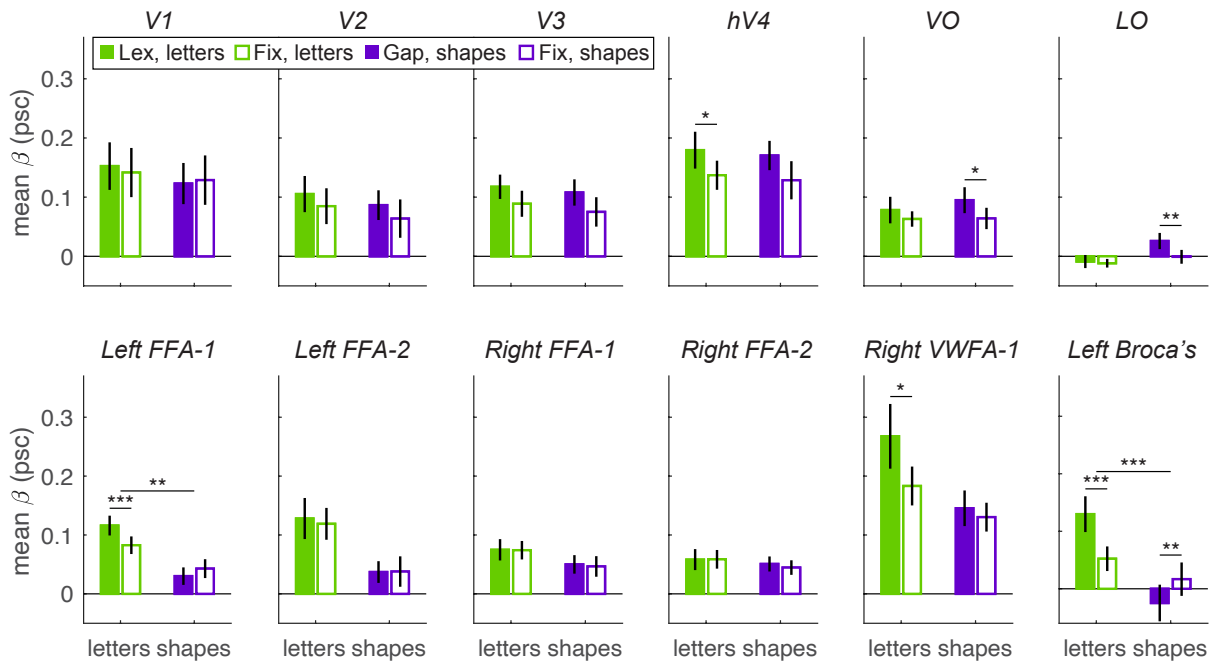

Figure 3: Mean BOLD responses, collapsed across stimulus position, in other areas. The top row is for retinotopic areas, collapsed across left and right hemispheres. Within V1-hV4, before collapsing across positions we extracted the response on each trial from vertices that had PRFs centered over that stimulus position (excluding ±9º, not localizable). Areas VO and LO were defined from published atlases and responses were averaged over all vertices within them. Abbreviations and significance stars as in Figure 2.

One face-selective area, left FFA-1, also responded more strongly to letters during the lexical than the fixation task, but none of the other three face areas showed any task effects or interactions with stimulus type. All the face-selective areas responded more strongly to letters than shapes (all $P<0.001$, BF>20), except right FFA-2 (BF=0.26). Right VWFA-1 behaved similarly to its sibling in the left hemisphere, but the task-by-stimulus interaction was not significant (see Table 1).

Finally, we also defined a putative Broca's area in the left frontal lobe by contrasting (in the average brain) responses to letters vs shapes, across both tasks and all stimulus positions (illustrated in **Figure 5A**). This area was at the border of the left ventral precentral sulcus and inferior frontal sulcus, overlapping Broadman's Area 44 and the pars opercularis, a region known to play a role in lexical decision (41, 42). Activity in Broca's area (bottom right panel of Fig. 3) showed the same task-by-stimulus interaction as the VWFAs: much larger response to letters in the lexical than fixation task, but smaller (near 0) responses to shapes in the gap than fixation task. For activity in Broca's area as a function of stimulus position, see **Supplementary Figure S2.**

Table 1. Statistics on task effects, stimulus effects, and their interaction on BOLD responses. Under "Hem", B = both hemispheres together, L = left, R = right. SE = standard error of the mean, *P* = t-test p-value; BF = Bayes Factor. Significant values are in bold.

			Task effect for Letters (Lex. - Fix.)			Task effect for Shapes (Gap - Fix.)			Interaction (Lex-Fix) - (Gap-Fix)		
ROI	**Hem**	**N**	**Mean (SE)**	***P***	**BF**	**Mean (SE)**	***P***	**BF**	**Mean (SE)**	***P***	**BF**
V1	B	15	0.01 (0.02)	0.86	0.3	-0.01 (0.02)	0.70	0.3	0.02 (0.02)	0.67	0.24
V2	B	15	0.02 (0.02)	0.40	0.4	0.02 (0.02)	0.23	0.5	-0.00 (0.04)	0.70	0.27
V3	B	15	0.03 (0.01)	0.05	1.5	0.03 (0.02)	0.06	1.4	-0.00 (0.02)	0.74	0.23
hV4	B	15	**0.04 (0.01)**	**0.02**	**3.6**	0.04 (0.03)	0.14	0.7	0.00 (0.03)	0.86	0.24
VO	B	15	0.02 (0.02)	0.68	0.3	**0.03 (0.01)**	**0.04**	**2.0**	-0.02 (0.02)	0.31	0.45
LO	B	15	0.00 (0.01)	0.55	0.3	**0.03 (0.01)**	**10^{-2}**	**5.8**	-0.02 (0.01)	0.13	0.61
VWFA-1	L	15	**0.16 (0.02)**	**10^{-4}**	**10^{4}**	**-0.04 (0.01)**	**10^{-3}**	**16**	**0.19 (0.02)**	**10^{-7}**	**10^{6}**
VWFA-1	R	11	**0.08 (0.03)**	**0.04**	**2.0**	0.02 (0.02)	0.43	0.4	0.07 (0.05)	0.16	0.78
VWFA-2	L	13	**0.11 (0.03)**	**10^{-2}**	**14**	**-0.05 (0.01)**	**10^{-3}**	**39**	**0.16 (0.04)**	**10^{-3}**	**76**
text-mfs	L	9	**0.08 (0.02)**	**10^{-3}**	**15**	-0.03 (0.01)	0.10	1.1	**0.11 (0.03)**	**10^{-3}**	**99**
FFA-1	L	14	**0.03 (0.01)**	**10^{-3}**	**51**	-0.01 (0.01)	0.2	0.6	**0.05 (0.02)**	**0.01**	**7.7**
FFA-1	R	15	0.00 (0.01)	0.87	0.3	0.00 (0.01)	0.65	0.3	-0.00 (0.02)	0.86	0.23
FFA-2	L	13	0.01 (0.02)	0.56	0.3	-0.00 (0.02)	0.91	0.3	0.01 (0.03)	0.83	0.25
FFA-2	R	12	0.00 (0.01)	0.73	0.3	0.01 (0.02)	0.53	0.4	-0.01 (0.03)	0.58	0.33
Broca's	L	15	**0.08 (0.01)**	**10^{-4}**	**380**	**-0.04 (0.01)**	**10^{-3}**	**26**	**0.12 (0.02)**	**10^{-4}**	**10^{5}**

In summary, the left VWFAs and left Broca's area showed a unique pattern of activity: compared to when stimuli were ignored during the fixation task, responses to letters were greatly enhanced during the lexical task, but responses to shapes were suppressed. As we discuss below, Broca's area could be the source of the task effects in the VWFAs, providing excitation when words are read and inhibition when non-lexical stimuli are processed for a non-linguistic task.

The VWFAs are sensitive to lexicality and more so during the lexical task

The highly specific task effects in the VWFAs suggest that, beyond encoding orthographic information, they execute computations that are specifically related to performing lexical tasks. To push this idea further, we separately analyzed responses to pseudowords and to real words. Previous studies have demonstrated that the VWFAs respond more strongly to pseudowords, perhaps because pseudowords are processed more effortfully or for more time (11, 15, 25, 26, 43, 44). Here we investigated whether the magnitude of this lexicality effect differed across tasks, and whether it varied with

stimulus positions. We pool data from VWFA-1 and VWFA-2, because their lexicality effects did not differ in either task (*Ps*>0.6, BFs<0.33).

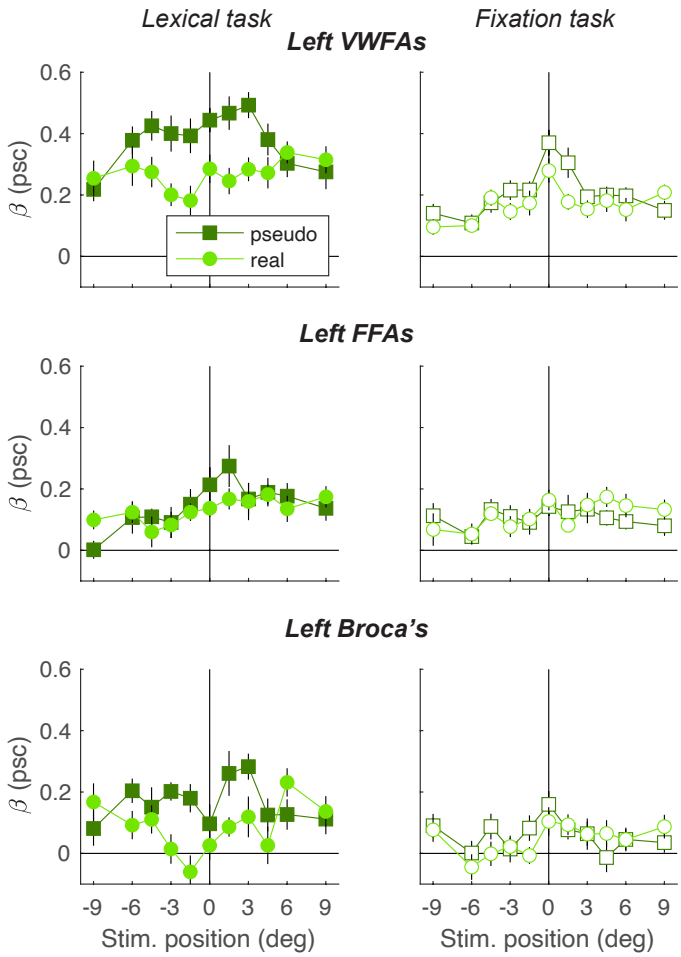

Figure 4: Lexicality effects. Mean BOLD responses to pseudowords (dark green squares) vs. real words (light green circles) in the left VWFAs (top row), the left FFAs (middle row), and Broca's area (bottom row). Data are separated by the two tasks: lexical decision in the left column, and fixation task in the right column. Error bars = +/- 1 SEM.

The mean responses in the VWFAs were larger for pseudowords than real words ($F(1,350)=43.6$, $P<10^{-9}$; BF=5×10^{7}), and overall larger in the lexical than fixation task ($F(1, 354)=140$, $P<10^{-26}$; BF=10^{20}). The lexicality effect was present in both tasks assessed separately, but significantly smaller in the fixation task (interaction $F(1,354)=12$, $P<0.001$; BF=22). In the lexical task, averaging across stimulus positions, responses to pseudowords were on average 1.63 times the response to real words (mean difference = 0.12 p.s.c; SEM=0.02, $t(14)=6.1$, $P=3\times10^{-5}$; BF=881). In the fixation task, the effect was less

than half as large: responses to pseudowords were 1.28 times the response to real words (mean difference = 0.04 p.s.c; SEM=0.014; t(14)=2.55, *P*=0.023; BF=2.83).

Also, the lexicality effect decreased with absolute stimulus eccentricity (F(1,354)=3.93, *P*=0.047). It is noteworthy that pseudowords evoked larger responses than real words only within 6º eccentricity. This corresponds to behavioral lexical decision task accuracy, which was near chance beyond 6º (see plot in Fig. 1D). Thus, although mean BOLD responses during the lexical task were high for stimuli across the visual field (Fig. 2), the VWFA's *differential* response to pseudowords and real words roughly mirrors participants' ability to distinguish those two categories of stimuli.

One interpretation of these data is that the VWFA's response to pseudowords is elevated only when the language system is engaged to recognize words (15). The residual effect present in the fixation task could be due to occasional shift of attention away from the fixation mark and onto the word (which seems to have happened only for words at the fovea or just to the right).

In the union of left FFA-1 and FFA-2 (Fig.4, middle row), we observed a main effect of task (F(1,354)=7.3, *P*=0.01, BF=3.1), but no effect of lexicality (F(1,354)=0.26, *P*=0.61, BF=0.1) or interaction (F(1,354) =1.68, *P*=0.19, BF=0.28). Responses in Broca's area (Fig. 4, bottom) were more like the VWFAs: similar main effects of task (lexical>fixation; F(1,354)=37.5, $P<10^{-8}$, BF=7×10^{3}), lexicality (pseudo>real; F(1,354)=13.8, $P<10^{-3}$, BF=27), and interaction (F(1,354)=6.8, *P*=0.009, BF=1.51). There was no significant lexicality effect during the fixation task in Broca's area (F(1, 176)=0.52, *P*=0.47, BF=0.15).

Task-dependent connectivity between VWFA and Broca's area

As shown above (Figs 2-4), the left VWFAs and Broca's area showed similar effects of stimulus content and task. One possible explanation is that the task effects in the VWFAs are top-down influences from language regions, including Broca's area. Such an explanation predicts correlated activity in the two areas, contingent on the task. Specifically, if the response on each trial in the VWFA is modulated by top-down input from Broca's area, then incidental trial-to-trial response fluctuations should be correlated between the two areas. To test that prediction, we did the following functional connectivity analysis with Broca's area as the "seed." For each surface vertex, we extracted the "residuals" in single-trial responses by subtracting out the across-trial

mean response for each stimulus type (real words, pseudowords, shape strings), task, and visual field position. Then we computed the correlation between each vertex's residuals and the mean residuals in Broca's area.

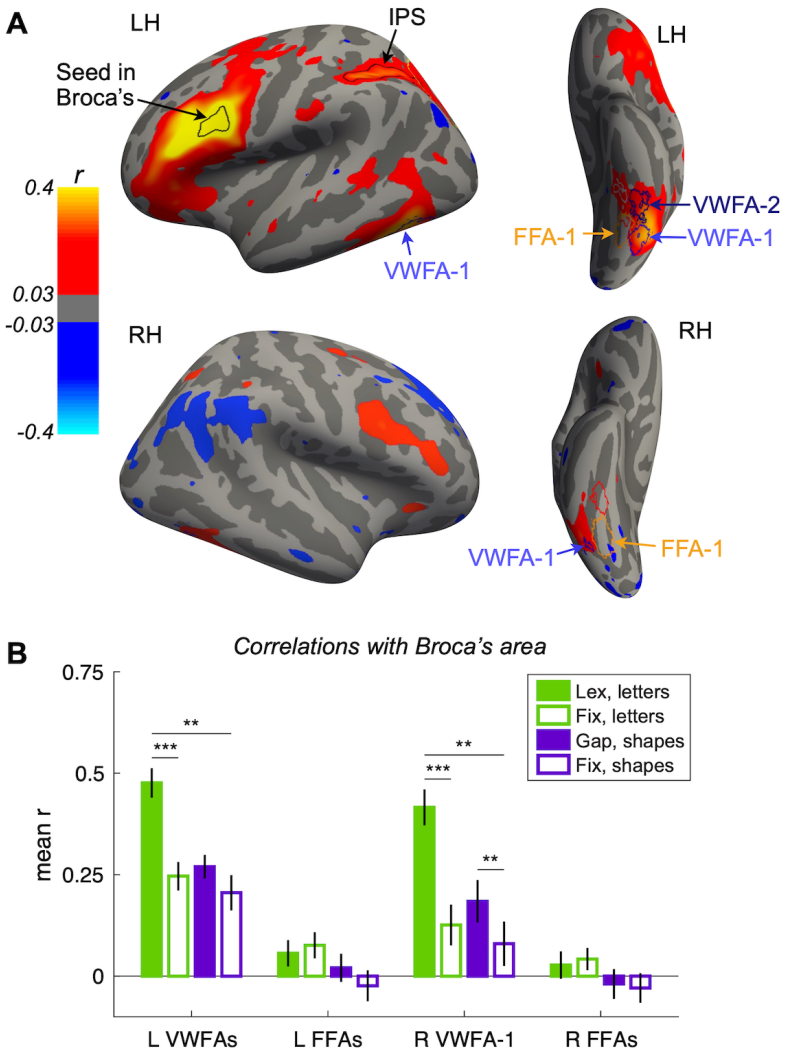

Figure 5: Functional connectivity with Broca's area. A: Maps of mean correlations with Broca's area during the lexical decision task, on Freesurfer's fsaverage surface. Before averaging, each subject's data were smoothed with a 2D Gaussian kernel (full-width at half-maximum = 5 mm). The data are masked to show only vertices where the correlation was significant (p<0.05, corrected for false discovery rate), peaking at r>=0.4. Lateral and ventral views of both hemispheres (LH and RH) are pictured. The "Broca's area" ROI is outlined in black, as is an IPS region drawn to encompass r>0.2. Just posterior to that, outlined in yellow, is IPS-0, -1 and -2. **B**: Across-subject mean correlation coefficients between Broca's area and several ventral temporal regions, in each task & stimulus condition. "VWFAs" is the average across VWFA-1 and -2, and "FFAs" is the average across FFA-1 and -2. Asterisks and abbreviations as in Figure 2.

Figure 5A shows surface parameter maps of those correlation coefficients, computed for trials with the lexical task, averaged over subjects. Broca's area is a hotspot because it correlates well with itself. The correlations are high in the left VWFA-s, but not in the neighboring face areas. Right VWFA-1 also shows significant correlation with Broca's, as does the intraparietal sulcus (IPS). A right hemisphere homologue of Broca's also appeared. These patterns are specific: it is not that the VWFAs correlate with many brain areas. **Supplementary Figure S5** demonstrates that when left VWFA-1 is the "seed" region, the same patches of cortex show significant correlation as when Broca's is the seed. For pairwise correlations between 13 different regions in each task condition, see **Supplementary Figures S6 and S7.**

More importantly, these across-region correlations depended jointly on stimulus type and task. **Figure 5B** plots across-subject mean correlation coefficients with Broca's area extracted from key ROIs. In the left VWFAs (mean of VWFA-1 and -2), the correlation was high (r=0.48) during the lexical decision task, but roughly half as strong in all three other conditions. When letters were on the screen but ignored (fixation task), the correlation between Broca's and the VWFAs was not even as strong as when non-letter shapes were attended (gap task). The effect of task on the correlation for trials with letters (lexical - fixation task) was large (t(14) = 6.54, $P=10^{-5}$, BF=1732), but there was little to no task effect for shapes (t(14) = 1.43, P=0.17, BF=0.61), and there was a strong interaction (t(14)=3.77, P=0.002, BF=21). Importantly, within the fixation task, the correlation when letters were presented was not any stronger than when shapes were presented (t(14)=1.29, P=0.22; BF=0.52).

When we analyzed data from the lexical task in the two left VWFA subregions separately, we found that VWFA-1 and VWFA-2 were correlated with Broca's to similar degrees (r=0.52 vs r=0.43; P=0.28, BF=0.47). Right VWFA-1 showed a similar correlation pattern as the left VWFAs, except with a significantly stronger correlation for responses to shapes during the gap than fixation task (t(10)=3.49, P=0.006; BF=9.4). The FFAs showed very little correlation with Broca's at all, and no effects of task, although the left FFAs did have a slightly stronger correlation when letters were present than shapes (F(1,56)=7.40, P=0.009). For a full analysis of activity in the IPS region outlined in black in Figure 5A, see **Supplemental Figure S8**. It responded more strongly when stimuli were attended than ignored, but roughly equally to letters and shapes. Its functional connectivity with Broca's area was highest during the lexical task, just like the VWFA.

DISCUSSION

By carefully manipulating both stimulus parameters and the participant's task, we revealed highly specific top-down effects on BOLD activity in ventral occipito-temporal cortex. In brief, instructing the participant to engage in a lexical decision task nearly doubled the VWFA's responses to letter strings across the visual field, compared to when the participant focused on the color of the fixation dot and ignored the letter strings. The difference in response between real words and novel pseudowords was also enhanced: the lexicality effect was barely present in the fixation task but quite large in the lexical task (for a related result, see ref. (15)). Remarkably, when the participant judged the location of a gap in a string of non-letter shapes, the VWFA's response was *reduced* compared to when those shapes were ignored during the fixation task. This pattern was specific to the VWFAs and Broca's area, and is the opposite of the expected enhancement for attended stimuli (which did arise in retinotopic areas VO and LO).

Functional connectivity shed further light on these findings. The trial-to-trial fluctuations in response magnitude (variance not explained by any of our stimulus or task parameters) were selectively correlated between the VWFAs, a putative left frontal Broca's area, and the IPS. Compared to other visual areas, Broca's had privileged connectivity to the VWFA, in all conditions. However, that activity correlation was roughly twice as strong during the lexical task as during the other tasks. The connectivity seems to require engagement in a word recognition task, rather than the mere physical presence of words, because during the fixation task the correlation was not higher when letters were presented than when shapes were presented.

Altogether, these results suggest that the cortical reading network is activated by voluntary engagement in a word recognition task, and can be suppressed by engaging in a non-linguistic task or ignoring presented words. One hypothesis that we favor is that Broca's area is a source of control for other parts of the reading network, including the VWFA. When Broca's area is engaged, it communicates with the VWFA and boosts activity there, especially when the letter string is unfamiliar (e.g., a pseudoword) or difficult to recognize (e.g., presented outside the fovea). These areas may communicate iteratively until the letter string is matched to an item in the mental lexicon, or can be discarded as a non-word. When the participant engages in a non-linguistic task that

requires isolating a single feature in a single shape (rather than integrating the identities of many shapes), Broca's area is less active and the VWFAs are suppressed.

The lexicality effect (pseudo>real words) in the VWFAs could be directly caused by feedback from Broca's area when the observer reads words. However, that hypothesis is not supported by direct intracranial recordings: sensitivity to lexical features emerges first in a mid-fusiform text-selective area (text-mfs), not later than in frontal areas (7). An alternate hypothesis is that excitatory feedback from frontal or parietal cortex is necessary to fully process written words, especially when the task is difficult, but that orthographic lexical access first occurs locally in ventral temporal cortex.

A previous study by two of the present authors is an interesting point of comparison (14). They also measured VWFA responses to words and other stimuli that were either task-relevant or irrelevant. The results were notably different from those reported above: the effect of making the stimuli task-relevant was to positively scale the responses to *all types* of stimuli. Several aspects of our design might account for the difference: first, the lexical decision task we used likely places stronger demands on the language system than the one-back memory task or face vs. word categorization task used by ref. (14). Second, our gap task has special properties: it requires extracting information from just one shape in a string that is crowded by irrelevant flanking stimuli. That is the opposite of what the VWFA is theoretically trained to do, namely, integrating the shapes in a string into a coherent whole. That could be why performing the gap task reduces VWFA activity. Taken together, these results highlight the importance of studying how specific tasks place varying demands on visual, language and attention systems to modulate the response properties of word-selective cortex.

One common finding in our study and its predecessors is the involvement of the IPS. Kay & Yeatman (2017) concluded that IPS (specifically, areas IPS-0 through IPS-2) integrates sensory information from VTC and sends feedback that boosts weak signals (14), see also (45, 46). These prior data fit well under the broader theory that the IPS is part of a dorsal attentional control system (47, 48). Note however that the specific regions (IPS-0/1/2) associated with visual attention (14) are more posterior than the part of the IPS that we found to correlate well with Broca's and the VWFA in the present study (see Figs. 5 and S4). This more anterior IPS is associated with letter position encoding (49), lexical processing (7), orthographic working memory (50), and even

semantic memory (51). Like us, other authors have found IPS to be functionally connected to both the VWFA and Broca's area (32, 36, 40, 52, 53). Therefore, it may do more than just boost sensory signals generally, but control language-specific modulations such as we observed, in concert with left frontal language regions.

Another insight provided by our data is that engaging in a lexical task boosts responses to words even when they are presented in peripheral vision. Previous studies have reported that the left VWFA has a limited "field of view" that extends only a few degrees to the left of fixation and drops off quickly between 5-10º to the right (21, 22). They measured responses to a sweeping bar during a fixation task. We found a much flatter spatial profile during the lexical decision task: single flashed words evoked large responses even at 9º eccentricity to the left and right. However, the *difference* in response between words and pseudowords was more limited to the central visual field, roughly matching behavioral accuracy. These data emphasize that VWFA's activity must be interpreted with respect to both the stimulus drive and top-down signals. The behaviorally relevant field of view that relates to how well the person can recognize words may not be a simple reflection of the overall BOLD response.

One important and challenging question arising from this study is: are the task effects due to "visual attention"? In a loose sense, they must be, because the tasks required the participant to pay attention to different things. But we can be more specific. Covert, endogenous spatial attention could play a role: the fixation task required a narrow focus of attention on just the very center of the screen, whereas both the lexical and gap tasks required an even distribution of attention across all 11 possible stimulus locations. But spatial attention alone cannot explain why the VWFA response to letters was enhanced, while the response to shapes was suppressed. Thus, we must invoke other mechanisms to explain our results. Another form of attention is feature-based: a boost in activity for neurons that are tuned to task-relevant features like colors or motion directions (5, 54, 55). Our tasks differed in the task-relevant features: red color in the fixation task; high-spatial frequency form, letter identities, and word familiarity in the lexical task; and tiny discontinuities across contours in the gap task. Could some combination of feature-based and spatial attention explain our data? Perhaps, but only if we extend our concept of feature-based attention to include a dimension along which the features relevant for the lexical and gap tasks are at opposite ends, with the features relevant for the fixation task in the middle.

Rather than drastically stretch the models of visual attention that have been elegantly applied to visual cortex in the past, we argue for a more inclusive view of task effects in visual cortex beyond attention. In other contexts, top-down modulations in ventral temporal cortex have proved more complex than the predicted effects of spatial and feature-based attention (45). Our data point to flexible and specific integration of visual and linguistic information that depends on top-down signals to support reading. Reading is a uniquely human skill that requires many hours of dedicated instruction to reshape cortical networks (6). Perhaps as a result of that rewiring, voluntary engagement in a word recognition task elicits communication between higher-level language areas and the specific regions of visual cortex that have become specialized to process text (the VWFAs). Those regions are suppressed during another stimulus-directed task for which the VWFA and language regions are not adept.

Overall, our data violate the predictions of general attentional enhancement and instead suggest activity in ventral visual cortex depends on a triple conjunction of task demands, stimulus content, and the selectivity of each brain region. Top-down enhancement is observed specifically in the VWFA only when words are presented and the observer is engaged in a lexical task. Top-down suppression is applied to the VWFA when processing non-linguistic stimuli. One caveat, however: some conditions remain to be tested that would confirm such a triple conjunction hypothesis. For instance, we have yet to test how the VWFA responds to non-linguistic stimuli presented while the participant is engaged in a lexical task. Data such as those will continue to refine our theories of what exactly the VWFA does during reading. More generally, our hope is that ongoing research will further map out the space of cognitive and sensory factors that determine activity in visual cortex, allowing flexible use of vision for any task.

MATERIALS AND METHODS

This study was approved by the Institutional Review Board of Stanford University and complies with all relevant ethical regulations. All participants gave written informed consent and were paid a fixed monetary reward.

Participants

15 volunteers (10 female) participated. Their ages ranged from 19 to 28 years (mean = 23.8), and 14 were right-handed. All had normal or corrected to normal vision, and no

history of dyslexia or other cognitive disorders. All scored at or above the population norm of 100 on the TOWRE-II tests of sight word efficiency and phonemic decoding efficiency (56): means (and SDs) = 120 (9) and 117 (9), respectively. One additional participant was excluded from the analyses for falling asleep in nearly every scan and performing near chance even for stimuli at fixation.

Equipment

We acquired MRI data at the Center for Cognitive Neurobiological Imaging at Stanford University on a 3T GE Discovery MR750 scanner (GE Medical Systems) using a 32-channel head coil. In each session we collected one T1-weighted structural scan with 0.9 mm isotropic voxel size. We acquired functional data with a T2* sensitive gradient echo EPI sequence with a multiplexing (multiband) factor of 3 to acquire whole-brain coverage (51 slices). The TR was 1.19 s, TE was 30 ms and flip angle was 62°. The voxel size was 2.4 mm isotropic.

Via a mirror mounted above their nose, participants viewed the stimuli on an LCD screen (total viewing distance = 280 cm). The display had a resolution of 1920 x 1080 pixels, refreshing at 60 Hz. We presented the stimuli with custom MATLAB software (MathWorks, Natick, MA, USA) and the Psychophysics Toolbox (57, 58). Throughout each scan we recorded monocular gaze position with an SR Research Eyelink 1000 tracker. Calibration was usually successful (details below), and even when it was not, participants believed their fixation was being monitored. Participants responded to the tasks by pressing two buttons on a response pad held in their right hand.

Main Experiment: Stimuli

Figure 1A shows two example stimuli. The screen's background luminance was set to 80% of its maximum. The visual displays on each trial consisted of a persistent fixation dot of diameter 0.11 degrees of visual angle (dva) at the screen center, and one black stimulus string that flashed for 150 ms at a random one of 11 positions along the horizontal meridian. There were two stimulus types: letter strings and shape strings. The letter strings were all composed of 4 letters in "Liberation Mono" (a monospaced font similar to Courier). The font size was set such that the "x" was 0.41 dva tall. The 4-letter strings were on average 1.69 dva wide (range 1.62-1.76), with 0.44 degrees between the centers of neighboring letters. The stimulus set contained 264 unique letter strings, half of which were pronounceable pseudowords with constrained bigram

statistics generated by MCWord (59). The other half were high-frequency real words of all syntactic categories (e.g., nouns and verbs). The mean frequency was 549 per million, ranging from 195 to 1,884.

Each shape string was composed of 4 black circles and squares, matched in size and spacing to the letters (height = 0.43 dva, spacing = 0.44 dva). Each shape was composed of black lines 5 pixels wide (the same as most letter contours). There were 16 unique strings, each composed of 4 shapes that were independently and randomly set to be a square or a circle. Before being presented, one of the inner two letters had a gap added either to the top or the bottom side. This gap was 0.17 dva wide, equal in size to the gap in the letter "c".

The fixation dot was usually dark gray in color (40% of maximum screen luminance), but on a random 50% of all trials it turned dark red during the 150 ms of stimulus presentation. When a stimulus was centered on fixation, the dot was superimposed onto the stimulus to remain visible.

Main experiment: Trials and tasks

Each 4-s trial was composed of one stimulus, a letter or shape string, presented for 150 ms, followed by a 3850 ms interval during which the subject could press a button to respond, followed immediately by the next trial. Trials came in blocks of 6. The stimulus type was constant within each block but varied randomly from block to block. Between blocks were blank periods of rest (no task except fixation on the dot), lasting 4, 6, or 8 seconds (randomly assigned).

The participant performed three different tasks at different times. Half of the runs (scans lasting ~6 minutes) were "attend-fixation" and half were "attend-stimuli." During "attend-fixation" runs, the participants ignored the letter and shape strings and performed the fixation task. The task was to press one of two keys to report whether or not the fixation dot turned red. The saturation of the red color (in HSV space) was controlled by a staircase to converge on the 80% correct detection threshold.

During the "attend-stimuli" runs, participants made judgments of the shape or letter strings while maintaining fixation on the dot but ignoring changes in its color. During blocks of trials with shape strings, the participant performed the gap task: to report whether the gap was on the top or bottom side. During blocks of trials with letter

strings, the participant performed the lexical decision task: to report whether the presented letter string was a real word (e.g., book) or a pseudoword (e.g., blus). The task-irrelevant changes in fixation dot color were "replayed" from the staircases in fixation task runs. Each subject saw each letter string once during the MRI experiment, and we took care to equalize the sets of stimuli presented at different locations in terms of metrics that could affect task difficulty and BOLD response (see **Supporting Information** for details).

Procedure

Participants practiced the lexical decision and gap tasks for at least 2 one-hour sessions outside the scanner, with immediate feedback about gaze fixation as well as auditory feedback about accuracy on each trial. They then participated in two MRI scanning sessions. The goal was to complete 4 runs of the localizer (see **SI**) and 3 runs of retinotopy in the first session, and 8 runs of the main experiment in the second.

The main experiment included 528 trials: 132 for each stimulus type (letters, shapes) and task (attend-stimuli, attend-fixation) combination. Within those, there 12 trials per visual field position. In a few cases we collected 1 fewer run than planned due to time constraints. We excluded scans in which any framewise displacement due to head motion exceeded 2.4 mm (1 voxel). This applied to 4 scans from 1 subject and 3 scans from a second subject. When computing statistics, we weighted each participant's data by the number of trials they completed.

MRI data preprocessing

We used *fMRIPrep* 20.2.1 (60) to carry out the following pre-processing steps: inhomogeneity correction and segmentation of the T1-weighted images; cortical reconstruction; susceptibility correction of BOLD scans, registration of functional to structural scans; motion correction; and resampling data to native surfaces and the fsaverage space. See the **SI** for details.

BOLD response estimation

For both the localizer and main experiment, we conducted GLMs to estimate BOLD responses to the stimuli on each trial. These responses (beta weights) are expressed in percent signal change, and reflect changes relative to the "blank" periods when the participant was simply fixating a dot on an otherwise empty screen. For the localizer

data, we used GLMdenoise (61) to estimate the across-trial mean beta weight for each stimulus category. For the main experiment, we used GLMsingle (62, 63) to estimate single-trial beta weights. The design matrix coded each 150 ms stimulus presentation as a separate event. Both GLMdenoise and GLMsingle optimize the assumed hemodynamic response functions and remove from the final estimations a set of noise regressors unrelated to the task and stimulus. Compared to GLMdenoise, GLMsingle estimates hemodynamic response functions on a per-voxel basis and also introduces ridge regression to improve stability and accuracy of single-trial beta weights.

ROI definition

We defined word- and face-selective regions of interests (ROIs) using data from a separate localizer experiment (**Figure 1C**). Participants viewed sequences of images from 4 different categories: faces, objects, letter strings, and false fonts (provided by (64, 65)). Full details are in the **SI.** We computed contrasts of BOLD responses during the localizer experiment (text vs. false fonts, faces and objects; faces vs. text, false fonts and objects). We drew the ROIs based upon a visualization of the contrast t-statistic (t>3) on each subject's native surface.

We defined three text-selective ROIs: (1) VWFA-1 is anterior to the V4/VO1 border, in the posterior occipito-temporal sulcus (OTS). (2) VWFA-2 is anterior to VWFA-1, usually in the OTS but sometimes extending onto the gyri on either side. In some subjects VWFA-2 and VWFA-1 appeared contiguous at the chosen contrast threshold, but they always had separate peaks of text selectivity. We based these regions upon a prior publication (18). (3) Finally, we noted in some participants (9/15) a third text-selective region medial to VWFA-2, near the mid-fusiform sulcus (MFS), which we termed text-mfs. The text-mfs region may be the same as reported in electrocorticography studies (11, 66) and an fMRI study (27). Some participants also had text-selective blobs posterior to VWFA-1, but we did not analyze them as they were highly variable and often extended into the area occupied by hV4.

In addition, we defined two face-selective regions in the fusiform gyrus: one relatively posterior and usually medial to VWFA-1, which we call FFA-1, and another medial to VWFA-2, which we call FFA-2 (similar to what others have called pFus-faces and mFus-faces; (12)). **Table 2** notes the number of participants who had each ROI in each

hemisphere, and **Figure 1D** displays the most likely locations of each ROI on the fsaverage surfaces. For images of each participant's ROIs, see **Supplemental Figure S9**.

We also defined early visual areas (bilateral V1, V2, V3, hV4, and VO1/2) with data from separate retinotopic mapping scans (67). To analyze responses during the main experiment in these retinotopic areas, we selected subsets of voxels corresponding to each stimulus position. Stimuli at +/-9 deg were excluded from analysis of retinotopic areas, because the displays used for retinotopic mapping did not extend out that far. See the **SI** for details. Although not clearly visible in our own retinotopy data, we also extracted the approximate locations of lateral occipital area LO-1/2 (68) based on a published atlas (69).

Table 2 Count of subjects with ROIs for each text- and face-selective region, in each hemisphere (LH: left hemisphere; RH: right hemisphere).

Region	LH	RH
VWFA-1	15	11
VWFA-2	13	5
Text-MFS	9	4
FFA-1	14	15
FFA-2	13	12

Finally, we defined a putative Broca's area in the left pre-central sulcus, based on data in the main experiment. For each subject, at each surface vertex, we computed the mean difference in responses to letters vs. shapes, collapsing across positions and task conditions. Those difference maps were warped to the fsaverage template surface, and then we computed across-subject statistics at each vertex. We defined Broca's area as a reliably text-selective blob where the t>5. We then back-transformed that ROI into each participant's native surface.

Statistical analyses

We used a combination of statistical tests to compare across-subject mean data to null hypotheses that predict no effect of some manipulation. In many cases we used linear mixed-effect models with random effects for participants. For models of the interaction of two or three predictor variables, the results are reported as repeated-measures

ANOVAs (F-statistics and p-values). For tests of the effect of a mean difference between two conditions, we conducted paired t-tests. We also report Bayes Factors (BFs) for each test, to quantify strength of evidence. The BF is the ratio of the probability of the data under the alternate hypothesis (that two conditions differ), relative to the probability of the data under the null hypothesis (70). For example, a BF of 10 indicates that the data are ten times more likely under the alternate hypothesis than the null hypothesis. We computed BFs using the bayesFactor toolbox in MATLAB (https://github.com/klabhub/bayesFactor: DOI: 10.5281/zenodo.4394422).

Functional connectivity analysis

We assessed whole-cortex functional connectivity with two "seed" regions: left VWFA-1 and Broca's area (see ROI definitions above). For each stimulus and task condition (e.g., real words during the lexical task), for each subject, for each surface vertex, we subtracted from each trial's response the mean across all trials with that type of stimulus at the same location in the same task. Real words and pseudowords were treated separately in this analysis. Then, for each stimulus/task condition, we computed the correlation between each vertex's de-meaned responses and the de-meaned responses in the seed area (averaged across its vertices). Each such correlation coefficient estimates the correlation in trial-to-trial variance not explained by overall effects of stimulus type, position, and task.

ACKNOWLEDGEMENTS

We are grateful to Brian Wandell, Kalanit Grill-Spector, and Anthony Norcia for help designing the stimuli and tasks, as well as to Vassiki Chauhan for advice on the manuscript. Funding provided by NIH R00 EY029366, R01 HD095861, and P41 EB027061; NSF IIS-1822683 and IIS-1822929.

DATA AVAILABILITY

Data and materials will be shared upon publication of this manuscript.

REFERENCES

1. P. R. Roelfsema, F. P. de Lange, Early Visual Cortex as a Multiscale Cognitive Blackboard. *Annu. Rev. Vis. Sci.* **2**, 131–151 (2016).
2. J. H. R. Maunsell, Neuronal Mechanisms of Visual Attention. *Annu. Rev. Vis. Sci.* **1**, 373–391 (2015).
3. A. C. Nobre, S. Kastner, D. M. Beck, *Neural Systems for Spatial Attention in the Human Brain* (2014) https:/doi.org/10.1093/oxfordhb/9780199675111.013.011.
4. M. Carrasco, Visual attention: The past 25 years. *Vision Res.* **51**, 1484–1525 (2011).
5. J. H. R. Maunsell, S. Treue, Feature-based attention in visual cortex. *Trends Neurosci.* **29**, 317–322 (2006).
6. J. D. Yeatman, A. L. White, Reading: The Confluence of Vision and Language. *Annu. Rev. Vis. Sci.* **7**, 487–517 (2021).
7. O. Woolnough, *et al.*, A Spatiotemporal Map of Reading Aloud. *J. Neurosci.* **42**, JN-RM-2324-21 (2022).
8. B. D. McCandliss, L. Cohen, S. Dehaene, The visual word form area: Expertise for reading in the fusiform gyrus. *Trends Cogn. Sci.* **7**, 293–299 (2003).
9. S. Caffarra, I. I. Karipidis, M. Yablonski, J. D. Yeatman, Anatomy and physiology of word - selective visual cortex : from visual features to lexical processing. *Brain Struct. Funct.* (2021) https:/doi.org/10.1007/s00429-021-02384-8.
10. R. Gaillard, *et al.*, Direct Intracranial, fMRI, and Lesion Evidence for the Causal Role of Left Inferotemporal Cortex in Reading. *Neuron* **50**, 191–204 (2006).
11. O. Woolnough, *et al.*, Spatiotemporal dynamics of orthographic and lexical processing in the ventral visual pathway. *Nat. Hum. Behav.* **5**, 389–398 (2021).
12. K. Grill-Spector, K. S. Weiner, The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* **15**, 536–548 (2014).
13. M. Ben-Shachar, R. F. Dougherty, G. K. Deutsch, B. A. Wandell, Differential sensitivity to words and shapes in ventral occipito-temporal cortex. *Cereb. Cortex* **17**, 1604–1611 (2007).
14. K. N. Kay, J. D. Yeatman, Bottom-up and top-down computations in word- and face-selective cortex. *Elife* **6**, 1–29 (2017).
15. Q. R. Mano, *et al.*, The role of left occipitotemporal cortex in reading: Reconciling stimulus, task, and lexicality effects. *Cereb. Cortex* **23**, 988–1001 (2013).
16. G. Lerma-Usabiaga, M. Carreiras, P. M. Paz-Alonso, Converging evidence for functional and structural segregation within the left ventral occipitotemporal cortex in reading. *Proc. Natl. Acad. Sci.* **115**, E9981–E9990 (2018).
17. F. Vinckier, S. Dehaene, A. Jobert, J. P. Dubus, M. Sigman, Hierarchical coding of letter strings in the ventral stream: Dissecting the inner organization of the visual word-form system. *Neuron* **55**, 143–156 (2007).

18. A. L. White, J. Palmer, G. M. Boynton, J. D. Yeatman, Parallel spatial channels converge at a bottleneck in anterior word-selective cortex. *Proc. Natl. Acad. Sci.* **116**, 10087–10096 (2019).
19. M. Szwed, *et al.*, Specialization for written words over objects in the visual cortex. *Neuroimage* **56**, 330–344 (2011).
20. A. M. Rauschecker, R. F. Bowen, J. Parvizi, B. A. Wandell, Position sensitivity in the visual word form area. *Proc. Natl. Acad. Sci.* **109**, E1568–E1577 (2012).
21. R. K. Le, N. Witthoft, M. Ben-Shachar, B. A. Wandell, The field of view available to the ventral occipito-temporal reading circuitry. *J. Vis.* **17**, 1–19 (2017).
22. G. Lerma-usabiaga, R. Le, C. Gafni, M. Ben-shachar, B. A. Wandell, Interpreting sensory and cognitive signals in the cortical reading network (2021).
23. E. A. Hirshorn, *et al.*, Decoding and disrupting left midfusiform gyrus activity during word reading. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 201604126 (2016).
24. L. S. Glezer, X. Jiang, M. Riesenhuber, Evidence for Highly Selective Neuronal Tuning to Whole Words in the "Visual Word Form Area." *Neuron* **62**, 199–204 (2009).
25. M. Kronbichler, *et al.*, Taxi vs. taksi: On orthographic word recognition in the left ventral occipitotemporal cortex. *J. Cogn. Neurosci.* **19**, 1584–1594 (2007).
26. M. Kronbichler, *et al.*, The visual word form area and the frequency with which words are encountered: Evidence from a parametric fMRI study. *Neuroimage* **21**, 946–953 (2004).
27. F. Bouhali, Z. Bézagu, S. Dehaene, L. Cohen, A mesial-to-lateral dissociation for orthographic processing in the visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 21936–21946 (2019).
28. L. Reich, M. Szwed, L. Cohen, A. Amedi, A ventral visual stream reading center independent of visual experience. *Curr. Biol.* **21**, 363–368 (2011).
29. G. Dzięgiel-Fivet, *et al.*, Neural network for Braille reading and the speech-reading convergence in the blind: Similarities and differences to visual reading. *Neuroimage* **231** (2021).
30. K. Siuda-Krzywicka, *et al.*, Massive cortical reorganization in sighted braille readers. *Elife* **5**, 1–26 (2016).
31. C. J. Price, J. T. Devlin, The interactive account of ventral occipitotemporal contributions to reading. *Trends Cogn Sci* **15**, 246–253 (2011).
32. L. Chen, *et al.*, The visual word form area (VWFA) is part of both language and attention circuitry. *Nat. Commun.* **10**, 1–12 (2019).
33. J. D. Yeatman, A. M. Rauschecker, B. A. Wandell, Anatomy of the visual word form area: Adjacent cortical circuits and long-range white matter connections. *Brain Lang.* **125**, 146–155 (2013).
34. F. Bouhali, *et al.*, Anatomical Connections of the Visual Word Form Area. *J. Neurosci.* **34**, 15402–15414 (2014).

35. Z. M. Saygin, *et al.*, Connectivity precedes function in the development of the visual word form area. *Nat. Neurosci.* **19**, 1250–5 (2016).
36. A. C. Vogel, F. M. Miezin, S. E. Petersen, B. L. Schlaggar, The putative visual word form area is functionally connected to the dorsal attention network. *Cereb. Cortex* **22**, 537–549 (2012).
37. W. D. Stevens, D. J. Kravitz, C. S. Peng, M. H. Tessler, A. Martin, Privileged functional connectivity between the visual word form area and the language system. *J. Neurosci.* **37**, 5288–5297 (2017).
38. J. Li, D. E. Osher, H. A. Hansen, Z. M. Saygin, Innate connectivity patterns drive the development of the visual word form area. *Sci. Rep.*, 1–12 (2020).
39. M. S. Koyama, *et al.*, Resting-state functional connectivity indexes reading competence in children and adults. *J. Neurosci.* **31**, 8617–8624 (2011).
40. D. López-Barroso, *et al.*, Impact of literacy on the functional connectivity of vision and language related networks. *Neuroimage* **213** (2020).
41. J. R. Binder, *et al.*, Neural correlates of lexical access during visual word recognition. *J. Cogn. Neurosci.* **15**, 372–393 (2003).
42. S. Heim, S. B. Eickhoff, A. K. Ischebeck, G. Supp, K. Amunts, Modality-independent involvement of the left BA 44 during lexical decision making. *Brain Struct. Funct.* **212**, 95–106 (2007).
43. J. S. H. Taylor, K. Rastle, M. H. Davis, Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. *Psychol. Bull.* **139**, 766–791 (2013).
44. B. Gagl, *et al.*, The lexical categorization model: A computational model of left ventral occipitotemporal cortex activation in visual word recognition. *PLoS Comput. Biol.* **18**, 1–33 (2022).
45. R. Y. Zhang, K. N. Kay, Flexible top-down modulation in human ventral temporal cortex. *Neuroimage* **218**, 116964 (2020).
46. L. Cohen, S. Dehaene, F. Vinckier, A. Jobert, A. Montavont, Reading normal and degraded words: Contribution of the dorsal and ventral visual pathways. *Neuroimage* **40**, 353–366 (2008).
47. T. Z. Lauritzen, *et al.*, Top-down flow of visual spatial attention signals from parietal to occipital cortex. *J. Vis.* **9(13):18**, 1–14 (2009).
48. M. a Silver, D. Ress, D. J. Heeger, a Michael, D. J. H. Topographic, Topographic Maps of Visual Spatial Attention in Human Parietal Cortex. *J. Neurophysiol.* **94**, 1358–1371 (2005).
49. O. Ossmy, M. Ben-Shachar, R. Mukamel, Decoding letter position in word reading. *Cortex* **59**, 74–83 (2014).
50. B. Rapp, J. Purcell, A. E. Hillis, R. Capasso, G. Miceli, Neural bases of orthographic long-term memory and working memory in dysgraphia. *Brain* **139**, 588–604 (2016).

51. K. J. Forseth, *et al.*, A lexical semantic hub for heteromodal naming in middle fusiform gyrus. *Brain* **141**, 2112–2126 (2018).
52. X. Wang, A. Caramazza, M. V. Peelen, Z. Han, Y. Bi, Reading without speech sounds: VWFA and its connectivity in the congenitally deaf. *Cereb. Cortex* **25**, 2416–2426 (2015).
53. M. Makuuchi, A. D. Friederici, Hierarchical functional connectivity between the core language system and the working memory system. *Cortex* **49**, 2416–2423 (2013).
54. A. L. White, M. Rolfs, M. Carrasco, Stimulus competition mediates the joint effects of spatial and feature-based attention. *J. Vis.* **15**, 1–21 (2015).
55. T. Liu, Feature-based attention: effects and control. *Curr. Opin. Psychol.* **29**, 187–192 (2019).
56. J. Torgesen, C. Rashotte, R. Wagner, TOWRE-2: Test of Word Reading Efficiency, 2nd Ed. *Austin, TX Pro-Ed* (1999).
57. D. H. Brainard, The psychophysics toolbox. *Spat. Vis.* **10**, 443–446 (1997).
58. D. G. Pelli, The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
59. D. A. Medler, J. R. Binder, MCWord: An on-Line orthographic database of the English language. (2005).
60. O. Esteban, *et al.*, fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).
61. K. N. Kay, A. Rokem, J. Winawer, R. F. Dougherty, B. A. Wandell, GLMdenoise: A fast, automated technique for denoising task-based fMRI data. *Front. Neurosci.* **7**, 247 (2013).
62. E. J. Allen, *et al.*, A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nat. Neurosci.* **25**, 116–126 (2022).
63. J. S. Prince, *et al.*, GLMsingle : a toolbox for improving single-trial fMRI response estimates. **5000** (2022).
64. V. Vildavski, L. Lo Verde, G. Blumberg, J. Parsey, A. Norcia, PseudoSloan: A perimetric-complexity and area-controlled font for vision and reading research. *J. Vis.* **21**, 2857 (2021).
65. C. Vidal, A. Content, F. Chetail, BACS: The Brussels Artificial Character Sets for studies in cognitive psychology and neuroscience. *Behav. Res. Methods* **49**, 2093–2112 (2017).
66. A. Lochy, *et al.*, Selective visual representation of letters and words in the left ventral occipito-temporal cortex with intracerebral recordings. *Proc. Natl. Acad. Sci.* **115**, E7595–E7604 (2018).
67. N. C. Benson, *et al.*, The Human Connectome Project 7 Tesla retinotopy dataset: Description and population receptive field analysis. *J. Vis.* **18**, 1–22 (2018).
68. J. Larsson, D. J. Heeger, Two Retinotopic Visual Areas in Human Lateral Occipital

Cortex. *J. Neurosci.* **26**, 13128–13142 (2006).
69. L. Wang, R. E. B. Mruczek, M. J. Arcaro, S. Kastner, Probabilistic maps of visual topography in human cortex. *Cereb. Cortex* **25**, 3911–3931 (2015).
70. J. N. Rouder, P. L. Speckman, D. Sun, R. D. Morey, G. Iverson, Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* **16**, 225–237 (2009).