Voluntary effort to read reshapes stimulus selectivity in the visual word form area and increases its communication with the language network

Vassiki S. Chauhan[1*], Krystal C. McCook[1], Alex L. White[1]

[1]Department of Neuroscience & Behavior
Barnard College, Columbia University
76 Claremont Ave
New York, NY 10027
USA

[*]Correspondence: vchauhan@barnard.edu

Abstract

Reading depends on a brain region known as the “visual word form area” (VWFA). Although this region responds most strongly to written words, its selectivity is not absolute, and it exhibits top-down modulations related to task demands. We used fMRI to characterize the combination of factors that most strongly activate VWFA. We presented three stimulus types (English words, pseudowords, and visually matched false fonts) while the participant performed three different tasks: discriminating the lexicality of the character string, discriminating the color of the character string, and discriminating the color of the fixation mark while ignoring the character strings. We found a strong interaction between stimulus type and task: compared to the fixation task, the VWFA’s response to text was elevated in the lexical task but not in the character string color task. Thus, the enhanced response requires voluntary effort to read words and does not automatically occur when words are attended. Moreover, this effect of reading effort is stimulus-selective: false fonts presented during the lexical task evoked *smaller* responses compared to when they were ignored. These patterns were absent in other visual regions but were present in a putative Broca’s area in the left frontal cortex. Additionally, we found strengthened functional connectivity between Broca’s area and the VWFA during the lexical decision task. In sum, strong activation of the VWFA is not merely dependent on the presence of words nor on engagement in a linguistic task, but on the conjunction of both.

Significance Statement

Neuroscientists have long debated whether a region of ventral temporal cortex, often termed the visual word form area (VWFA), is specialized for identifying written words or whether it serves more general language functions. We used fMRI to characterize what drives VWFA activity: seeing words, attending visually to words, or explicitly trying to read words. We found the VWFA prefers letter strings over visually-matched shapes even when attention is withdrawn, but it is maximally engaged only when letter strings are present and the participant voluntarily tries to read them. These response patterns, combined with functional connectivity analyses, support a hybrid model: the VWFA has inherent visual selectivity for text, but its function is actively reshaped by language processing.

Introduction

Over the course of history, developments in human culture have shaped our environment and our brains adapted to meet these changing demands. A unique example is a brain region in ventral occipito-temporal cortex that becomes specialized for processing written words when an individual learns to read (McCandliss et al. 2003; Dehaene et al., 2010; Dehaene and Cohen 2011). This region is often termed the visual word form area (VWFA). While the VWFA's plasticity during literacy acquisition is well established (Dehaene et al., 2015), comparatively little is known about how its stimulus selectivity changes in the short term to meet task demands.

The nature of functional specialization in the VWFA is debated. Some researchers characterize it primarily as a visual region. It may serve as the "orthographic lexicon", detecting familiar letter strings and preparing them for subsequent linguistic comprehension (Glezer et al. 2009; Dehaene & Cohen 2011; Yeatman & White 2021). Others emphasize its top-down activation during language tasks and integration of linguistic information across modalities (Price & Devlin 2011; Qin et al. 2021; Dębska et al. 2023). We contribute to this lively discussion by examining how task demands shape the VWFA's selectivity for visual inputs as well as its communication with the canonical language network.

Like other areas around it, the VWFA has strong selectivity for visual stimulus properties. While it responds above baseline to many types of visual inputs (such as shapes and symbols), it demonstrates a clear preference for text (Muayqil et al. 2015; Vinckier et al. 2007; White et al. 2023). It is also sensitive to visual properties of written words, such as length (Woolnough et al. 2021), position (Rauschecker et al. 2012), contrast (Kay and Yeatman 2017) and other perceptual degradations (Cohen et al. 2008). Beyond purely visual properties, the VWFA is also sensitive to learned statistics of letter combinations within words (Woolnough et al. 2021; Gagl 2022), and how common a word is in the language (e.g, Kronbichler et al. 2004; White et al. 2019).

It is also well-established that the VWFA's activity is influenced by attentional allocation and cognitive task demands. For instance, words evoke larger responses when they fall within the focus of spatial attention than when they are ignored (White et al. 2019; Kay & Yeatman 2017). However, top-down modulations of VWFA activity differ from those in surrounding regions and are distinct from general forms of visual attention. Specifically, those modulations can both enhance the response to words and suppress the response to non-letter shapes, compared to when stimuli are ignored (White et al. 2023). Moreover, there is evidence for purely top-down language influences. For example, when participants perform a naming task, responses are elevated both for words and line drawings compared to during a visual size discrimination task (Mano et al. 2013). Some studies have even reported VWFA activations during language tasks without visual inputs (Reich et al. 2011; Striem-Amit et al. 2012). Indeed, it has been argued that VWFA's function is supported by connections with prefrontal cortex, auditory cortex and the attention network (Yeatman et al. 2012; Stevens et al. 2017; Chen et al. 2019; Li et al. 2020).

As both visual properties and a range of cognitive processes influence VWFA's activity, it is challenging to model the VWFA's functional specialization. In this study, we used functional magnetic resonance imaging (fMRI) to examine the *interaction* of stimulus features and task demands. We ask: what maximally drives activity in the VWFA? Simply viewing words,

attending to the visual properties of words, or explicitly engaging in a linguistic task? We answer these questions by fully crossing several types of visual stimuli with several types of tasks. We then relate the effects of task demands to communication with other parts of the language network as assessed via functional connectivity.

On each trial, a character string flashed on the screen for 150 ms, which was either a real English word, a pronounceable pseudoword, or a string of false font characters. The words and pseudowords were presented in two familiar fonts (Courier New and Sloan). For each of those real fonts we used a false font of unfamiliar characters with matched visual features (Vidal et al. 2017; Vildavski et al. 2022). Example stimuli are in **Figure 1B.** The three stimulus types were randomly intermingled during three different tasks. The *lexical decision task* was to report whether the stimulus was a real word or not. The *stimulus color task* was to report whether the character string was red or gray. The *fixation color task* was to report whether the fixation dot turned red while the stimulus flashed on the screen. The character strings were task-irrelevant during the fixation task, which provided our baseline measure of ignored stimulus responses. The lexical decision and stimulus color tasks required attending to the stimuli, but only the lexical decision task required recognizing the words. Both the stimulus color task and fixation color task required judgments of the same visual feature – color – but differed in whether the character strings were attended or not.

Our predictions are: if only stimulus features influence VWFA responses, we would expect higher responses to words than false fonts by the same degree across all tasks. In contrast, if only linguistic task demands influence VWFA responses, we would expect to see highest responses during the lexical decision task regardless of stimulus type. Other hypotheses involve interactions between the task demands and stimulus type; for instance, words may evoke stronger responses when task-relevant than when ignored, whereas non-linguistic stimuli get suppressed when task-relevant. Indeed, recent work has shown that the VWFA's response to non-letter shapes is suppressed when they are attended than ignored (White et al. 2023). However, those shapes were not perfectly matched in visual features to real letters, and when they were attended, it was during a simple shape judgement task. Thus, we do not yet know whether the suppression of VWFA responses is driven by performing non-linguistic tasks, or attending to non-linguistic stimuli in any task. The present study fills that gap by using a fully crossed design that measures the response to false fonts as well as words in three different tasks.

We examined BOLD response magnitudes in regions of interest (ROIs) defined precisely in each individual's brain, based on independent localizer scans (example stimuli in **Figure 1A**). That allows us to separately analyze VWFA sub-regions: VWFA-1 and VWFA-2 in posterior and anterior portions of the occipito-temporal sulcus (respectively), as well as a mid-fusiform text-selective region (MFUS-Text). Some previous studies suggest that these subregions play differing functional roles (Lerma-Usabiaga et al. 2018; White et al. 2019; Woolnough et al. 2021). We also defined ROIs within the language network with a separate localizer scan (Fedorenko et al. 2010). ROIs in an example participant's brain are shown in **Figure 1B.**

Results

Behavior

The participant's task on each trial was to push one of two buttons to report whether or not they saw a 'target': real words in the lexical decision task; red letters in the stimulus color task; and a brief change in the fixation dot's color from gray to red. In both color tasks, an adaptive staircase continuously adjusted the saturation of red color to each participant's detection threshold. The character string's position varied randomly across trials, either centered on the fixation mark, 3º left, or 3º right. The stimulus types were randomly intermingled, and the task conditions were blocked.

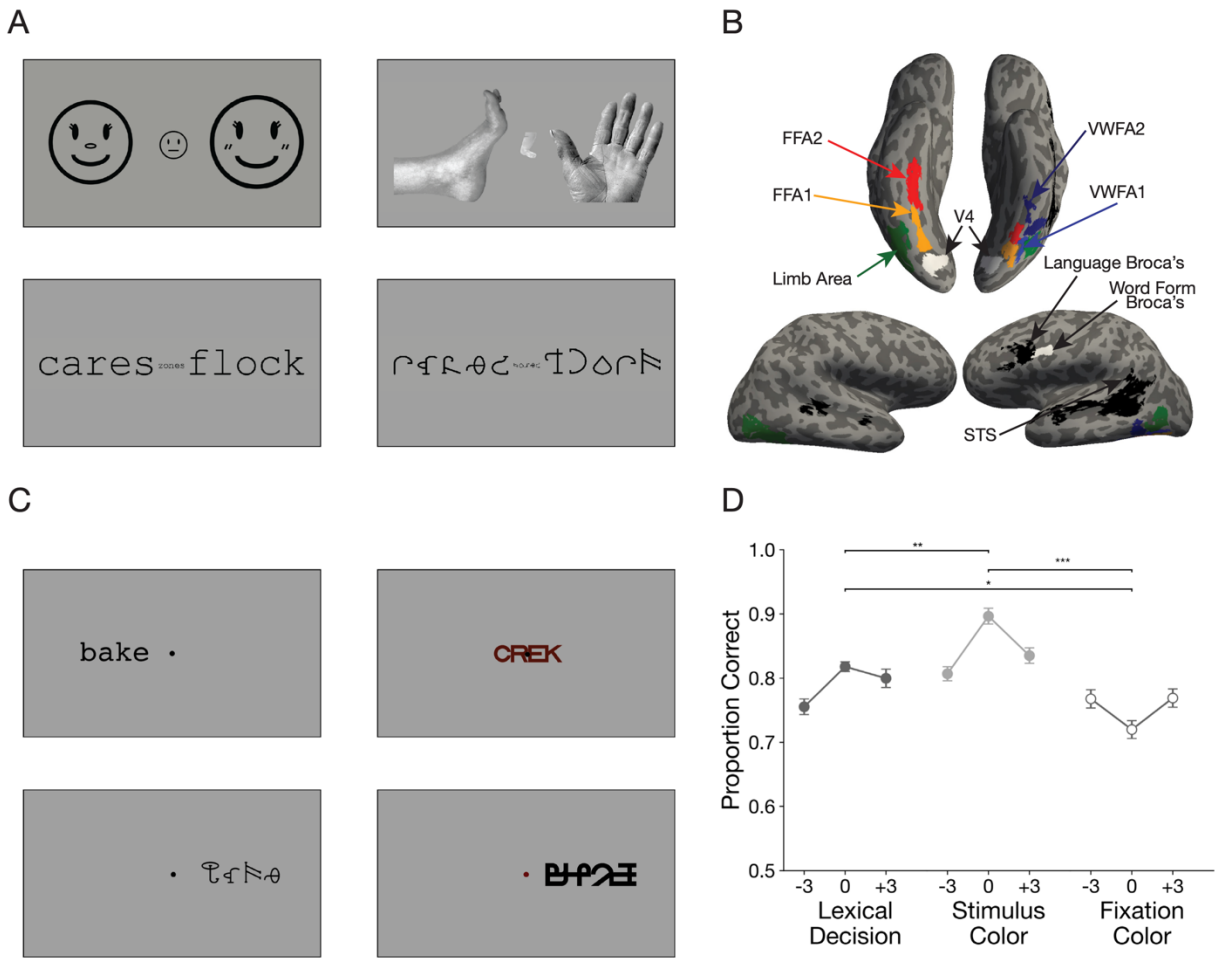


Figure 1: Experimental Paradigm. (A) Example stimuli in the localizer scan for category-selective regions in the ventral temporal cortex. Schematic faces, created by Yeoul Kwon from Noun Project under the CC BY 3.0 license, are used as placeholders for real faces from the experiment **(B)** Regions of interest (ROIs) in one example participant. Regions in color were defined using the localizer in panel A. Regions in black were defined from the language localizer scan, and regions in white were obtained from a previous study (White et al., 2023) **(C)** Example tasks and stimuli in the main experiment. **(D)** Mean accuracy in the three behavioral tasks plotted as a function of the character string's position in degrees of visual angle. Error bars are ±1 SEM, stars and horizontal lines depict statistically significant task effects (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.0001$)

Figure 1D plots mean task accuracy in each condition. Using linear mixed models with subjects as random effects, we analyzed task accuracy as a function of the task, stimulus type, and stimulus location. We found a main effect of all three experimental manipulations (Task: $F(2,33)=37.77$, $p<0.001$, Stimulus Type: $F(2,23)=109.84$, $p<0.001$, Stimulus Location: $F(2,46)=6.12$, $p=0.004$), and two-way interactions between task and stimulus type ($F(4,375)=117.09$, $p<0.001$), as well as between task and location ($F(4,375)=9.8$, $p<0.001$). Accuracy was highest in the stimulus color task, and lowest in the fixation color task (all pairwise comparisons $p<0.03$). In the lexical decision and stimulus color tasks, accuracy was higher when the stimuli were presented at fixation (0º) than 3º in the parafovea, but accuracy in the fixation color task was relatively impaired by characters appearing at fixation (**Figure 1D**). Post-hoc tests (corrected for false discovery rate) revealed better lexical decision performance when real words were presented in the right compared to the left hemifield ($t=2.31$, $p=0.03$), consistent with prior results (Huey, 1900; Mishkin & Forgays, 1952) .

The lexical decision task enhances the VWFA's response to text and suppresses its response to false fonts

As we expected, we found overall higher responses in the left VWFAs (VWFA- and VWFA-2) to words and pseudowords (collectively referred to as "text") than to false fonts across all tasks (**Figure 2A and 2B**). The main effect of stimulus, collapsing across tasks, was significant in both areas ($F(1,14)=17.68$, $p < 0.001$; $F(1,15)=35.73$, $p < 0.001$, respectively). We also found an interaction between task type and stimulus category in both areas (VWFA-1: $F(2,101)=27.14$, $p < 0.001$; VWFA-2: $F(2,108)= 28.22$, $p < 0.001$). The preference for text over false fonts was significant even during the fixation color task, when the stimuli were ignored (VWFA-1: $t(22)=2.06$, $p = 0.06$, VWFA-2: $t(29)=3.03$, $p = 0.008$; averaging over VWFA-1 and -2: $t(25)=2.71$, $p = 0.015$).The statistics for pairwise comparisons between tasks for each stimulus type are reported in **Table 1**. Here, we highlight two important results.

First, compared to when the stimuli were ignored during the fixation color task, the VWFA's response to text (real words and pseudowords) was enhanced during the lexical decision task, but not during the stimulus color task. The mean differences from the fixation task baseline are shown in **Figure 2C** and **2D**, where dark green bars show the large enhancement in the lexical task and light green bars show the minimal effect during the stimulus color task. Word recognition is sometimes described as an "automatic" process (Stroop, 1935; Price et al. 1996; Brown et al. 2002; Ruthruff et al. 2008; Pattamadilak et al. 2017; Joo et al. 2021). That would predict that directing attention to written words, even in order to simply extract visual properties, would fully engage the VWFA. In contrast, our results demonstrate that engagement in an explicitly lexical task is necessary to drive responses higher than the baseline visual response to ignored words.

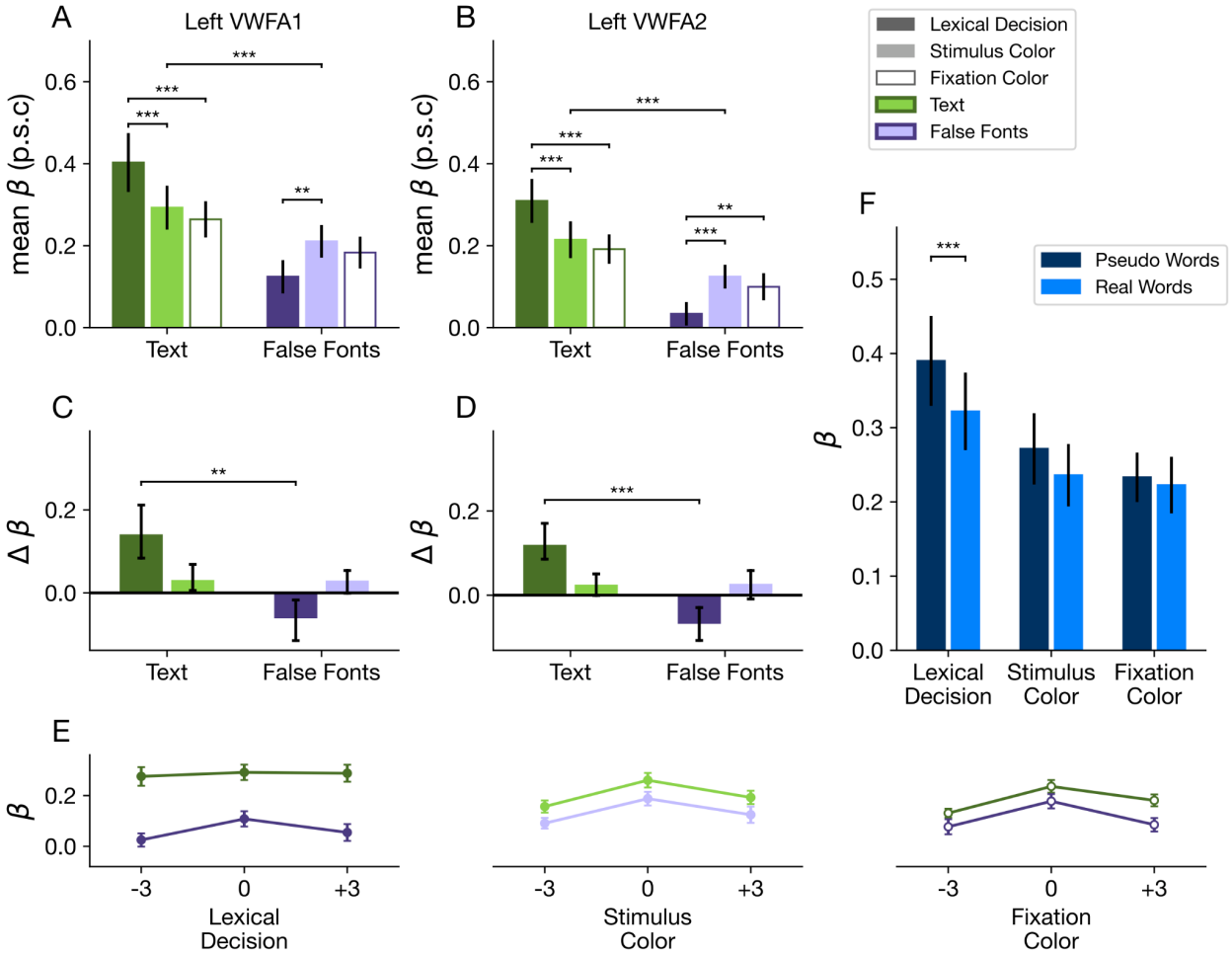


Figure 2: The VWFA's activity reflects an interaction of stimulus features and task demands. (A, B) Mean percent signal change (p.s.c.) as a function of stimulus type and task condition in VWFA-1 and VWFA-2, respectively. "Text" includes both real words and pseudowords (which are separately analyzed in panel F). Error bars represent ±1 SEM. Horizontal bars and stars indicate statistical significance for the main effect of stimuli and pairwise t-tests between tasks. **(C, D)** Task effects computed as the mean difference in percent signal change from the fixation task baseline. Error bars indicate 95% confidence intervals. Horizontal bars and stars indicate a significant difference in the effect of engaging in the lexical task (compared to fixation task) on responses to text vs. false font. (The effect of the stimulus color task did not differ across stimulus types). **(E)** Responses as a function of stimulus position, separately for each task and stimulus type, collapsed across VWFA-1 and VWFA-2. Error bars are ±1 SEM. **(F)** Responses to words vs. pseudowords in the three tasks, collapsed across VWFA-1 and VWFA-2. The lexicality effect is only observable during the lexical decision task. Error bars are ±1 SEM.

Second, for false fonts, we observed the *opposite* task effect: responses during the lexical decision task were *suppressed* compared to the other two tasks. See the dark purple bars in **Figure 2C** and **2D**. This result suggests that VWFA responses are suppressed compared to the baseline while the brain is engaged in linguistic processing but is presented with non-preferred, albeit visually matched, stimuli. It is worth noting that attending to false fonts to judge their color had no measurable effect compared to ignoring them (light purple bars).

ROI	Stimulus	Task	Mean ΔPSC	SEM	95% CI	t-value	p-value
VWFA1	Text	LD - FC	0.15	0.02	0.09, 0.2	6.05	<0.001
		LD - SC	0.12	0.02	0.07, 0.17	5.77	<0.001
		SC - FC	0.03	0.02	0.02, 0.08	1.47	0.147
	False Fonts	LD - FC	-0.06	0.03	-0.14, 0.01	2.03	0.07
		LD - SC	-0.09	0.03	-0.16, -0.02	3.31	0.004
		SC - FC	0.03	0.03	-0.04, 0.1	1.04	0.3
VWFA2	Text	LD - FC	0.12	0.02	0.07, 0.16	6.5	<0.001
		LD - SC	0.1	0.02	0.05, 0.14	5.67	<0.001
		SC - FC	0.02	0.02	0.01, 0.07	1.35	0.18
	False Fonts	LD - FC	-0.07	0.02	-0.12, 0.01	2.71	0.01
		LD - SC	-0.09	0.02	-0.15, -0.03	3.88	0.001
		SC - FC	0.03	0.02	-0.03, 0.08	1.05	0.3
Word Form Broca's	Text	LD - FC	0.11	0.02	0.07, 0.15	6.66	<0.001
		LD - SC	0.09	0.02	0.06, 0.13	6.05	<0.001
		SC - FC	0.01	0.02	-0.02, 0.05	0.91	0.37
	False Fonts	LD - FC	-0.1	0.02	-0.15, 0.04	4.27	<0.001
		LD - SC	-0.11	0.02	-0.17, -0.06	5.27	<0.001
		SC - FC	0.02	0.02	-0.03, 0.07	0.9	0.37
Language Broca's	Text	LD - FC	0.07	0.01	0.04, 0.10	5.73	<0.001
		LD - SC	0.07	0.01	0.04, 0.10	6.02	<0.001
		SC - FC	0	0.01	-0.03, 0.03	0.03	0.98
	False Fonts	LD - FC	-0.05	0.02	-0.09, 0.004	2.66	0.01
		LD - SC	-0.06	0.02	-0.10, -0.02	3.47	0.002
		SC - FC	0.01	0.02	-0.03, 0.05	0.75	0.46

Table 1: Mean difference in percent signal change across pairs of tasks, separately for text and false font stimuli, in the two VWFAs and Broca's areas. For each ROI, the six p-values are corrected for false discovery rate. Abbreviations: LD = lexical decision task; FC = fixation color task; SC = stimulus color task.

One potential concern is that differences in task difficulty could explain the different BOLD response magnitudes. However, task accuracy (**Figure 1D**) suggests that all tasks were moderately challenging (mean accuracy between 70% and 90% correct). Moreover, the stimulus color task was the easiest and the fixation color task was the hardest. The VWFA results, in contrast, show indistinguishable levels of activity during these two tasks for both text and false fonts. The biggest differences are driven by the lexical decision task, which had intermediate accuracy, and in opposite directions for different types of stimuli. Thus, an account based only on task effort or difficulty cannot explain our BOLD results.

These two left VWFA subregions, VWFA-1 and VWFA-2, showed similar effects in all our analyses. (A third region, MFUS-Text, is shown in **Figure 3F**). Therefore, for the remainder of the analyses, we collapse across left VWFA-1 and VWFA-2 (referred to as "VWFAs").

Figure 2E shows that VWFAs are sensitive to the visual field position of the words (Rauschecker et al. 2012, White et al. 2019, White et al. 2023). In general, the VWFAs prefer stimuli at fixation (0º) over stimuli presented 3º to the left or right (Main effect of stimulus location: $F(2,24)=29.7$, $p < 0.001$). However, that spatial sensitivity is greatly reduced for text presented during the lexical decision task (green line in **Figure 2E**). During that condition, there were no significant difference in pairwise comparisons of stimulus positions (all $t<1$, $p>0.7$). Thus, voluntary effort to read enhances the response especially when words are in the parafovea where legibility is reduced (Cohen et al. 2008; White et al. 2023).

Lastly, **Figure 2F** shows that the VWFAs respond differently to real words and pseudowords, but only during the lexical decision task (stimulus by task interaction: $F(2,45)=3.5$, $p=0.038$). Consistent with prior studies, during the lexical decision task, pseudowords evoked larger responses ($t(42)=3.7$, $p = 0.001$), perhaps because processing them takes more time than processing familiar words (Gagl et al. 2022; White et al 2023). However, that "lexicality effect" is absent when the stimuli are ignored, suggesting a key role for top-down input for the VWFA to be sensitive to higher-level lexical features of the stimuli.

The task and stimulus type interactions are absent from other occipital and temporal regions

In addition to VWFA-1 and VWFA-2, we analyzed responses in six other regions of interest, in the left occipital and temporal lobes. See **Figure 3** (right hemisphere regions are shown in Supplementary Figure 1). Within the ventral occipito-temporal cortex, we analyzed V4, a posterior and anterior FFA-1 and FFA-2, a limb-selective region (OTS-Limbs) and another text-selective region within the fusiform gyrus (MFUS-Text). Additionally, we analyzed responses in a large swath of the superior temporal sulcus (STS) that was activated in the language localizer. We did not find a significant effect of task, stimulus type, or an interaction between task and stimuli within any of these ROIs. Thus, the pattern of activity we found in the VWFA is unique among the visual regions we analyzed and also absent in the STS. Statistics for all ROIs are reported in the Supplementary Table 2. After correcting for multiple comparisons, the only marginally significant effect was observed in the left MFUS-Text (**Figure 3F**), where responses to text were higher in the lexical task than the other two tasks ($p=0.07$).

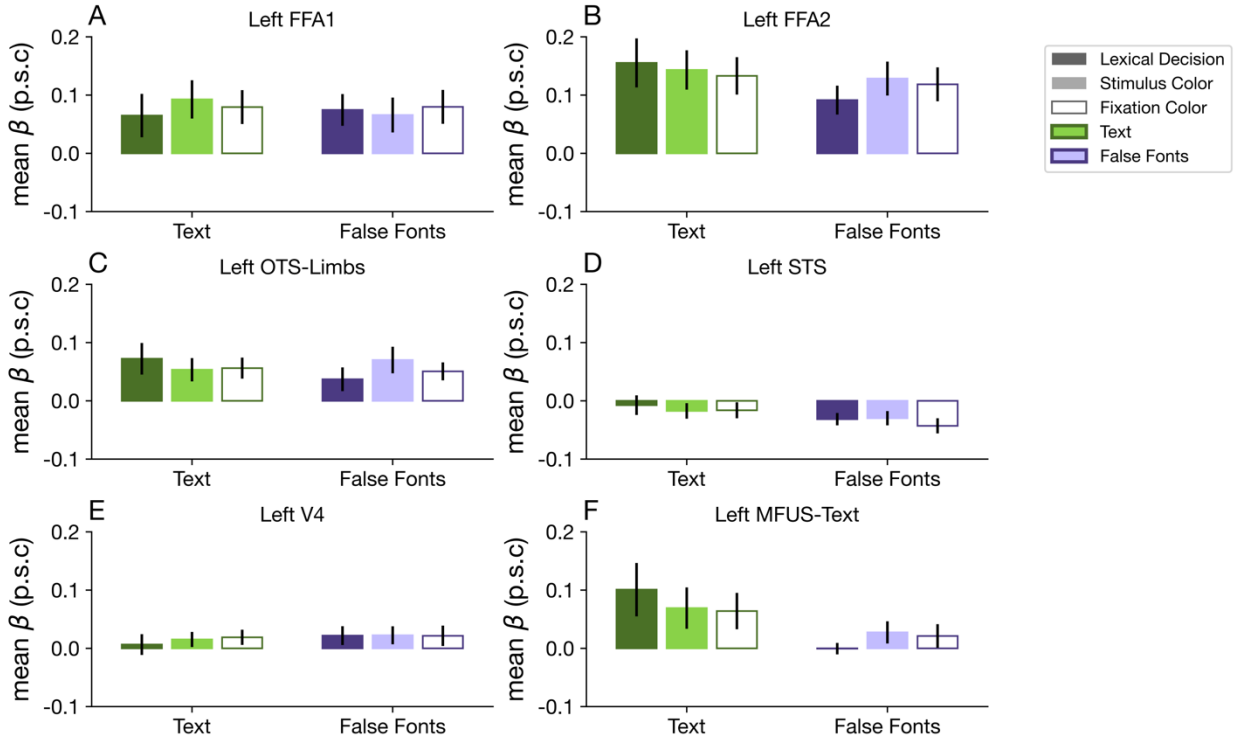


Figure 3: Task and stimulus effects are absent in other regions of left occipital and temporal lobes. Each panel plots mean BOLD responses in one ROI, formatted as in Figure 2A.

Broca's area shows strong task-dependent enhancement and suppression

We also localized two language processing regions in the left frontal cortex. The first we refer to as "*Language Broca's*," because it was defined in each individual by our independent language localizer (see Methods). It was located in the inferior frontal sulcus (sometimes extending into the inferior frontal gyrus). The second region we refer to as "*Word Form Broca's*" and it was defined (in an average brain) from a previous study contrasting the response to written words and non-letter shape strings (White et al, 2023). Its location was slightly posterior to Language Broca's, in the precentral sulcus. A similar contrast on our main experiment, contrasting text stimuli with false fonts, produced an overlapping region. We used the Word Form Broca's from the independent dataset in the previous study to avoid using the same data for defining an ROI and analyzing mean responses. The locations of the two Broca's regions relative to each other (in the average brain) are included in Supplementary Figure 2, and also depicted in **Figure 1B** in an example subject.

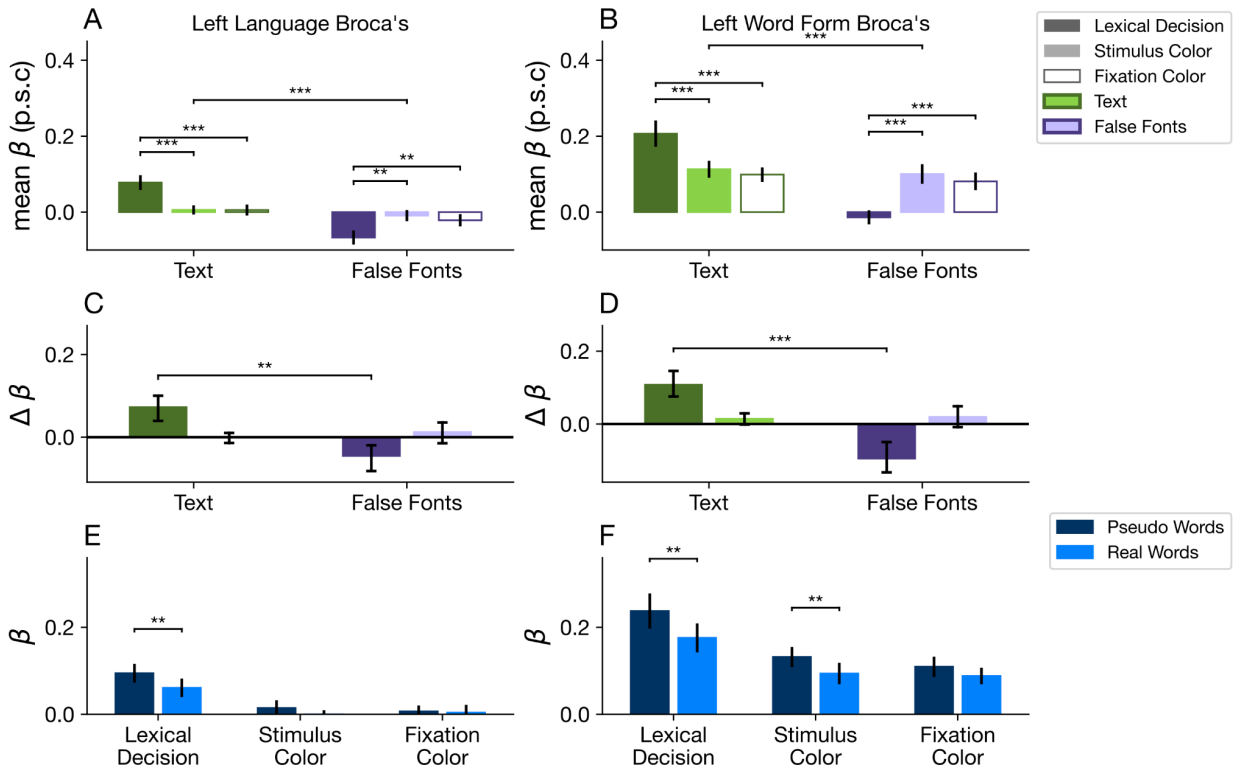


Figure 4: Task and stimulus effects in two left frontal regions. (**A, B**) Mean percent signal change as a function of stimulus type and task in two frontal regions in the left hemisphere. Format as in Figure 2A. **(C, D)** Difference in mean percent signal change from fixation task baseline. Format as in Figure 2C; error bars indicate 95% confidence intervals. **(E, F)** Responses to real words vs. pseudowords in each task (format as in Fig 2E).

Figure 4 shows that in both these Broca's area sub-regions, we found interactions between stimulus type and task that went in the same direction as in the VWFAs. However, response magnitudes in the more anterior Language Broca's area were different: it did not respond positively to stimuli overall (**Figure 4A**). Its response differed from baseline only during the lexical decision task, but in opposite directions for text vs. false fonts: a strong increase for text, and a decrease below baseline for false fonts (**Figure 4C**). It also showed sensitivity to lexicality only during the lexical task (**Figure 4E**). The response magnitudes in Word Form Broca's were more similar to the VWFAs, with generally positive responses above baseline (**Figure 4B**). It showed the same task effects: a selective enhancement for text and suppression for false fonts (down to 0) during the lexical decision task (**Figure 4D**). It also responded more strongly to pseudowords, except during the fixation task (**Figure 4F**).

One interpretation of these data is that the Language Broca's area functions as a *control region* that applies positive or negative modulations to other parts of the network. This region is engaged only during the lexical task. If letter strings are presented during the lexical task, its activity rises and it in turn excites other regions involved recognizing words: the VWFAs as well as the more posterior Word Form Broca's. If it detects false fonts, its activity decreases below baseline and the other regions are suppressed as well. To further understand how activity in these areas co-vary, we analyzed functional connectivity between them.

Modulations of VWFA activity are accompanied by task-dependent functional connectivity with the language network

We performed the functional connectivity analysis by correlating the *fluctuations* in trial-evoked responses, after subtracting out the mean response in each condition. For each surface vertex, we averaged the responses for each unique combination of task, stimulus type (independently for real words, pseudowords and false fonts) and stimulus location. We then removed this mean response from every single trial's beta response that corresponded to the same unique condition, and performed the subsequent analysis with these demeaned *residual* responses. Specifically, we computed correlations coefficients in these residual responses across regions, separately for each task and stimulus condition.

Figure 5A shows a whole-brain map of correlations between each surface vertex and the left Language Broca's area as the seed region. This represents functional connectivity just during the lexical task (including trials with all stimulus types). We observed high positive correlations between Language Broca's area and much of the left superior temporal sulcus (STS) and intraparietal sulcus (IPS), consistent with prior results (White et al. 2023). Functional connectivity was also high with the occipito-temporal sulcus, where the VWFAs are, in both hemispheres. Interestingly, there was also a hot spot of connectivity with a homologous 'Broca's area' in the right frontal cortex. In Supplementary Figure 3-5, we report analogous analyses with VWFA-1, VWFA-2 and Word Form Broca's as the seed regions.

In addition to whole brain functional connectivity, we also computed mean correlations between the left VWFAs (collapsing VWFA-1 and -2) and left language Broca's area, separately for each task and stimulus condition. The results are in **Figure 5B.** We found a main effect of stimulus type ($F(1,15)=18.53$, $p < 0.001$) and an interaction between task and stimulus type ($F(2,108)=25.90$, $p < 0.001$). We observed highest correlations between the two regions during the lexical decision task, for both types of stimuli, and lowest correlations during the fixation task. Statistical results are reported in **Table 2**. The functional connectivity patterns were similar for VWFA-1 and VWFA-2, analyzed separately (Supplementary Figure 6).

It is especially noteworthy that when false fonts were presented, functional connectivity between the VWFAs and Broca's was also *increased* during the lexical task compared to the fixation task. This goes in the opposite direction as the mean overall response magnitudes, which were *decreased* for false fonts during the lexical task (**Figure 2A-B)**. This suggests that the suppression of response magnitudes to false fonts in the VWFAs could be caused by communication with Language Broca's area.

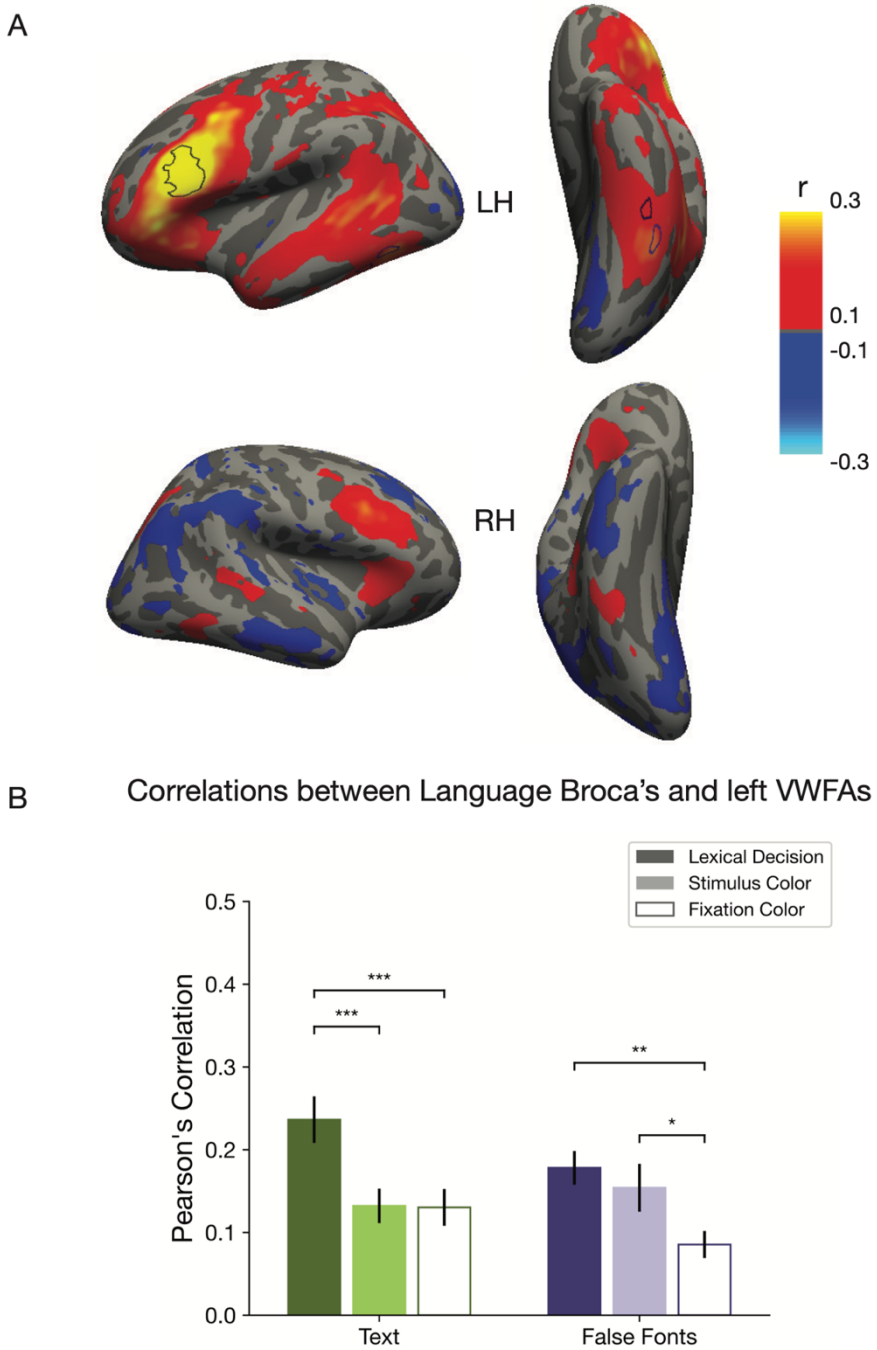


Figure 5: Functional connectivity shows task-dependent communication between Broca's area and the VWFAs. (A) Whole-brain connectivity with the seed region in left language Broca's area (black outline) during the lexical decision task for all stimulus types. The color map represents Pearson's correlation coefficients (minimum $r>0.11$, $q<0.05$, FDR corrected). Correlations were highest with a homologous Broca's region in the right hemisphere, the left superior temporal sulcus, intraparietal sulcus, and occipito-temporal sulcus in both hemispheres. The two black outlines on the left ventral surface represent the peaks of the contrast between words and all other categories in the category localizer experiment, in the average brain. Thus, they roughly correspond to the most likely locations of VWFA-1 and VWFA-2. (**B**) Average correlation coefficients between left language Broca's area and the left VWFAs (collapsing VWFA-1 and VWFA-2). The correlations were highest during the lexical decision

task for both types of stimuli. Error bars are +/- 1 SEM. Horizontal lines and stars indicate statistical significance. (*=p<0.05, **=p<0.01, ***=p<0.001)

Stimulus	Task	Mean Δ r	SEM	95% Confidence Interval	t-value	p-value
Text	LD - FC	0.12	0.02	0.06, 0.16	5.68	< 0.001
	LD - SC	0.1	0.02	0.05, 0.16	4.89	< 0.001
	SC - FC	0	0.02	-0.05, 0.06	0.08	0.94
False Fonts	LD - FC	0.09	0.03	0.03, 0.16	3.61	0.002
	LD - SC	0.02	0.03	-0.04, 0.09	0.87	0.39
	SC - FC	0.07	0.03	0.00, 0.14	2.41	0.029

Table 2: Statistics on how functional connectivity between the language Broca's Area and the left VWFAs (collapsing over VWFA-1 and VWFA-2) differed across pairs of tasks. This is calculated separately for trials with text stimuli (both words and pseudowords) and false fonts. "Mean Δ r" is the mean difference in the correlation coefficient r across the two tasks listed in each row. P-values are FDR corrected for multiple comparisons. Abbreviations: LD = lexical decision task; FC = fixation color task; SC = stimulus color task.

Discussion

The VWFA is maximally engaged when the participant views letter strings and tries to read them

We investigated the combination of stimulus features and task demands that maximally engage the VWFA. The results show that the VWFA's activity can neither be accounted for by visual stimulus selectivity alone, nor by top-down modulation alone. The strongest activation requires specific stimuli (strings of familiar letters) and specific task demands (explicitly trying to read words). This interaction between bottom-up and top-down factors is unique to the VWFA among visually-driven areas, and yields some surprising patterns. Primarily, engaging in the lexical task increased the VWFA's BOLD response to words and pronounceable pseudowords (with an even greater increase for pseudowords), but *suppressed* the response to visually-matched novel characters (false fonts). Attending to either type of character string in order to judge its color had little to no impact on the response, compared to when the stimuli were ignored. These stimulus-specific and task-dependent modulations are likely due to interactive feedback from regions involved in processing language, as we discuss in more detail below.

Possible explanations for what drives activity within the VWFA

The VWFA's functional specialization has been heavily debated (Dehaene & Cohen 2011; Price & Devlin 2011; Rossion & Lochy 2022; Dębska et al. 2023). By fully crossing task demands and stimulus types, our results effectively rule out several explanations for VWFA activity. Here we consider four such hypotheses.

Word-likeness of visual stimuli

One hypothesis is that the VWFA's activity is explained by bottom-up tuning for visual stimuli that are the most like known words. That predicts, in our study, that responses should be weakest for false fonts, intermediate for pseudowords, and highest for real words, in all three tasks. Our data only partly match that prediction. The VWFA did prefer real letter strings over false fonts with visually-matched features, no matter the task. But that selectivity was greatly magnified during the lexical task (compared to both other tasks), in part because false fonts were actively suppressed. More importantly, we observed highest responses for *pseudowords*, consistent with prior findings (e.g., Kronbichler et al. 2004), but that elevation of responses to pseudowords over real words only occurred during the lexical decision task (**Figure 2F**). These results suggest that the VWFA's sensitivity to *lexical* attributes of letter strings is contingent on top-down cognitive influences during linguistic tasks.

Top-down boost for attended stimuli

Throughout visual cortex, task-relevant stimuli evoke larger responses than task-irrelevant or ignored stimuli. These modulations could be described as additive or multiplicative modulations of the stimulus-driven response, with the magnitude of the modulation depending on the match between the tuning properties of neurons and which locations or features are attended (e.g., Reynolds & Heeger, 2009). One hypothesis is that such attentional mechanisms boost responses to task-relevant stimuli in the VWFA, but only when the stimuli are text. For our study, that hypothesis would predict a higher response to words and pseudowords when they are task-relevant to than when they are ignored, and no modulation by attention for false fonts. This prediction is violated in two key ways, evidenced by the interaction effect in the stimulus types and task demands. First, there was no difference in the response to text in the stimulus color task, when the letter strings were task-relevant, than in the fixation color task, when they were ignored (light green bars in **Figure 2A-D**). Second, the responses to false fonts, which are non-preferred stimuli for the VWFA, were *lower* when they were task-relevant during the lexical decision task than when they were ignored during the fixation color task (**Figure 2A-D**, dark purple bars). Together, these results show that the top-down modulations in the VWFA *cannot* easily be explained by visual attention, or modulations based on the task-relevance of the stimuli.

This conclusion is consistent with a recent study by White et al. (2023), but with several key differences. They also found enhancement of the VWFA response to attended words, but suppression of the response to attended strings of shapes (squares and circles). But the stimulus type and task co-varied in blocks, so the shape strings were never presented during the lexical task and vice versa (as we did here). Moreover, in that study, the participant's task for the shape strings was to report the location of a gap in one of those shapes. That result could mean that the VWFA is suppressed during engagement in a non-linguistic task with non-letter shapes. If that were generally true, then in the present study we should have found suppression of the response to false fonts during the stimulus color task. In contrast, we only found that suppression during the lexical decision task. One way to reconcile these findings is that the VWFA is suppressed when the task requires processing the shapes of letter-like but unfamiliar shapes. That was true both in the gap task in White et al. 2023, and in the lexical decision task reported above.

Automaticity of word recognition

Automaticity of word recognition implies that as long as a word is legible and some attentional resources are available, all lexical and semantic information about it is fully processed (Carr 1992; Brown et al. 2002). Consider the Stroop effect, in which the task-irrelevant meaning of a word interferes with judgements of the color of the ink (Stroop 1935). In its simplest form, such a theory would predict equivalent responses to written words in the lexical decision and stimulus color tasks. In contrast, we found that engaging in the stimulus color task had a negligible effect compared to the fixation task. Therefore, within the VWFA, seeing a word is not the same as reading it. Full-fledged VWFA activation requires voluntary effort to read the words.

One potential criticism of the comparison between the lexical and stimulus color tasks is that the stimulus color task could be accomplished by attending to and processing only one letter, whereas all letters must be processed together for the lexical task. That may be true, but two factors make it an unlikely strategy. First, the position of the whole character string was unpredictable and it was presented for just 150 ms, so the participant could not focus spatial attention on the location of just one letter. Second, the saturation of the red color was set to each participant's detection threshold, so the optimal strategy would be to integrate the color signal across all the letters to best detect it.

Top-down linguistic effort

This hypothesis pertains to the role of the VWFA as an orthographic lexicon that links visual and linguistic networks of the brain (Yeatman & White 2021; Caffarra et al. 2021). The 'linguistic effort' hypothesis proposes that when the participant engages in the lexical task, the VWFA's response to each stimulus gets increased by an amount proportional to how difficult it is to match the item to a known word or reject it as a non-word. That can account for the elevated response to pseudowords over real words during the lexical task: while familiar real words can be quickly matched to items in the mental lexicon, plausible pseudowords require an extended search through the lexicon before being rejected, and that results in a greater BOLD response (Gagl et al. 2022). However, this hypothesis cannot account for the suppression of false fonts during the lexical task. False fonts elicit a modest response during the fixation task, when *no* effort is devoted to processing any of the character strings. During the lexical task, the false font strings were task-relevant, and behavioral data shows that they were relatively easy for subjects to quickly reject as non-targets. This hypothesis would then predict that false font strings evoke roughly the same response in the lexical task as during the fixation task, or perhaps slightly higher, contrary to the active suppression that we observed.

Thus, none of the four hypotheses just described fully account for the response magnitudes we observed in the VWFA. Before proposing our own explanation, we must consider the patterns of functional correlation that shed further light on the VWFA's functional role.

The VWFA's activity is shaped by communication with the language network

We found robust functional connectivity between the VWFA and several regions in the language network (**Figure 5A** and Supplementary Figures 3-4). Moreover, we discovered higher correlations between the VWFA and Broca's area during the lexical decision task than the other two tasks, across all stimulus types (**Figure 5B**). This task-dependent modulation of functional connectivity was unique to the VWFA within ventral temporal cortex, similar to a previous report (White et al. 2023). Importantly, this reflects correlated activity both when responses are enhanced within the VWFA, as well as when responses are suppressed—as in the case of false fonts during the lexical decision task. The seed region used for this analysis, which we called Language Broca's area, was not modulated by stimulus type except during the lexical decision task, and in fact reveals negative percent signal change during the presentation of false fonts (**Figure 4A** and **4C**). This suggests that suppressed VWFA response to false fonts during the lexical decision task is due to top-down feedback from Broca's area.

Whole-brain functional connectivity with Broca's area revealed vast swaths of high correlations across different areas along the cortical surface. These maps recovered areas that are recruited by the language localizer (Mahowald & Fedorenko 2016), such as the superior temporal sulcus (STS), as well as regions that participate in the allocation of visual attention, such as the inferior parietal sulcus (IPS). Correlations with the STS showed no modulation by task. Additionally, we found homologous regions to Broca's in the right hemisphere, as well as a patch in the right ventral temporal cortex homologous to the anterior parts of the left VWFA.

Functional segregation in frontal language regions

We observed interesting differences between the activity patterns of two frontal language regions. The first region, which we call "Language Broca's", was located more anteriorly in the inferior frontal sulcus and obtained using the language localizer scans (Mahowald & Fedorenko 2016). The second region, "Word Form Broca's", in the precentral sulcus, was obtained from a previous study (White et al. 2023), using the contrast between text and non-letter shapes. While language Broca's did not seem to be engaged by any stimuli except during the lexical decision task, the word form Broca's demonstrated a similar response profile to the VWFA. Other researchers have speculated about functional segregation within the IFG along the dorsal-ventral axis, as well as the anterior-posterior axis (Hagoort 2014). Our results suggest that the anterior patch of the frontal cortex we analyzed (Language Broca's) is more involved in task demands and executive control than actually processing the stimuli.

Our proposal for the functional role of the VWFA

The VWFA's functional role is widely contested. Even the label ascribed to this area is debated. The term "visual word form area" (VWFA) is preferred by researchers who conceptualize its function as identifying a particular type of visual input and then communicating an abstract representation to the language network (McCandliss et al. 2003; Dehaene and Cohen 2011). Others propose that the VWFA serves a more general and multi-modal language function

(Debska et al, 2023) prefer terms such as the basal temporal language area (BTLA) (Price and Devlin 2003).

We propose a hybrid model. Our results support the notion that the VWFA is primarily a visual region devoted to matching letter strings to known items in an orthographic lexicon (Glezer et al. 2009; Dehaene and Cohen 2011; Woolnough et al. 2021). Its bottom-up responses are selective for familiar letters over stimuli with carefully matched visual features even when attention is withdrawn. However, the VWFA is not automatically engaged by the presence of familiar letter strings. Its selectivity for text is greatly enhanced during linguistic tasks, in part due to active suppression of other types of visual stimuli. Also, its sensitivity to higher-order lexical features of letter strings depends critically on the participant engaging in an explicitly linguistic task. These aspects of its function are dependent on communication with other parts of the language network, as our functional connectivity analyses revealed.

Limitations and future directions

Methods that offer greater temporal resolution, such as magnetoencephalogaphy, will be necessary to reveal *when* the task effects arise in frontal regions and the VWFA, relative to the onset of a word. Our results demonstrate a high degree of spatial specificity of cognitive modulations in the VWFA, which did not occur in nearby category-selective regions. Nonetheless, there may be analogous cognitive effects in those other visual regions during tasks optimized for them, such as face recognition in the FFA. Indeed, a region near to Broca's area in the right hemisphere is active during face recognition tasks (Guntupalli et al. 2017; Visconti di Oleggio Castello et al. 2017), and perhaps it modulates the FFA in a way similar to the linguistic influences we observed in the VWFA.

We found similar activity patterns in VWFA-1 and VWFA-2. While previous studies have reported the same (White et al. 2023), others have emphasized the importance of functional subdivisions within text selective patches in the ventral surface of the brain (Lerma-Usabiaga et al., 2018). We did find a patch within the left fusiform gyrus (text-MFS) that clearly preferred words over false fonts but showed weaker task effects than the VWFA-1 and -2 in the occipito-temporal sulcus. Prior work has emphasized the importance of this region in reading (Woolnough et al. 2021). However, we were only able to locate this area in half our sample, and the fMRI signal of anterior temporal cortex is impoverished by susceptibility artifacts. More work is needed to determine how these regions divide the labor of reading, especially in light of their functional and anatomical connectivity with regions within the language network (Lerma-Usabiaga et al. 2018).

Conclusions

We conclude that activity in the VWFA is determined by an interaction between visual stimulus selectivity and top-down input from frontal language regions. The VWFA is maximally activated when letter strings are presented *and* the observer is voluntarily looking for known words. It is not sufficient for words to be presented or even attended visually. Moreover, engagement in a lexical task does not always enhance the VWFA's response; in fact, it suppresses the response to unfamiliar characters. These findings open the door to future work on how the VWFA

communicates with other language areas during more naturalistic reading tasks, as well as comparisons to other specialized parts of the visual system.

Methods

Participants

We recruited 17 participants (Age range 19 to 38, 21.12 ± 4.44, 4 self-identified as male, 1 left-handed) from the Barnard College and Columbia University student body. We chose the sample size on the basis of a power analysis on an interaction effect observed in a recent study using a similar experimental paradigm (White et al., 2023). The study was approved by the Internal Review Board at Barnard College, Columbia University. All participants provided written informed consent, acquired digitally, prior to participating in the study and were monetarily compensated for their participation. All participants were screened for MR compatibility prior to recruitment, and we ensured that all participants had learned English before the age of five.

To ensure high quality of data, we used the following criteria for excluding functional runs and participants. If the participant moved by a distance greater than 2 voxels (4 mm) within a single run, that run was excluded from analysis. Additionally, if the participant responded in less than 50% of the trials in the main experiment, that run was removed. Finally, if half the runs met any of these criteria for a single subject, that entire participant was dropped from the dataset. Using these constraints, the analysis reported here is based on data from 16 subjects.

Equipment

We collected MRI data at Zuckerman Institute, Columbia University on a 3T Siemens Prisma scanner using a 64 channel head coil. The experiment involved three MRI sessions per participant, with an additional hour of behavioral training and screening for MR compatibility. In each MR session, we acquired a T1 weighted structural scan, with voxels measuring 1 mm isometrically. We acquired functional data with a T2* echo planar imaging sequences with multiband echo sequencing (SMS3) for whole brain coverage. The TR was 1.5s, TE was 30 ms and the flip angle was 62°. Voxel size of 2 mm isotropic.

Stimuli were projected through an LCD screen that the participants viewed through a mirror that the images were reflected on to, with a viewing distance of 142 cm. The display had a resolution of 1920 by 1080 pixels, and a refresh rate of 60 Hz. We presented the stimuli using custom code written in MATLAB using Psychophysics Toolbox. Throughout the scan, we recorded monocular gaze position using SR Research Eyelink 1000 tracker. Participants believed that their eyes were being tracked even when the calibration or validation procedure failed. Participants responded via button press using three buttons on an MR safe response pad, which they held in their right hand.

Main Experiment

Stimuli

Figure 1A shows example stimuli. A small dark fixation dot at screen center was present throughout all trials (0.11 degrees of visual angle, or dva, diameter) against a white background (95% of screen maximum). The stimuli were four element long character strings from three different categories: real English words (different parts of speech), pseudowords (meaningless but pronounceable strings) and visually matched false fonts. The pseudowords were generated using MCWord (Medler and Binder, 2005) to match the bigram and trigram statistics of the real English words. All stimulus strings comprised four characters. For the full list of stimuli, see Appendix 1.

Two different fonts were used for the presentation of real words and pseudowords: Sloan and Courier New. The false fonts, PseudoSloan and BACS2, created by (Vildavski et al. 2022) and (Vidal et al. 2017), were chosen to match the visual properties of these fonts such as the amount of ink, number of edges in letters and the amount of curvature in the orthography of the Latin alphabet. All four fonts were scaled such that the height of the letter 'x' was 0.41 dva. The distance between centers of neighboring letters was roughly 0.53 dva (ranging 0.52–0.55 across fonts). Each word was 2 dva wide, and between 0.41 and 0.78 dva tall. The color of the character string was dark gray on 67% of trials (70% contrast against the background) or dark red on 33% of trials. The dark red was roughly the same luminance as the dark gray. Similarly, on independently selected 33% of trials, the fixation dot turned from dark gray to dark red during the 150 ms that the character string appeared. On the remaining trials the fixation dot remained dark gray. See the description of tasks below for how the colors were controlled.

On each trial, a single character string flashed for 150 ms at one of three locations: centered at fixation, 3 dva left, or 3 dva right). The stimulus was followed by a blank with only the fixation mark present for 3850 ms, during which the participant had the opportunity to respond via button press. After every five trials with a stimulus, there was a rest period (no task except fixation on the dot). The duration of the rest period was either 4, 6 or 8 s in duration (randomly and uniformly selected).

Each character string was presented exactly once to each participant during the experiment. The sets of words used for each participant were also balanced to equalize the orthographic neighborhood size, log lexical frequency, mean reaction time and accuracy estimated from the English Lexicon Project (Balota et al. 2007) across the three presentation locations used in the experiment, to minimize effects of differences in task difficulty. We generated 6 sets of word to location assignments, and randomly used one out of 6 for each participant within our experiment.

Task

Participants performed three different tasks during different runs, two of which involved attending to the character strings, and one that was designed to encourage participants to ignore them. In the ***lexical decision task,*** participants reported whether the character string on each trial was a real word or not. In the ***stimulus color task***, participants reported whether the color of

the character string was red or gray. In the ***fixation color task***, participants reported whether or not the fixation dot turned red. This fixation task served as the baseline for neural responses: the same stimuli were presented as in the other two tasks, but the participants were required to ignore them as they tried to detect a subtle, threshold-level color change that occurred at the exact same time the character string appeared. For the stimulus color task and the fixation color task, the saturation of the red color was updated using an adaptive staircase that converged at the 80% correct detection threshold. During runs when the character (or fixation dot) color were not task-relevant, the saturation levels across trials were replayed from the staircase in another color task run.

Across all three tasks, the probability of a real word appearing was 0.33. The probability that the character string on each trial was red was 0.33. Similarly, the probability that the fixation dot turned red was 0.33. All three of those events were independent of each other, and consistently distributed across tasks. Thus, for all three tasks, the probability of a "target" (a real word, a red character string, or a red fixation dot) was 0.33, and the stimuli were matched across tasks. All that changed across runs was which stimulus attributes were task-relevant.

Participants were instructed to respond via button press with their right index finger for "yes" responses (word or not, red or not), and right middle finger for "no" responses. Participants were instructed to respond within 3s after each stimulus and to prioritize accuracy over speed.

Procedure

Before the scanning session, participants were trained on all tasks in a testing room outside the scanner. They performed a minimum of 24 trials of each task, and were also trained to calibrate the eye-tracker and to maintain fixation. Instructions reminded them to respond as accurately as possible.

The main experiment consisted of fifteen 5-minute runs, with 5 runs for each of the three tasks. The order in which the tasks were presented was:

LD SC FC SC FC LD FC LD SC LD FC SC SC LD FC

where LD is lexical decision, SC is stimulus color, and FC is fixation color task. This order was chosen such that each task was equally likely to follow both of the others. As described previously, the stimuli that each participant saw within these runs were chosen from one out of six possible counterbalanced orders.

Each run began with written instructions on the screen, informing the participant which task they were to perform. They pushed a button to proceed, and then were quizzed on which task they were to perform in that run. The run would not start until the participant answered correctly, to ensure that they performed the correct task.

There was no feedback about response accuracy after each trial, but at the end of each run, text appeared on the screen to report their percent correct, along with the percentage of trials in which they failed to make a response.

The sequence of trials within each run was counterbalanced and uniquely generated for each participant. There were 45 trials in each run, and the category of the character string stimulus was randomized. In summary, there were 27 unique conditions, defined by stimulus type, stimulus location, and task. There were 25 trials per condition.

Localizer for visual category-selective ventral temporal regions

Stimuli

Participants viewed sequences of images each of which contained 3 items of one category: words, pseudowords, false fonts, faces, and limbs (examples in **Figure 1A**). This localizer was also used in White et al., 2023, which describes the stimuli in detail. The major change here is the addition of the "limbs" category instead of objects. The grayscale limb images included hands, arms, feet and legs (adapted from Stigliani et al. 2015). We used these particular stimuli to elicit responses in other high level visual areas in the ventral temporal cortex, beyond the VWFA, such as face and limb regions.

Task

Participants performed a one-back repetition detection task.On 33% of the trials, the exact same set of images flashed twice in a row. The participant's task was to push a button with their right index finger whenever they detected such a repetition.

Procedure

Each participant performed 4 runs of the localizer task. Each run consisted of 77 four-second trials, lasting for approximately 6 minutes. Each category was presented 56 times through the course of the experiment.

Language Network Localizer

Stimuli

In order to map the entire language responsive network that can be engaged by written text, we used a language localizer (Mahowald & Fedorenko 2016). Details about the localizer are provided by the authors who developed the resource, briefly, the stimuli included visually present written words and pronounceable pseudowords. The words were presented as meaningful sentences, while the control, non-linguistic condition formed jabberwocky phrases. Both conditions were presented using rapid serial visual presentation of 12 unique stimuli.

Task and Procedure

An icon of a hand would appear following the presentation of twelve character strings, as a prompt for the participant to make a motor response via button press. They were instructed to read the stimuli as they normally would outside the scanner. Participants performed three runs of the language localizer task, and each run lasted for approximately 6 minutes. Each run included 16 trials, each trial lasted for 6s, beginning with a blank screen for 100ms, followed by the presentation of 12 words or pseudowords for 450ms each (5400s total), followed by a response prompt for 400ms and a final blank screen for 100ms. Each run also included 5 blank trials which lasted for the same duration as regular trials.

Study Procedure

The entire study required four sessions per participant. The first session was dedicated to obtaining informed consent, MR safety screening and training the participant on the tasks and eye-tracking The first two MRI sessions were identical: T1 structural scan, 6 runs of the main experiment, 2 runs of the ventral visual localizer, and finally fieldmap acquisition. One run of the localizer was interspersed after every 2 runs of the main experiment. In the third session, participants performed the final 3 runs of the main experiment, and 3 runs of the language localizer.

Analysis of behavioral performance

We calculate accuracy as the proportion of trials with correct responses made within 3.85 seconds after stimulus onset. On an average of 8% of trials (SEM=2%), the participant did not make a response within that window before the trial ended. Those trials were excluded from analyses of accuracy and response time.

Preprocessing

We pre-processed our MRI data with *fMRIPrep* 21.0.1 (Esteban, Markiewicz, et al. (2018); Esteban, Blair, et al. (2018); RRID:SCR_016216), which is based on *Nipype* 1.6.1 (K. Gorgolewski et al. (2011); K. J. Gorgolewski et al. (2018); RRID:SCR_002502).

Preprocessing of B_0 inhomogeneity mappings

A total of 3 fieldmaps were found available within the input BIDS structure for this particular subject. A B_0 nonuniformity map (or *fieldmap*) was estimated from the phase-drift map(s) measure with two consecutive GRE (gradient-recalled echo) acquisitions. The corresponding phase-map(s) were phase-unwrapped with prelude (FSL 6.0.5.1:57b01774).

Anatomical data preprocessing

A total of 3 T1-weighted (T1w) images were found within the input BIDS dataset. All of them were corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al. 2010), distributed with ANTs 2.3.3 (Avants et al. 2008, RRID:SCR_004757). The T1w-reference was then skull-stripped with a *Nipype* implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 6.0.5.1:57b01774, RRID:SCR_002823, Zhang, Brady, and Smith 2001). A T1w-reference map was computed after registration of 3 T1w images (after INU-correction) using mri_robust_template (FreeSurfer 6.0.1, Reuter, Rosas, and Fischl 2010). Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, RRID:SCR_001847, Dale, Fischl, and Sereno 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al. 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* [Fonov et al. (2009), RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym].

Functional data preprocessing

First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 6.0.5.1:57b01774, Jenkinson et al. 2002). The estimated *fieldmap* was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field coefficients were mapped on to the reference EPI using the transform. BOLD runs were slice-time corrected to 0.706s (0.5 of slice acquisition range 0s-1.41s) using 3dTshift from AFNI (Cox and Hyde 1997, RRID:SCR_005927). The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve and Fischl 2009). Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al. (2014)) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al. (2002)). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al. 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi et al. 2007). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The

implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, the aCompCor masks are subtracted a mask of pixels that likely contain a volume fraction of GM. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's *aseg* segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the *k* components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al. 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. The BOLD time-series were resampled into standard space, generating a *preprocessed BOLD run in MNI152NLin2009cAsym space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. The BOLD time-series were resampled onto the following surfaces (FreeSurfer reconstruction nomenclature): *fsaverage*, *fsnative*. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos 1964). Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

BOLD Response Estimation

To estimate the BOLD responses to each trial, we used GLMSingle in Python (Prince et al. 2022), which optimizes the hemodynamic response function for every voxel or node in the brain. For the main experiment, the design matrix consisted of a column for each combination of task, stimulus and presentation location. For the functional localizer, each unique category of stimulus received its own column. For the language localizer, the design matrix included two columns corresponding to the sentence and jabberwocky conditions. Before estimating the BOLD responses, the timeseries for each run were upsampled from 1.5s to 1s using an interpolation function from pyslicetime (Kay et al. 2020). We used GLMSingle to estimate the response to each individual trial of each condition for each surface node in each participant's fsnative space. To obtain single trial beta responses for each participant, we used the GLMSingle model that found the optimal hemodynamic responses function for each surface node, estimated cross validated nuisance regressors computed from repetitions of the same experimental condition, and regularized responses for each node to remove correlated noise across single trials.

Region of Interest Definition

All regions analyzed in this study were obtained using one of three procedures: 1) Visual localizer (regions depicted in color in **Figure 1B**), 2) Language localizer (regions depicted in black in **Figure 1B**) or from an independent study using a separate pool of subjects scanned in a different location (regions depicted in white in **Figure 1B**). Details about the participants we were able to identify these ROIs in are reported in Supplementary Table 1.

We defined regions of interest (ROIs) in the ventral temporal cortex by contrasting the beta responses to different image categories in the visual localizer. For each category-selective region (words, faces, and limbs), we computed t-statistics for the contrast of that category vs. all others using a paired-sample t-test. The t-values were visualized on each participant's inflated cortical surfaces, with the threshold at $t \geq 3$.

We identified up to three text-selection areas (the VWFAs) in each hemisphere. VWFA-1 and VWFA-2 were in the occipitotemporal sulcus, VWFA-1 being anterior to hV4 and VWFA-2 was a separate region anterior to VWFA-1. The face-selective areas, FFA-1 and FFA-2 were medial to the VWFAs, within the fusiform gyrus. FFA-2 was anterior to FFA-1 but sometimes contiguous with it. The limb-selective area was lateral to the VWFAs, defined in the occipito-temporal sulcus (Grill-Spector and Weiner 2014) but in some subjects contiguous with a patch on the lateral side of the brain (encompassing what is sometimes called the extrastriate body area (Downing et al. 2001). The ROI for area hV4 was taken from a previous study (White et al. 2023), defined in fsaverage space and remapped into each individual's native surface.

We also defined two language-responsive regions in the left frontal lobe ("Broca's area"). First, using our data from the language localizer, we used a similar procedure as the one described above, with a contrast between sentence and jabberwocky conditions. For each individual, we defined a region with a peak in the inferior frontal gyrus as the "Language Broca's Area" (ref) (black frontal region in **Figure 1B**). The second left frontal area, we took from a previous study that contrasted responses to words vs. shapes (White et al. 2023). This we call "Word Form Broca's". It was also in the left precentral sulcus, but slightly posterior (white frontal region in **Figure 1B**). This area was defined in fsaverage space and remapped into each individual's native surface. Finally, we also leveraged the language localizer to define an ROI in the posterior superior temporal sulcus (STS).

In order to obtain the group level ROIs in **Figure 5A** and Supplementary Figures 2-5, we used the following approach. For the VWFAs, we projected individually defined contrasts between text and all other categories in the visual localizer to the template fsaverage space. We then concatenated these contrasts across all subjects, and performed a one sample t-test across subjects to determine the nodes in which responses between text and other categories were significantly different from zero. We performed an FDR correction for multiple comparisons, and only saved the nodes with significant t-values. We used this map to define labels for VWFA-1 and -2 in the fsaverage space. For the Language Broca's area, we used a similar approach but with the contrast between meaningful sentences and jabberwocky in the language localizer.

Statistical analysis

For determining which of our experimental manipulations significantly affected responses in the functional ROIs defined by us, we used linear mixed effect models. The dependent variable was the mean beta response for each subject, and the fixed effects included the stimulus type, location and task. The random effects included slopes and intercepts for these variables across participants. The F-statistics and p-values reported were obtained from a type three sum of squares table upon performing repeated measures ANOVA using orthogonalized contrasts on the fixed effects, and the t-values for post hoc tests are reported with FDR corrected p-values. These analyses were implemented using the python package pymer4 (Jolly, E. 2018).

Functional Connectivity Analysis

Similar to previous studies (Kay and Yeatman, 2017; White et al. 2023), we evaluated across-region correlations in the single-trial beta responses, from which low frequency drifts and other nuisance regressors had already been removed, rather than the raw time series. Specifically, we examined correlations in the fluctuations of activity across trials of the same condition (defined by task, stimulus type, and location}. To do so, we demeaned responses for trials that had the same combination of task, stimulus type and location from each surface node, for each participant. Real words and pseudowords were treated separately in this analysis to account for the higher responses to pseudowords overall.

First, we were interested in specifically exploring the changes in functional connectivity across tasks between pairs of ROIs that have been previously reported as being relevant in processing written text, the VWFAs and the Broca's seed region. We obtained these results by simply averaging the correlation values obtained for nodes within each individual's VWFAs, and averaged these values across subjects, separately for each task. The statistical analysis was similar to the one described for analyzing beta responses, except in the fact that correlation values were used as the dependent variable. Second, we obtained a whole brain estimate of correlations between the Broca's seed region and every other cortical vertex for all our participants. For this analysis, we first projected correlation coefficients during the lexical decision task from each subject's anatomical space to the fsaverage space. We then computed a one sample t-test across subjects to assess if the correlations were significantly different from zero. Using the FDR corrected p-values for each vertex for this test as a threshold, we obtained a map of nodes that were significantly correlated with the seed region during the lexical decision task. These analyses were performed using custom code in python, making use of statistical functions in the libraries in scipy and statsmodels.

To examine the differences in functional connectivity across tasks, we factored out differences corresponding to events and used the residual betas responses for subsequent analysis. We accomplished this by calculating the mean value for the timepoints where the same stimulus category was presented in the same task at the same location, and subtracted this value from each trial that used this combination of experimental manipulations, separately for every node in the cortex. Real words and pseudowords were treated as independent stimulus categories for this analysis. Once the residual beta values were obtained for each node in the brain, we examined functional connectivity across the cortex with a "seed" region (In **Figure 5A**, left Language Broca's is the seed). We averaged the responses from all nodes within the seed region, and then

calculated Pearson's correlation coefficient between it and beta response residuals for every individual node on the cortical surface, separately for each condition. These results are reported in two ways.

References

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends in cognitive sciences*, *7*(7), 293-299.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., ... & Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *science*, *330*(6009), 1359-1364.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in cognitive sciences*, *15*(6), 254-262.

Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience*, *16*(4), 234-244.

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*(2), 199-204.

Price, C. J., & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends in cognitive sciences*, *15*(6), 246-253.

Yeatman, J. D., & White, A. L. (2021). Reading: the confluence of vision and language. *Annual Review of Vision Science*, *7*, 487-517.

Qin, L., Lyu, B., Shu, S., Yin, Y., Wang, X., Ge, J., Siok, W. T. & Gao, J. H. (2021). A heteromodal word-meaning binding site in the visual word form area under top-down frontoparietal control. *Journal of Neuroscience*, *41*(17), 3854-3869.

Dębska, A., Wójcik, M., Chyl, K., Dzięgiel-Fivet, G., & Jednoróg, K. (2023). Beyond the Visual Word Form Area–a cognitive characterization of the left ventral occipitotemporal cortex. *Frontiers in Human Neuroscience*, *17*.

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*(1), 143-156.

Muayqil, T., Davies-Thompson, J., & Barton, J. J. (2015). Representation of visual symbols in the visual word processing network. *Neuropsychologia*, *69*, 232-241.

White, A. L., Kay, K. N., Tang, K. A., & Yeatman, J. D. (2023). Engaging in word recognition elicits highly specific modulations in visual cortex. *Current Biology*, *33*(7), 1308-1320.

Woolnough, O., Donos, C., Rollo, P. S., Forseth, K. J., Lakretz, Y., Crone, N. E., ... & Tandon, N. (2021). Spatiotemporal dynamics of orthographic and lexical processing in the ventral visual pathway. *Nature Human Behaviour*, *5*(3), 389-398.

Rauschecker, A. M., Bowen, R. F., Parvizi, J., & Wandell, B. A. (2012). Position sensitivity in the visual word form area. *Proceedings of the National Academy of Sciences*, *109*(24), E1568-E1577.

Kay, K. N., & Yeatman, J. D. (2017). Bottom-up and top-down computations in word-and face-selective cortex. *elife*, *6*, e22341.

Cohen, L., Dehaene, S., Vinckier, F., Jobert, A., & Montavont, A. (2008). Reading normal and degraded words: Contribution of the dorsal and ventral visual pathways. *Neuroimage*, *40*(1), 353-366.

Gagl, B., Richlan, F., Ludersdorfer, P., Sassenhagen, J., Eisenhauer, S., Gregorova, K., & Fiebach, C. J. (2022). The lexical categorization model: A computational model of left ventral occipito-temporal cortex activation in visual word recognition. *Plos Computational Biology*, *18*(6), e1009995.

Kronbichler, M., Hutzler, F., Wimmer, H., Mair, A., Staffen, W., & Ladurner, G. (2004). The visual word form area and the frequency with which words are encountered: evidence from a parametric fMRI study. *Neuroimage*, *21*(3), 946-953.

White, A. L., Palmer, J., Boynton, G. M., & Yeatman, J. D. (2019). Parallel spatial channels converge at a bottleneck in anterior word-selective cortex. *Proceedings of the National Academy of Sciences*, *116*(20), 10087-10096.

Reich, L., Szwed, M., Cohen, L., & Amedi, A. (2011). A ventral visual stream reading center independent of visual experience. *Current Biology*, *21*(5), 363-368.

Striem-Amit, E., Cohen, L., Dehaene, S., & Amedi, A. (2012). Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron*, *76*(3), 640-652.

Mano, Q. R., Humphries, C., Desai, R. H., Seidenberg, M. S., Osmon, D. C., Stengel, B. C., & Binder, J. R. (2013). The role of left occipitotemporal cortex in reading: reconciling stimulus, task, and lexicality effects. *Cerebral Cortex*, *23*(4), 988-1001.

Yeatman, J. D., Dougherty, R. F., Ben-Shachar, M., & Wandell, B. A. (2012). Development of white matter and reading skills. *Proceedings of the National Academy of Sciences*, *109*(44), E3045-E3053.

Stevens, W. D., Kravitz, D. J., Peng, C. S., Tessler, M. H., & Martin, A. (2017). Privileged functional connectivity between the visual word form area and the language system. *Journal of Neuroscience*, *37*(21), 5288-5297.

Chen, L., Wassermann, D., Abrams, D. A., Kochalka, J., Gallardo-Diez, G., & Menon, V. (2019). The visual word form area (VWFA) is part of both language and attention circuitry. *Nature communications*, *10*(1), 5601.

Li, J., Osher, D. E., Hansen, H. A., & Saygin, Z. M. (2020). Innate connectivity patterns drive the development of the visual word form area. *Scientific reports*, *10*(1), 18039.

Vidal, C., Content, A., & Chetail, F. (2017). BACS: The Brussels Artificial Character Sets for studies in cognitive psychology and neuroscience. *Behavior Research Methods*, *49*, 2093-2112.

Vildavski, V. Y., Verde, L. L., Blumberg, G., Parsey, J., & Norcia, A. M. (2022). PseudoSloan: A perimetric-complexity and area-controlled font for vision and reading research. *Journal of Vision*, *22*(10), 7-7.

Lerma-Usabiaga, G., Carreiras, M., & Paz-Alonso, P. M. (2018). Converging evidence for functional and structural segregation within the left ventral occipitotemporal cortex in reading. *Proceedings of the National Academy of Sciences*, *115*(42), E9981-E9990.

Fedorenko, E., Hsieh, P. J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *Journal of neurophysiology*, *104*(2), 1177-1194.

Huey, E. B. (1900). On the psychology and physiology of reading. I. *The American Journal of Psychology*, *11*(3), 283-302.

Mishkin, M., & Forgays, D. G. (1952). Word recognition as a function of retinal locus. *Journal of experimental psychology*, *43*(1), 43.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643.

Price, C. J., Wise, R. J. S., & Frackowiak, R. S. J. (1996). Demonstrating the implicit processing of visually presented words and pseudowords. *Cerebral cortex*, *6*(1), 62-70.

Brown, T. L., Gore, C. L., & Carr, T. H. (2002). Visual attention and word recognition in Stroop color naming: Is word recognition" automatic?". *Journal of Experimental Psychology: General*, *131*(2), 220.

Ruthruff, E., Allen, P. A., Lien, M. C., & Grabbe, J. (2008). Visual word recognition without central attention: Evidence for greater automaticity with greater reading ability. *Psychonomic Bulletin & Review*, *15*, 337-343.

Pattamadilok, C., Chanoine, V., Pallier, C., Anton, J. L., Nazarian, B., Belin, P., & Ziegler, J. C. (2017). Automaticity of phonological and semantic processing during visual word recognition. *NeuroImage*, *149*, 244-255.

Joo, Sung Jun, Kambiz Tavabi, Sendy Caffarra, and Jason D. Yeatman. "Automaticity in the reading circuitry." *Brain and Language* 214 (2021): 104906.

Cohen, L., Dehaene, S., Vinckier, F., Jobert, A., & Montavont, A. (2008). Reading normal and degraded words: Contribution of the dorsal and ventral visual pathways. *Neuroimage*, *40*(1), 353-366.

Rossion, B., & Lochy, A. (2022). Is human face recognition lateralized to the right hemisphere due to neural competition with left-lateralized visual word recognition? A critical review. *Brain Structure and Function*, *227*(2), 599-629.

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*(2), 168-185.

Carr, T. H. (1992). Automaticity and cognitive anatomy: is word recognition" automatic"?. *The American journal of psychology*, 201-237.

Caffarra, S., Karipidis, I. I., Yablonski, M., & Yeatman, J. D. (2021). Anatomy and physiology of word-selective visual cortex: from visual features to lexical processing. *Brain Structure and Function*, 1-15.

Mahowald, K., & Fedorenko, E. (2016). Reliable individual-level neural markers of high-level language processing: A necessary precursor for relating neural variability to behavioral and genetic variability. *Neuroimage*, *139*, 74-93.

Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Current opinion in Neurobiology*, *28*, 136-141.

Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *Neuroimage*, *19*(3), 473-481.

Guntupalli, J. S., Wheeler, K. G., & Gobbini, M. I. (2017). Disentangling the representation of identity from head view along the human face processing pathway. *Cerebral Cortex*, *27*(1), 46-53.

Visconti di Oleggio Castello, M., Halchenko, Y. O., Guntupalli, J. S., Gors, J. D., & Gobbini, M. I. (2017). The neural representation of personally familiar and unfamiliar faces in the distributed system for face perception. *Scientific reports*, *7*(1), 12237

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. *Behavior research methods*, *39*, 445-459.

Prince, J. S., Charest, I., Kurzawski, J. W., Pyles, J. A., Tarr, M. J., & Kay, K. N. (2022). Improving the accuracy of single-trial fMRI response estimates using GLMsingle. *Elife*, *11*, e77599.

Kay, K., Jamison, K. W., Zhang, R. Y., & Uğurbil, K. (2020). A temporal decomposition method for identifying venous effects in task-based fMRI. *Nature methods*, *17*(10), 1033-1039.

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536-548.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470-2473.

Jolly, E. (2018). Pymer4: Connecting R and Python for linear mixed modeling. *Journal of Open Source Software*, *3*(31), 862.

Abraham, Alexandre, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gael Varoquaux. 2014. "Machine Learning for Neuroimaging with Scikit-Learn." Frontiers in Neuroinformatics 8. https://doi.org/10.3389/fninf.2014.00014.

Avants, B. B., C. L. Epstein, M. Grossman, and J. C. Gee. 2008. "Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain." Medical Image Analysis 12 (1): 26–41. https://doi.org/10.1016/j.media.2007.06.004.

Behzadi, Yashar, Khaled Restom, Joy Liau, and Thomas T. Liu. 2007. "A Component Based Noise Correction Method (CompCor) for BOLD and Perfusion Based fMRI." NeuroImage 37 (1): 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042.

Cox, Robert W., and James S. Hyde. 1997. "Software Tools for Analysis and Visualization of fMRI Data." NMR in Biomedicine 10 (4-5): 171–78. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L.

Dale, Anders M., Bruce Fischl, and Martin I. Sereno. 1999. "Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction." NeuroImage 9 (2): 179–94. https://doi.org/10.1006/nimg.1998.0395.

Esteban, Oscar, Ross Blair, Christopher J. Markiewicz, Shoshana L. Berleant, Craig Moodie, Feilong Ma, Ayse Ilkay Isik, et al. 2018. "fMRIPrep." Software. https://doi.org/10.5281/zenodo.852659.

Esteban, Oscar, Christopher Markiewicz, Ross W Blair, Craig Moodie, Ayse Ilkay Isik, Asier Erramuzpe Aliaga, James Kent, et al. 2018. "fMRIPrep: A Robust Preprocessing Pipeline for Functional MRI." Nature Methods. https://doi.org/10.1038/s41592-018-0235-4.

Fonov, VS, AC Evans, RC McKinstry, CR Almli, and DL Collins. 2009. "Unbiased Nonlinear Average Age-Appropriate Brain Templates from Birth to Adulthood." NeuroImage 47, Supplement 1: S102. https://doi.org/10.1016/S1053-8119(09)70884-5.

Gorgolewski, K., C. D. Burns, C. Madison, D. Clark, Y. O. Halchenko, M. L. Waskom, and S. Ghosh. 2011. "Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python." Frontiers in Neuroinformatics 5: 13. https://doi.org/10.3389/fninf.2011.00013.

Gorgolewski, Krzysztof J., Oscar Esteban, Christopher J. Markiewicz, Erik Ziegler, David Gage Ellis, Michael Philipp Notter, Dorota Jarecka, et al. 2018. "Nipype." Software. https://doi.org/10.5281/zenodo.596855.

Greve, Douglas N, and Bruce Fischl. 2009. "Accurate and Robust Brain Image Alignment Using Boundary-Based Registration." NeuroImage 48 (1): 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060.

Jenkinson, Mark, Peter Bannister, Michael Brady, and Stephen Smith. 2002. "Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images." NeuroImage 17 (2): 825–41. https://doi.org/10.1006/nimg.2002.1132.

Klein, Arno, Satrajit S. Ghosh, Forrest S. Bao, Joachim Giard, Yrjö Häme, Eliezer Stavsky, Noah Lee, et al. 2017. "Mindboggling Morphometry of Human Brains." PLOS Computational Biology 13 (2): e1005350. https://doi.org/10.1371/journal.pcbi.1005350.

Lanczos, C. 1964. “Evaluation of Noisy Data.” Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis 1 (1): 76–85. https://doi.org/10.1137/0701007.

Power, Jonathan D., Anish Mitra, Timothy O. Laumann, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. 2014. “Methods to Detect, Characterize, and Remove Motion Artifact in Resting State fMRI.” NeuroImage 84 (Supplement C): 320–41. https://doi.org/10.1016/j.neuroimage.2013.08.048.

Reuter, Martin, Herminia Diana Rosas, and Bruce Fischl. 2010. “Highly Accurate Inverse Consistent Registration: A Robust Approach.” NeuroImage 53 (4): 1181–96. https://doi.org/10.1016/j.neuroimage.2010.07.020.

Satterthwaite, Theodore D., Mark A. Elliott, Raphael T. Gerraty, Kosha Ruparel, James Loughead, Monica E. Calkins, Simon B. Eickhoff, et al. 2013. “An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data.” NeuroImage 64 (1): 240–56. https://doi.org/10.1016/j.neuroimage.2012.08.052.

Tustison, N. J., B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee. 2010. “N4itk: Improved N3 Bias Correction.” IEEE Transactions on Medical Imaging 29 (6): 1310–20. https://doi.org/10.1109/TMI.2010.2046908.

Zhang, Y., M. Brady, and S. Smith. 2001. “Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm.” IEEE Transactions on Medical Imaging 20 (1): 45–57. https://doi.org/10.1109/42.906424.

Supplementary Material

Region of Interest	Hemisphere	Number of Subjects

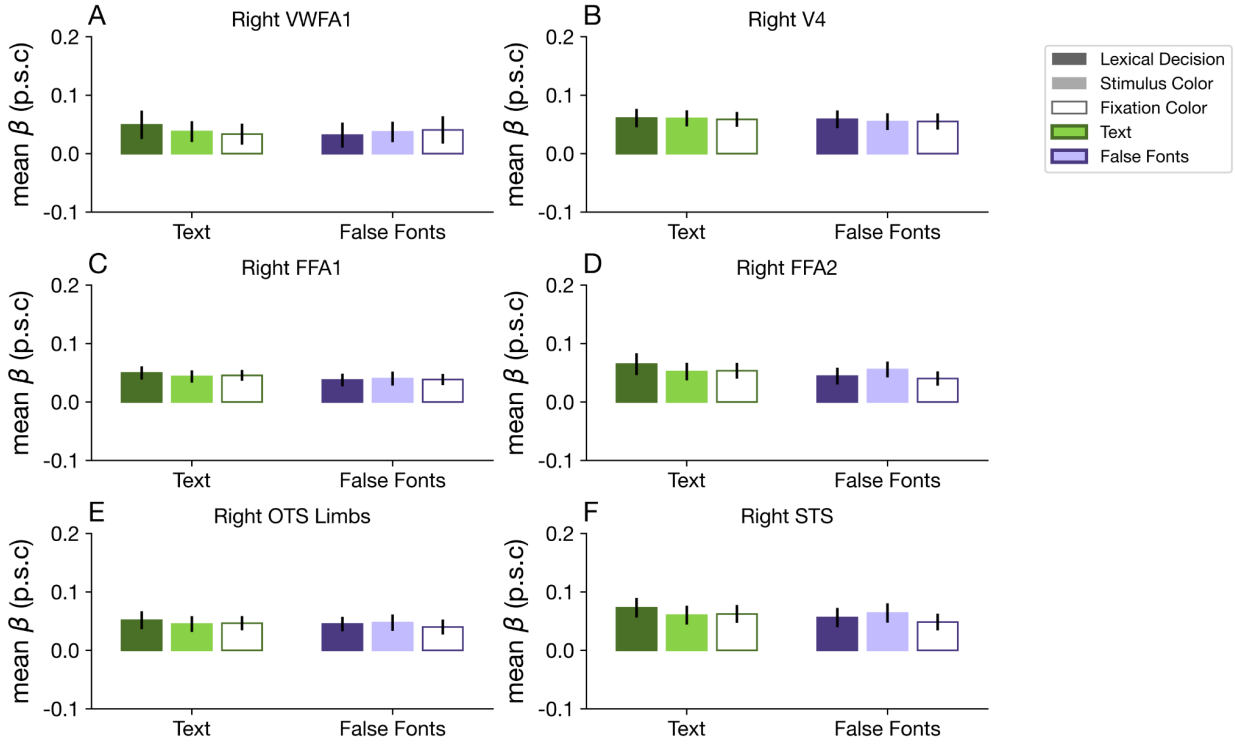
VWFA1	Left	15
VWFA1	Right	9
VWFA2	Left	16
VWFA2	Right	4
Language Broca's	Left	16
Language Broca's	Right	16
FFA1	Left	16
FFA1	Right	16
FFA2	Left	14
FFA2	Right	16
OTS-Limbs	Left	16
OTS-Limbs	Right	16
STS	Left	16
STS	Right	14
MFUS-Text	Left	8
MFUS-Text	Right	2

Supplementary Table 1: Number of participants in whom each ROI could be defined, reported separately for each hemisphere. Only the ROIs using the visual category localizer and the language localizer are reported in the table, as the ROIs defined using data from White et al. 2023 were defined for all subjects.

ROI	**Stimulus**	**Task**	**Mean PSC Difference**	**SEM PSC**	**95% Confidence Interval**	**t-value**	**p-value**
FFA2	Text	LD - FC	0.025	0.02	-0.01, 0.06	1.65	0.72
		LD - SC	0.014	0.02	-0.02, 0.05	0.95	0.72
		SC - FC	0.012	0.02	-0.02, 0.05	0.81	0.72
	False Fonts	LD - FC	-0.03	0.02	-0.08, 0.02	1.48	0.72
		LD - SC	-0.04	0.02	-0.09, 0.01	2.07	0.72

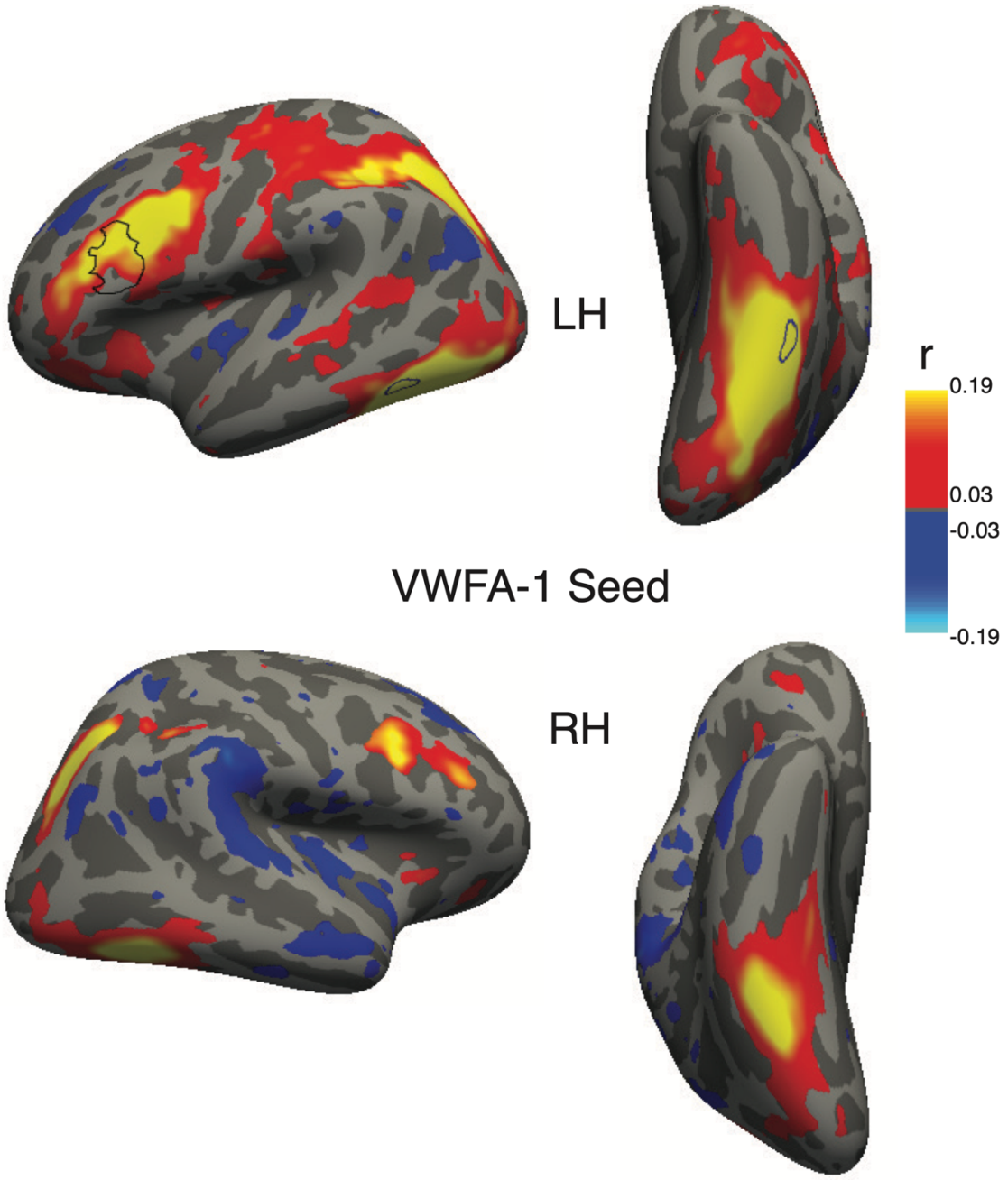
		SC - FC	0.01	0.02	0.04, 0.06	0.55	0.86
OTS-Limbs	Text	LD - FC	0.02	0.01	-0.01, 0.04	1.45	0.72
		LD - SC	0.02	0.01	-0.005, 0.043	1.96	0.72
		SC - FC	-0.003	0.01	-0.03, 0.02	0.28	0.91
	False Fonts	LD - FC	-0.014	0.01	-0.05, 0.02	0.95	0.72
		LD - SC	-0.033	0.01	-0.07, -0.001	2.52	0.48
		SC - FC	0.019	0.01	-0.013, 0.05	1.46	0.72
STS	Text	LD - FC	0.009	0.007	-0.009, 0.03	1.2	0.72
		LD - SC	0.01	0.007	-0.008, 0.03	1.35	0.72
		SC - FC	-0.001	0.007	-0.02, 0.02	0.15	0.96
	False Fonts	LD - FC	0.01	0.01	-0.01. 0.04	1.11	0.72
		LD - SC	-0.002	0.01	-0.03, 0.02	0.16	0.96
		SC - FC	0.013	0.01	-0.01, 0.04	1.28	0.72
V4	Text	LD - FC	-0.01	0.009	-0.04, 0.01	1.38	0.81
		LD - SC	-0.009	0.009	-0.03, 0.14	0.95	0.81
		SC - FC	-0.004	0.009	-0.03, 0.02	0.42	0.89
	False Fonts	LD - FC	0.0	0.01	-0.03, 0.03	0.03	0.98
		LD - SC	-0.001	0.01	-0.03, 0.03	0.05	0.98
		SC - FC	0.001	0.01	-0.03, 0.03	0.08	0.98
text-MFS	Text	LD - FC	0.07	0.02	0.02, 0.12	3.68	0.07
		LD - SC	0.06	0.02	0.01, 0.11	3.29	0.07
		SC - FC	0.01	0.02	-0.04, 0.06	0.54	0.86
	False Fonts	LD - FC	-0.04	0.03	-0.11, 0.02	1.6	0.72
		LD - SC	-0.06	0.03	-0.12, 0.01	2.10	0.72
		SC - FC	0.012	0.03	-0.05, 0.08	0.47	0.89

Supplementary Table 2: Statistical results for stimulus and task differences for all control ROIs analyzed in the study. P-values corrected for multiple comparisons using FDR correction within ROIs, adjusted P-values corrected for multiple comparisons across ROIs using FDR correction reported in the final column of the table.

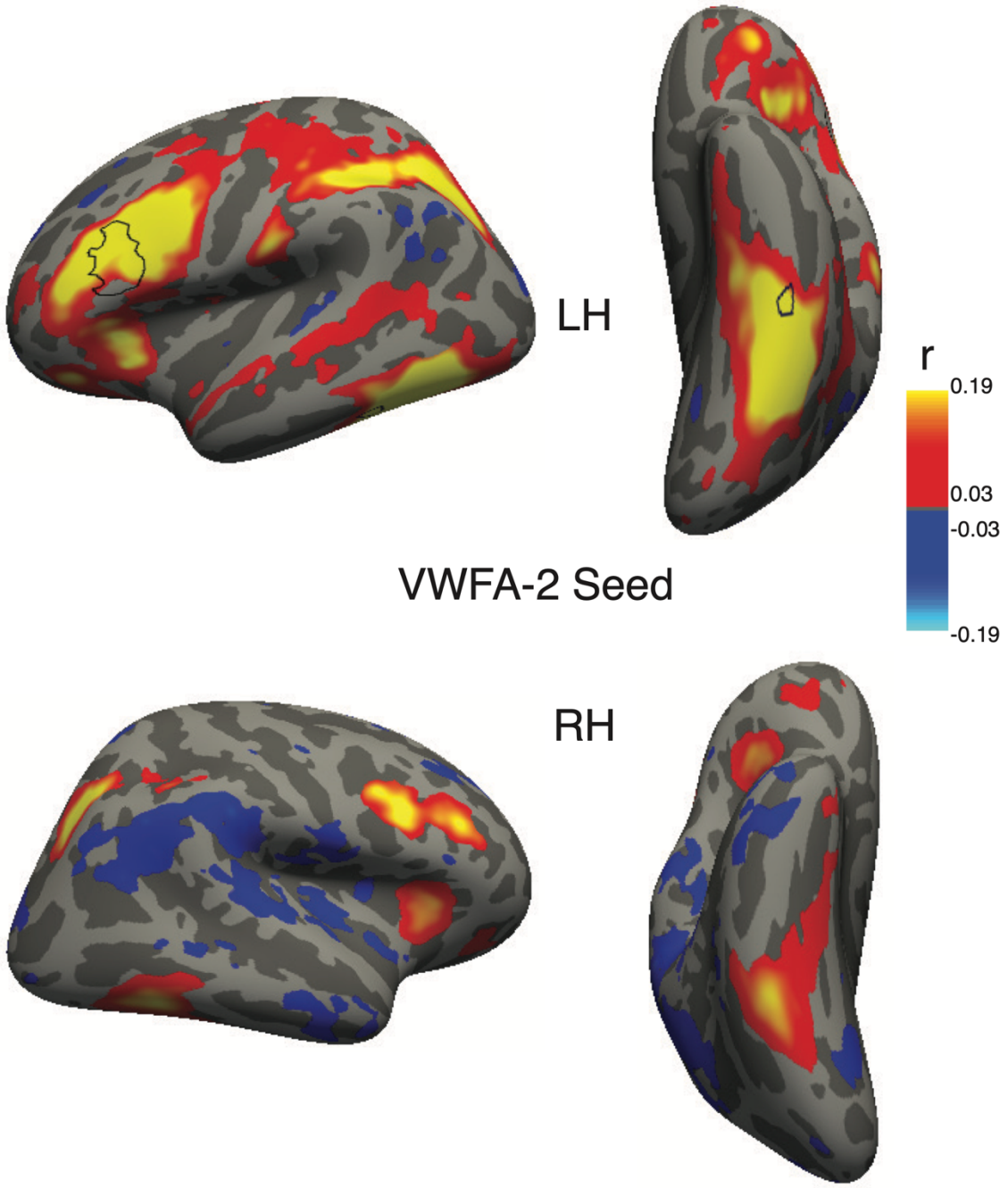


Supplementary Figure 1: BOLD responses in right hemisphere control regions. Mean percent signal change as a function of stimulus and task types in control ROIs in the right hemisphere. Error bars represent standard error of the mean.

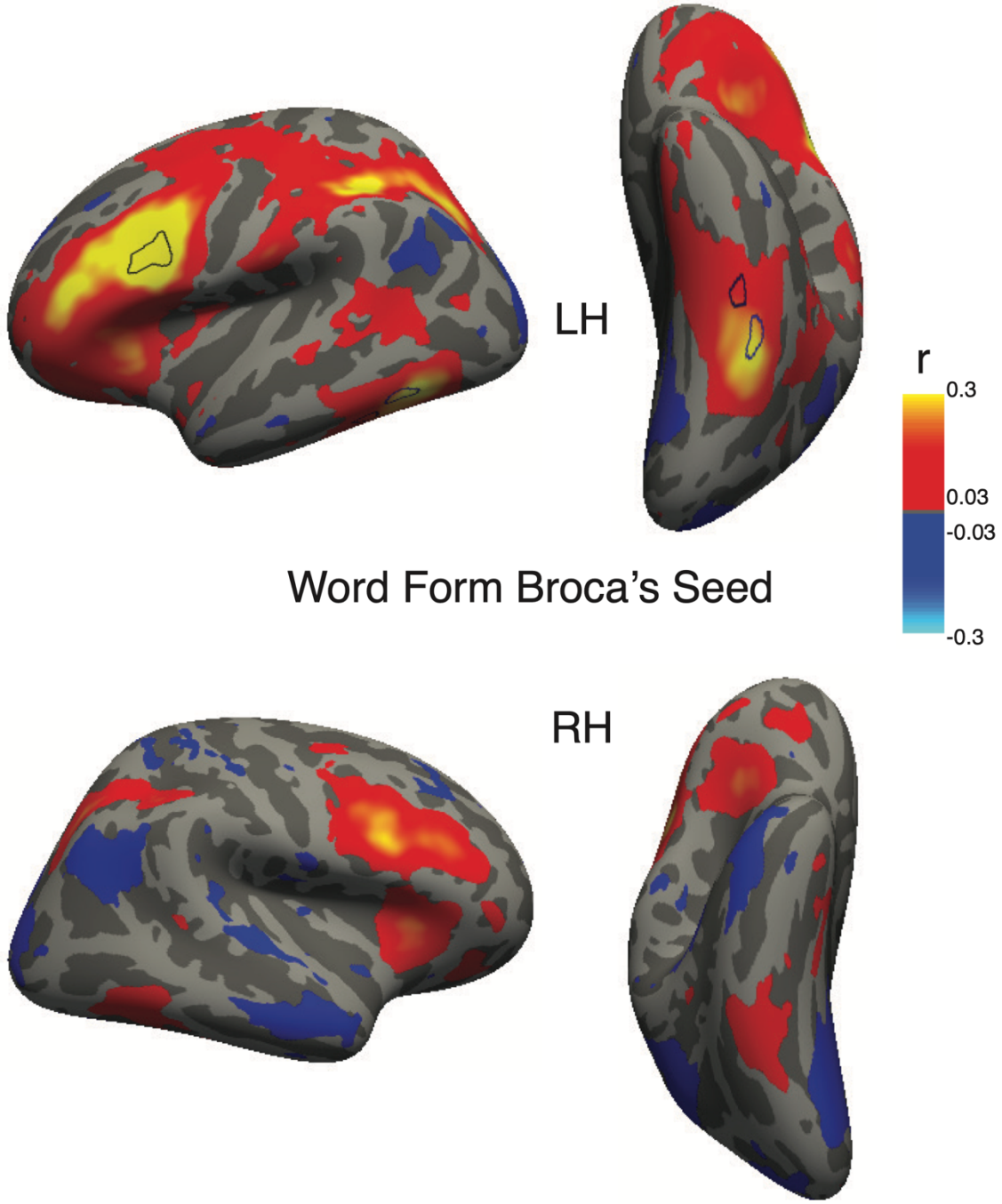
Supplementary Figure 2: Location of the two Broca's areas in fsaverage space. Black: Language Broca's defined using language localizer. White: Word Form Broca's defined using data from White et al. 2023



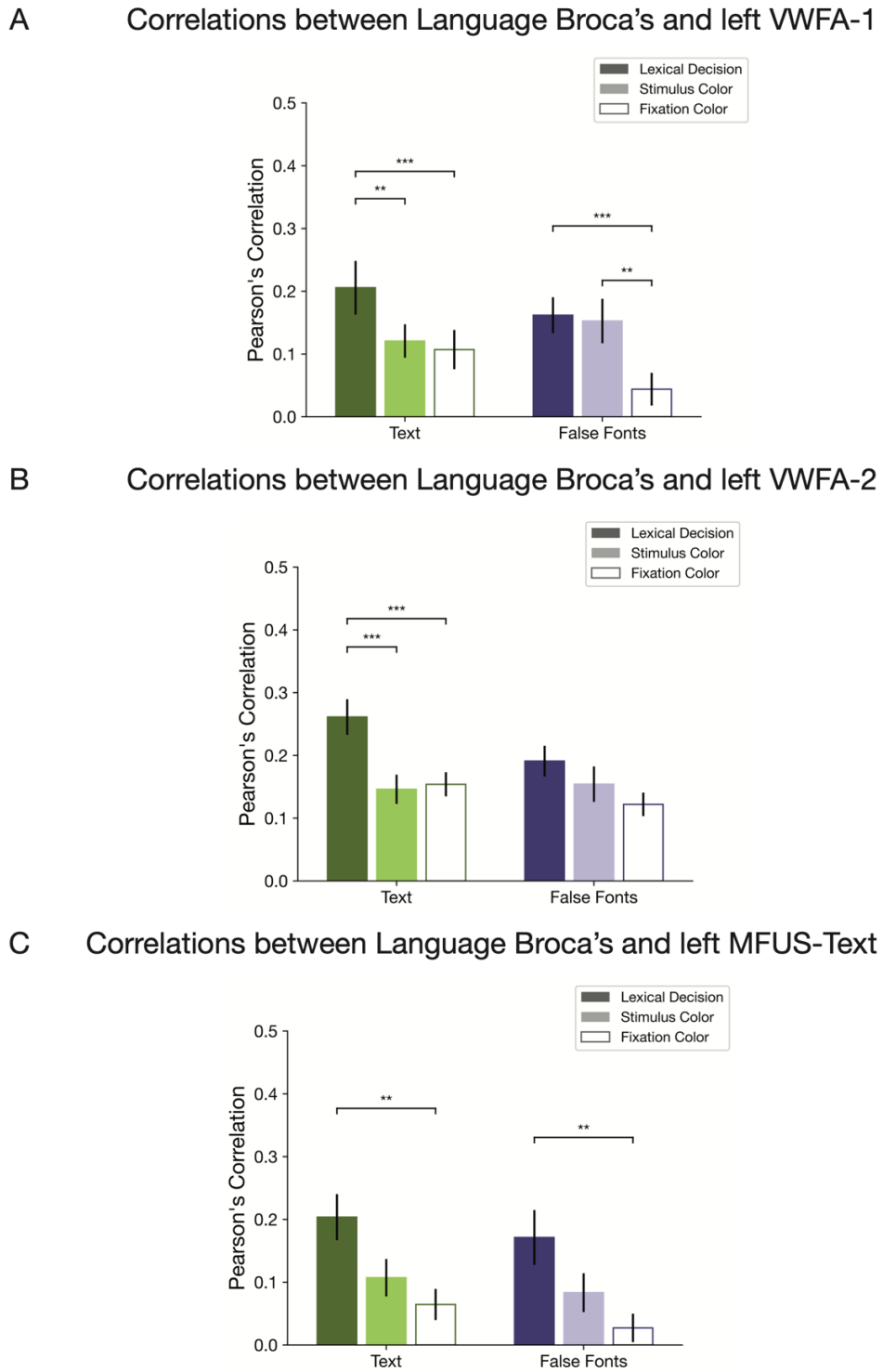
Supplementary Figure 3: Whole-brain connectivity with the seed region in left VWFA-1 during the lexical decision task. The color map represents Pearson's correlation values (minimum r>0.03, p<0.05, FDR corrected).



Supplementary Figure 4: Whole-brain connectivity with the seed region in left VWFA-2 during the lexical decision task. The color map represents Pearson's correlation values (minimum r>0.03, p<0.05, FDR corrected).



Supplementary Figure 5: Whole-brain connectivity with the seed region in left Word Form Broca's during the lexical decision task. The color map represents Pearson's correlation values (minimum r>0.03, p<0.05, FDR corrected).



Supplementary Figure 6: Average correlation coefficients between left language Broca's area and (**A**) left VWFA-1 (**B**) left VWFA-2 and (**C**) left MFUS-Text. The correlations were highest during the lexical decision task for both types of stimuli. Error bars are +/- 1 SEM. Horizontal lines and stars indicate statistical significance. (*=p<0.05, **=p<0.01, ***=p<0.001)