Visual Word Form Area demonstrates individual and task-agnostic consistency but inter-individual variability

Jamie L. Mitchell[1,2]*, Mia Fuentes-Jimenez[1], Hannah L. Stone[3], Maya Yablonski[1,4], Jason D. Yeatman[1,2,4]

[1]Graduate School of Education, Stanford University, Stanford, CA, USA.
[2]Department of Psychology, Stanford University, Stanford, CA, USA.
[3]Department of Psychological & Brain Sciences, University of California, Santa Baraba, CA, USA.
[4]Division of Developmental-Behavioral Pediatrics, Department of Pediatrics, Stanford University School of Medicine, Stanford,California, USA.

Abstract

Ventral Occipital Temporal Cortex (VOTC) is home to a mosaic of categorically-selective functional regions that respond to visual stimuli. Within left VOTC lies the Visual Word Form Area (VWFA) - a text-selective region that develops as an individual learns to read. While there is consistency in the general location of text-selective responses - within the posterior portion of the occipitotemporal sulcus (OTS) - there is also substantial variability across individuals in the size and precise anatomical location. Moreover, there is an ongoing debate over the extent to which the VWFA a) encodes the visual features of text, versus b) is driven by the task of reading. Using functional magnetic resonance imaging, we scanned adults and children as they completed two tasks while viewing text, pseudo fonts, faces, objects, limbs. Experimental tasks were: (1) a fixation task where they were instructed to ignore the stimuli while making psychophysical judgements on the fixation dot; (2) a one-back task where they were instructed to attend to the stimuli and detect repeats. We found that a consistent VWFA location could be identified on each individual's cortical surface using either task. At the same time, the response to text during the one-back task evoked a larger territory of text-selective response (leading to a larger ROI) than the fixation task. However, when averaged in template space, text-selective cortex could not be identified due to variability in the relative locations of text-, face-, object-, and limb-selective cortex. Thus, for both children and adults, a text-selective region (i.e., VWFA) can be localized on the individual cortical surface. This resulting region consistently responds to text irrespective of cognitive task, but text-selective responses are masked when averaged in template space due to variability in the exact configuration of category-selective regions.

* corresponding author contact: jamiel12@stanford.edu
Key Words: visual word form area, individual differences, fMRI, functional organization, ventral occipitotemporal cortex, category selectivity

1 Introduction

The ability to derive meaning from symbols on a page relies on the successful processing of visual stimuli before visual information can be conveyed to language regions of the brain (Yeatman & White, 2021). One brain area commonly studied in reading development is a region of visual cortex, known as the Visual Word Form Area or VWFA, that is tuned to the visual features that comprise a person's native orthography (Cohen et al., 2002; Cohen & Dehaene, 2004). This text-selective patch of cortex, which develops in the posterior and lateral portion of the left ventral occipitotemporal cortex (VOTC), is generally localized to the occipital temporal sulcus (OTS) (Grill-Spector & Weiner, 2014; Yeatman et al., 2013). VWFA is widely recognized as a pivotal neural structure for proficient reading and is hypothesized to be instrumental in the initial stages of word recognition, identifying words and letters from lower-level visual shapes before their association with phonology or semantics (McCandliss et al., 2003; Yeatman & White, 2021). However, since the original report of a "visual word form area" localized to high-level visual cortex based on fMRI (Cohen et al., 2000), there has been a debate over whether this region of cortex a) encodes the visual features of text versus b) is driven by the task of reading (or language more broadly) and should not be considered a visual region per se (Dalski et al., 2023; Grotheer et al., 2018; Price & Devlin, 2003, 2011). A primary reason this debate has persisted is due to methodological inconsistencies in the experiments that have been used to study the VWFA as well as the analytic approaches that are used to define this region.

There are inconsistencies in the literature about the proper way to isolate VWFA. Previous studies suggest that the VWFA typically emerges with a peak of activation around MNI coordinates x = -43, y = -54, z = -12 (McCandliss et al., 2003). This consistency in general location suggests a robust anatomical anchor for the region. Many studies have confirmed that functional localizers can successfully isolate a patch of cortex that selectivity responds to text on the individual cortical surface (Grill-Spector & Weiner, 2014; White et al., 2019; Yeatman et al., 2013), however it is important to distinguish this from the claim of its functional consistency across different tasks. More specifically, although a region can be reliably found with a specific localizer, its precise boundaries might shift, and its activity profile can be dynamically modulated depending on the specific task demands or stimuli presented. Some have claimed that task, rather than stimulus, is the primary factor driving localization implying that there is no consistent text-selective patch of cortex (Dębska et al., 2023; Price & Devlin, 2011; Reich et al., 2011; Striem-Amit et al., 2012).

Despite the broad consensus on the VWFA's involvement in reading, fundamental questions persist concerning its exact functional role (Price & Devlin, 2003; Vogel, Miezin, et al., 2012), its degree of specificity (Price & Devlin, 2011), and the intricate neural circuitry that supports its computations (Yeatman et al., 2013). A central point of contention revolves around whether the VWFA is exclusively specialized for processing written words or if its activity extends to other visual categories and is primarily driven by task demands (Dehaene & Cohen, 2011; Haxby et al., 2001; Price & Devlin, 2003). Fiez and Petersen (1998) noted that the use of various tasks to study the same region only partially replicate and suggest that the use of different tasks across several studies leads to variability among results. Others have asserted that the "purported VWFA" is actually the result of top-down signals from language regions and

has nothing to do with encoding the visual features of text (Dębska et al., 2023; Price & Devlin, 2011; Reich et al., 2011; Striem-Amit et al., 2012). This challenge is further compounded by findings in lower-level visual cortex, where tasks and arousal states have been shown to significantly influence blood-oxygen-level-dependent (BOLD) response (Burlingham et al., 2022; Roth et al., 2020). VWFA possesses an inherent selectivity for orthography, but its functional profile dynamically adapts under the top-down control of the language network when task demands require it (Chauhan et al., 2024; White et al., 2023). This dynamic modulation of functional characteristics, rather than a fixed, static response, contributes to the complexity in fully characterizing the anatomy and function of the VWFA.

A significant contributor to conflicting findings regarding VWFA specificity is the common practice of employing group-level analyses in template space (i.e., MNI - Evans et al., 1993 or fsaverage - Fischl et al., 1999) or template-based ROIs (Glezer & Riesenhuber, 2013). Aligning individual participants to an “average brain” assumes that functional regions have the same spatial configuration across each individual; the averaging process smears BOLD signals across cortical areas, including non-word-selective regions, since the VWFA is located slightly differently in each individual. This directly explains why some studies, particularly those relying on group-level analyses in template space, might report less specificity (or even non-existence) for the VWFA (Vogel et al., 2013; Vogel, Miezin, et al., 2012; Vogel, Petersen, et al., 2012), while others utilizing individual localizers demonstrate strong word-selectivity (Glezer & Riesenhuber, 2013; Mitchell et al., 2025). Consequently, template ROIs might be inadequate to analyze individual response profiles, because they fail to precisely identify the text-selective cortex within individuals, thus signals are contaminated by neighboring regions with dramatically different selectivity patterns. In other words, an area falling within a template ROI for one person may fall outside it for another. The inherent variability in the precise anatomical location and size of the VWFA across individuals means that averaging functional data across participants can effectively "wash out" the selectivity of the region and give the appearance that it responds to other stimuli (e.g., faces or objects) or is driven by various tasks.

VOTC is home to multiple regions that selectively respond to specific categories of visual images including faces, places, bodies and, of course, text (Grill-Spector & Weiner, 2014). Some argue that object recognition is the result of distributed activation patterns throughout this broad swath of cortex (Haxby et al., 2001), while others argue that discrete patches of cortex within VOTC are responsible for recognition of specific categories (Kanwisher et al., 1997). While there is still debate about the modularity of VOTC, it is well documented that VWFAs emerge developmentally later than other categorically-selective regions (Ben-Shachar et al., 2011; E. C. Kubota et al., 2019). As individuals learn to read, dynamic changes unfold within VOTC resulting in the emergence of a VWFA. Reading is an evolutionarily recent skill, which has left humans with an insufficient timeframe for developing an innate brain region that specializes in visual word recognition. Unlike evolutionarily ancient categories such as faces, which humans (and non-human primates) have specialized in recognizing for their entire evolutionary history (processed in regions like the Fusiform Face Area, FFA), text processing must necessarily leverage and repurpose pre-existing neural circuitry. The neuronal recycling hypothesis offers a compelling explanation for this phenomenon (Dehaene & Cohen, 2007). It proposes that reading acquisition recycles or repurposes a region of the ventral visual cortex that was originally involved in more general object recognition or other visual processing

(Dehaene-Lambertz et al., 2018; E. Kubota et al., 2024; Nordt et al., 2021). The very mechanism of this recycling provides a fundamental explanation for individual variability in the VWFA. The precise pre-existing neural biases and the specific "path" of this recycling process may differ slightly across individuals, leading to observable variations in the VWFA's size, exact anatomical location, and functional tuning. This contrasts with more evolutionarily ancient systems, which might exhibit less inter-individual variability in their core functional organization, thereby supporting the assertion that text processing represents a unique case in functional organization.

While the consistent and reproducible localization of the VWFA across individuals and writing systems is frequently asserted (Dębska et al., 2021), a significant degree of variability in its size and precise anatomical location among individuals is also well-documented; likely due in-part to the VWFA's unique experience-dependent developmental trajectory. As noted by Glezer and Riesenhuber (2013), obscuring of individual characteristics through group-level approaches leads to the very controversies surrounding VWFA that is observed in the literature, yet this practice is still widely utilized over a decade after their initial observation. Despite the extensive research on the VWFA, critical gaps remain in our understanding of its variability across individuals and its functional characteristics under different cognitive tasks. Prior studies, including those by Glezer and Riesenhuber, have employed group-level analyses, which obscure the nuances of individual differences in VWFA localization and functioning. As a result, there is insufficient clarity on how the VWFA's selectivity might differ between skilled readers and those with reading difficulties, such as dyslexia. Furthermore, the implications of these individual differences for our understanding of the developmental trajectory of reading skills and the potential influence of varying task demands have not been sufficiently explored. By expanding on Glezer and Riesenhuber's work, this study aims to illuminate the complexities of the VWFA and its neighboring visual regions, ultimately contributing to a more nuanced understanding of the neural basis of reading. In this study, we expand on Glezer and Riesenhuber's work and test two specific hypotheses regarding the consistency and variability of the VWFA and surrounding regions:

Hypothesis 1: Within an individual, there is a consistent location within VOTC that responds to text, irrespective of the cognitive task.

Hypothesis 2: The precise configuration of this text-selective region relative to face-, object-, and limb-selective regions will vary across individuals, such that group averages in template space will mix signals among multiple regions leading to misleading results.

We address these two hypotheses in two separate samples: (1) literate adults who are highly skilled readers and (2) children varying in age and reading ability, including those with dyslexia. Participants were scanned while engaging in different tasks involving different visual categories, including real words, pseudo words, faces, objects, and limbs. By adopting an individual-level analysis approach, categorically-selective ROIs were defined on each individual's native cortical surface. We expand on the approach of Glezer and Riesenhuber (2013), examining metrics of size, task effect, and text-selectivity to further investigate the ways methodological choices can have major effects of the study of the reading circuitry and explore how these effects are amplified in a pediatric population.

2 Methods

2.1 Participants

To understand how the anatomical and varying functional landscape affects group-level analyses of the VWFA, 82 children (ages 7-13; 39 female) and 14 adults (ages 21-34; 6 female) underwent functional magnetic resonance imaging (fMRI) scans. Child data was collected as part of a larger longitudinal study (see Mitchell et al., 2025; Stone et al., 2025). A total of 87 children and 15 adults were recruited, however data for 5 children and 1 adult were not included in the current analysis (see Data Inclusion for more details). All adult participants gave written informed consent and child participants gave informed verbal assent accompanied with parent informed written consent in compliance with the Stanford University Institutional Review Board. Participants all reported normal or corrected-to-normal vision and self-identified as either a monolingual English speaker or as using English for at least 60% of their daily communication. See Table 1 for more detail on participant demographic information.

Participant Demographics		
	Adults	**Children**
N	14	82
Gender (f/m)	6/8	39/42
Mean age ± SD	25.93 ± 3.60	10.07 ± 1.31
Multilingual Count	4	30

Table 1 | Participant Information
Demographic information separated by age group. Gender, multilingual status (speaking more than one language on a daily basis), are collected as self-report responses.

2.2 Experimental Design

The functional scans consisted of two runs of two separate functional localizer tasks (one-back and fixation) in which participants were shown several images of different types of visual stimuli; an experiment that largely followed the localizer design described in White et al., 2023. In both tasks, participants were asked to fixate on a colored dot at the center of the screen. During the one-back task, in which participants attended to the stimuli, participants were instructed to respond by pressing a button on a response box every time the image on the screen repeated. During the fixation task, in which participants ignored the stimuli, participants were instructed to respond every time the fixation dot at the center of the screen changed color. In each run, images repeated and the fixation dot changed color at random in 33% of trials. Thus, the stimulus presentation and trial structure was kept identical for both tasks with the only difference being the instructions given to direct the task.

Each run of the experiment consisted of 65 4-second trials. Each trial contained a sequence of 4 images from the same stimulus category presented for 800 ms with 200 ms of blank fixation in between. Main categories of stimuli included text, pseudo fonts, objects, faces, and limbs. A “blank” category was also included and trial order was randomized. See Mitchell et al. (2025) and Stone et al. (2025) for more information on the experiment.

2.3 MRI Acquisition

Structural and functional images were collected at the Stanford Center for Cognitive and Neurobiological Imaging (CNI) on a 3 T GE Discovery MR750 fitted with a Nova-Medical 32-channel head coil. A whole-brain anatomical scan was acquired using a T1-weighted magnetization-prepared rapid gradient echo sequence with a 0.9 mm^3 resolution. Functional data was collected using four runs of an echo-planar imaging sequence (2.4 mm^3 voxels, repetition time of 1.19 seconds, echo time of 30 ms, flip angle of 62°, field of view of 20.2^o, with 51 slices collected with a multiband of 3). Participants viewed the screen via a series of mirrors mounted to the head coil on the scanner reflecting onto a 47” 3D LCD Resonance Technology display mounted at the back of the magnet bore with a visual distance of 256.1 cm from the first mirror and a total of 277.1 cm from the participant's eye.

To maximize data quality and minimize potential confounds, we monitored 1) motion in real time via FIRMM Real-Time Motion Tracking, 2) participant engagement and focus through a webcam3) responses during the task, and 4) potential artifacts in the fMRI data. If any data quality issues were detected (e.g., the participant closing their eyes, moving, etc.), the scan was ended early, and researchers communicated with the participant to ensure attentiveness before restarting the scan.

2.4 MRI Data Preprocessing and BOLD Response Estimation

Data preprocessing and blood oxygen level dependent (BOLD) signal calculation were completed using the pipeline outlined in Mitchell et al. (2025) and Stone et al. (2025). In short, we used fMRIprep (Esteban et al., 2019) to preprocess functional data, incorporating anatomical data aligned to the AC-PC axis and processed via FreeSurfer for segmentation and surface reconstruction. BOLD runs underwent various corrections, including head motion estimation, slice-time correction, and co-registration to the anatomical reference. The resulting BOLD time-series were resampled into surface space, with quality assessment procedures in place to exclude noisy runs based on framewise displacement and participant attention. BOLD responses were estimated using a general linear model that incorporated confound regressors, providing results in percent signal change relative to blank trials. T1w images used for preprocessing and visualizations were taken from the first time point where a high-quality image was successfully acquired from each child. Adult participants underwent an identical preprocessing procedure.

2.5 Data Inclusion

87 children were originally recruited and participated in a larger longitudinal project. For the present study, data from a single time-point was used for analysis. Participants completed 4 runs of a functional localizer at every visit and the first time-point in which a child participant successfully completed all four runs of the experiment with high data quality (as defined by the exclusion criteria described below) was used for the current study. Adults completed a single visit and again participants who completed all four runs of the experiment with high data quality were included in the present study. Participants were excluded from the study if they 1) fell asleep or failed to keep their eyes open during any run of the experiment 2) had a mean framewise displacement (FD) of 0.5 mm or greater on any given run 3) if more than 30% of volumes had a FD of 0.5 mm or greater 4) did not complete all four runs of the experiment (two runs of the fixation task and two runs of the one-back task).

2.6 Categorical Response Estimation

In order to define functional regions of interest (ROIs) in each participant's native surface, a general linear model (GLM) was fit to the data using the Nilearn python package (Abraham et al., 2014; Nilearn contributors et al., 2024). Beta estimates from the GLM were then used to compute contrast maps to create surface meshes of category-specific responses. First, a GLM was fit to all four runs of the experiment to produce a single set of category-level responses. Using response estimates from this GLM, three different types of activation maps were created: 1) t-statistic maps were used for ROI definition and were created by taking the t-statistics from a contrast map comparing activation to each category relative to the weighted average of all other categories (i.e. Text > Other), 2) contrast maps were used used for group-level modeling and were created by taking the contrast values from the above contrast maps, 3) Activation or percent signal change (PSC) maps were used to calculate category-specific response estimation and were created by taking the contrast values from a comparison of each category relative to baseline (blank screen during the experiment; i.e. Text > Blank).

Next, to investigate category-selectivity for each task separately, the data were split in half and a second GLM was fit to the one-back task runs and the fixation-task runs separately to derive activation maps by task. Once again, t-statistic maps were derived in the same manner as before to define task-specific ROIs and activation maps (in units of PSC) were created and used to analyze task-specific activations.

2.7 Functional Regions of Interest

All functional ROIs were manually drawn using t-statistic maps on the Freesurfer (Fischl, 2012) Native surface. All regions of interest (ROIs) were established within the confines of the ventral occipitotemporal cortex (VOTC), specifically located posterior to the anterior tip of the occipitotemporal sulcus (OTS), lateral to the collateral sulcus, anterior to the posterior transverse collateral sulcus, and medial to, as well as including, the OTS. T-statistic maps were thresholded at t >3 and all ROIs were defined by grabbing any vertex that fell within specific anatomical boundaries which had a t-statistic at or above the threshold. VWFA ROIs were identified as all vertices that met the threshold value in a text > other contrast, and were located on the left occipitotemporal sulcus and the lateral region of the fusiform gyrus (any cortical area lateral to the mid-fusiform sulcus). FFA ROIs were defined as the vertices that met the threshold value in a faces > other contrast, and were found within the fusiform gyrus and mid-fusiform sulcus. Continuous clusters of activation that extended beyond the defined anatomical boundaries, either medially or laterally, were constrained by these anatomical limits. Two sets of ROIs were defined with this criteria: First, VWFAs and FFAs were drawn using all four runs of data for each participant. Second, task-specific ROIs (denoted tVWFA and tFFA) were defined using the same approach, but using functional maps created for each task separately.

In addition to the native ROIs described above, we created group-level ROIs to assess the degree of spatial overlap across participants. To do so, all native ROIs were projected to average space. Then, a probabilistic map was created by calculating the overlap of each ROIs in average space and dividing by the total number of participants. Finally, these probabilistic maps, calculated separately for both VWFA and FFA in the child and adult participant groups, were thresholded at probability of 0.2 (meaning that more than 20% of participants had a native

ROI at a given vertex) to yield the group ROIs (E. Kubota et al., 2023; Rosenke et al., 2021) . These ROIs are labeled aVWFA and aFFA for adult group average VWFA and FFA respectively, and cVWFA and cFFA for child group VWFA and FFA.

Finally, several literature based, functionally-defined ROIs are used for further analyses to compare effects of native and group labels. VWFAs and FFAs from Rosenke et al. (2021) and Kubota et al. (2023), which were defined using a similar probabilistic map approach as the one described above, were used. For the ease of clarification, pOTS-characters ROI and the combination of the pFus-faces and mFus-faces ROIs from Rosenke and colleagues will be referred to in this paper as rVWFA and rFFA respectively. Similarly, the combination of the pOTS-words and mOTS-words ROIs and the pFus-faces and mFus-faces ROIs from Kubota and colleagues will be referred to as kVWFA and kFFA respectively.

3 Results

3.1 Text-Selective Cortex Can Be Identified in Individual VOTCs

To assess how group level analyses may obscure individual level results related to VWFA, or other small category-selective regions dispersed among larger and more stable regions, we first ensured that VWFA exists on an individual level in our participants.To do so, we identified category-selective regions by contrasting neural response to each category against all other categories (for example, text versus the weighted sum of faces, objects, limbs, and pseudo-fonts, thresholded at $t > 3$) for every vertex in each hemisphere. ROIs were manually drawn on the cortical surface of each participant. Figure 1 shows this process with the projection of the Text > Other t-statistic map and both the VWFAs and FFAs of 6 sample participants (See Supplemental Figure S1 for the visualization of the Faces > Other map in the same individuals). Using this process, we were able to identify a VWFA in 70 of the 82 child participants and in 13 of the 14 adult participants. This is unsurprising as previous work has shown that children only develop a VWFA as they attain a certain reading level (Ben-Shachar et al., 2011; Dehaene-Lambertz et al., 2018; Feng et al., 2022) and our child participants had a large range of reading ability.

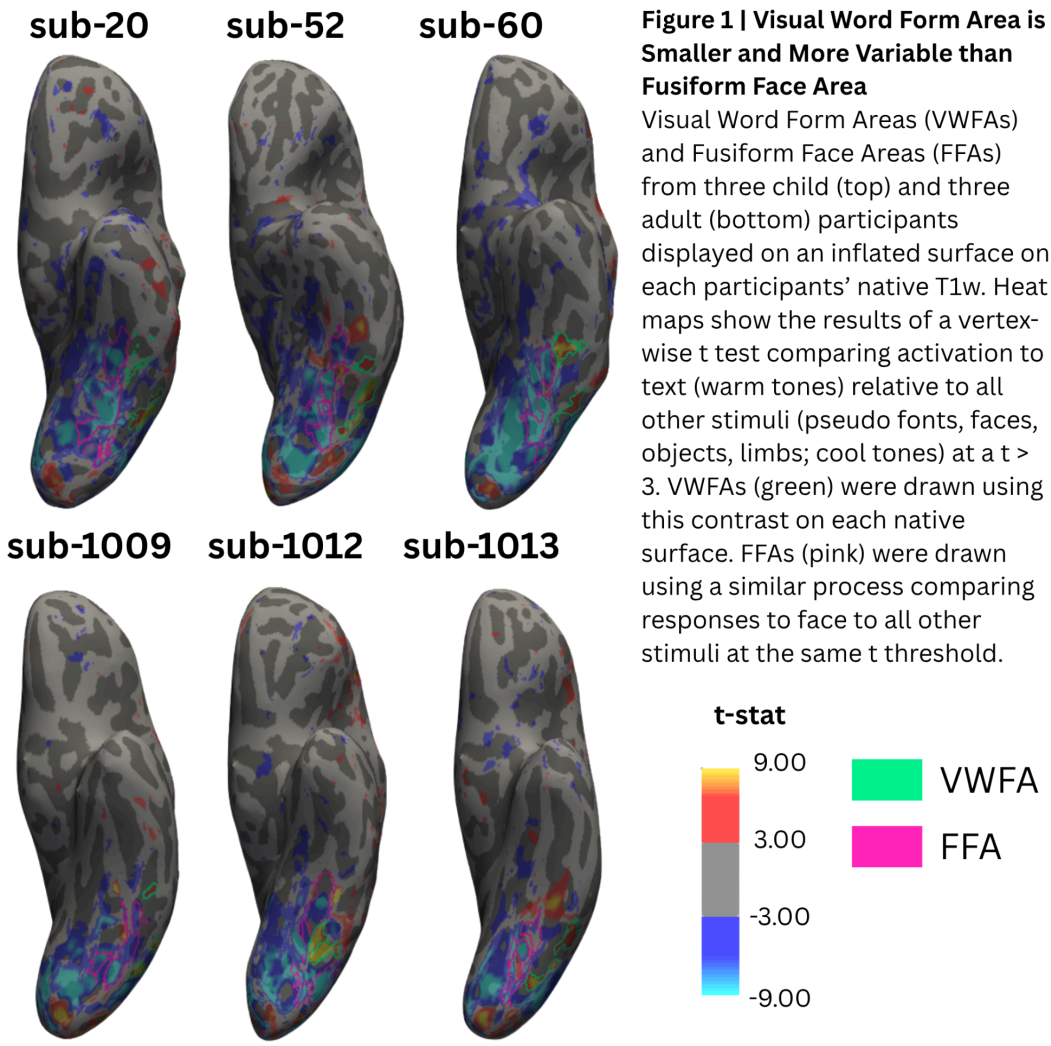

Figure 1 | Visual Word Form Area is Smaller and More Variable than Fusiform Face Area
Visual Word Form Areas (VWFAs) and Fusiform Face Areas (FFAs) from three child (top) and three adult (bottom) participants displayed on an inflated surface on each participants' native T1w. Heat maps show the results of a vertex-wise t test comparing activation to text (warm tones) relative to all other stimuli (pseudo fonts, faces, objects, limbs; cool tones) at a t > 3. VWFAs (green) were drawn using this contrast on each native surface. FFAs (pink) were drawn using a similar process comparing responses to face to all other stimuli at the same t threshold.

Figure 1 | Visual Word Form Area is Smaller and More Variable than Fusiform Face Area
Visual Word Form Areas (VWFAs) and Fusiform Face Areas (FFAs) from three child (top) and three adult (bottom) participants displayed on each participant's native inflated surface. Heat maps show the results of a vertex-wise t test comparing activation to text (warm tones) relative to all other stimuli (pseudo fonts, faces, objects, limbs; cool tones) at a t > 3. VWFAs (green) were drawn using this contrast on each native surface. FFAs (pink) were drawn using a similar process comparing responses to face to all other stimuli at the same t threshold.

3.2 VWFA is Consistent Within Individuals

We next sought to determine if regions identified as the VWFA consistently and stably respond to text across different experiments. To do this, we split our entire dataset in half by task. We defined VWFAs in each individual using, first, only data from the one-back task and second, using only data from the fixation task. Thus each individual had regions defined on independent data with different cognitive tasks. Figure 2a shows the boundaries of these task-specific VWFAs (tVWFAs) in 10 sample participants.

To quantify the consistency of the region identified as the VWFA, we calculated a dice similarity coefficient (DSC; Dice, 1945; Sørensen, 1948) to determine how similar one-back task VWFAs (otVWFAs) were to fixation task VWFAs (ftVWFA; Table S1). This was done with the following formula:

$$DSC = \frac{2\,|\,Ob \cap Fx\,|}{|\,Ob\,| + |\,Fx\,|} \quad (1)$$

where $Ob \cap Fx$ is the total the number of vertices that intersect between both tVWFAs, and $|\,Ob\,| + |\,Fx\,|$ is the sum of the total number of vertices in the surface mesh (which is twice the size of the participants native hemisphere). We found that all participants had an average DSC of 0.465 (sd = 0.262, sem = 0.031), indicating nearly half of the vertices overlap between tasks within an individual. We also found that adults had a higher DSC than children (see Table S1), indicating that this task effect may be reduced developmentally or as reading experience matures. This suggests that, despite some differences, viewing text consistently activates the same patch of VOTC within the individual (Figure 2a).

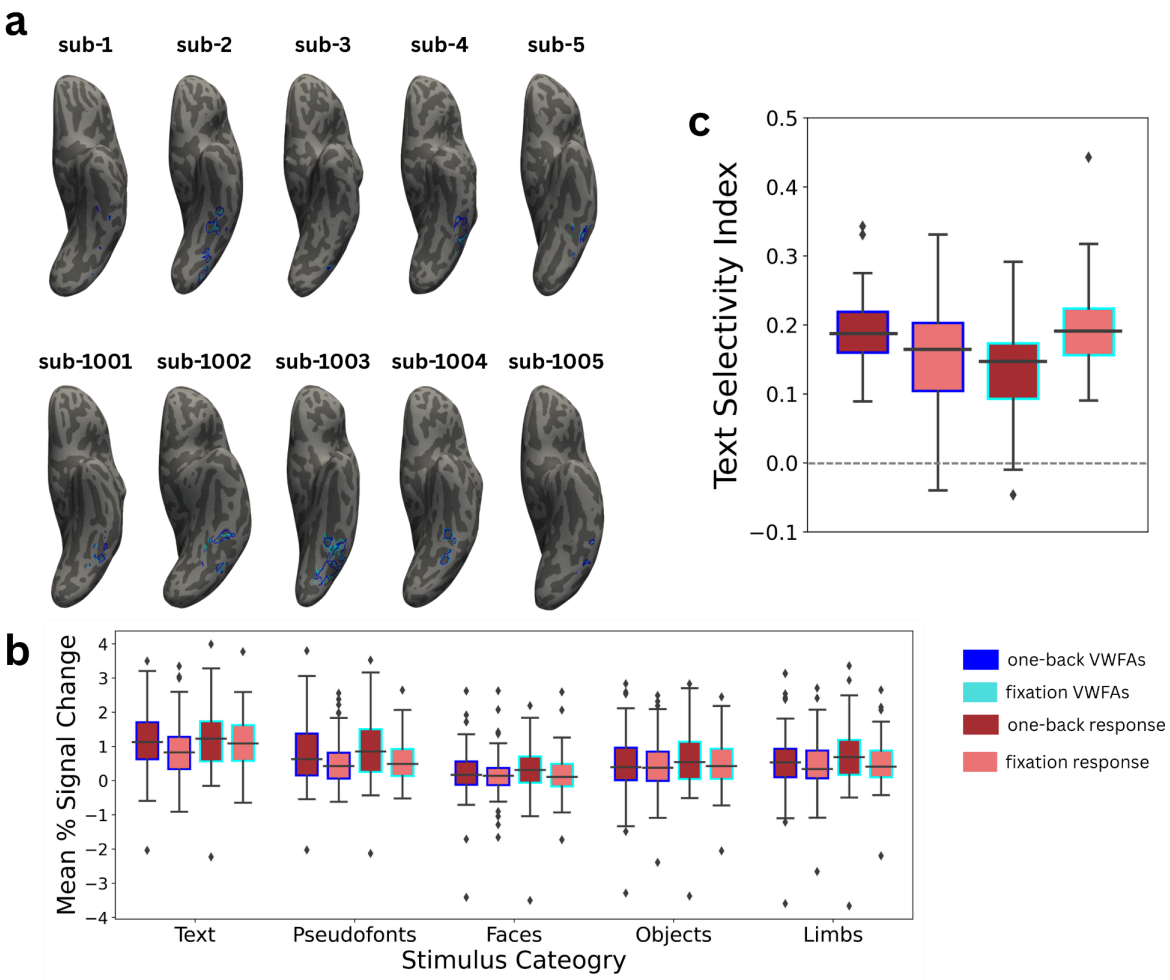

Figure 2 | Visual Word Form Area is Selective to Text Regardless of Cognitive Task
a, Visual Word Form Areas (VWFAs) from five child (top) and five adult (bottom) participants displayed on an

inflated surface on each participants' native T1w. Blue VWFAs represent text-selective regions drawn using only data from the one-back task while cyan VWFAs represent text-selective regions drawn using only data from the fixation color change task. **b**, Neural response to visual categories (text, pseudo fonts, faces, objects, and limbs) in units of percent signal change separated by response during each task and within each task-specific VWFA. **c**, Neural tuning in terms of a text-selectivity index separated by task and within each task-specific VWFA. Data for both plots combined all available data. Box plots display the median and quartiles of the data. Outliers are visualized in panels b and c as diamonds above and below the whiskers for each box plot.

We next analyzed whether the ROI defined based on one task was responsive and selective to text across tasks. To accomplish this, we first measured how an individual's BOLD response to visual stimuli varied between otVWFA and ftVWFA by calculating the percent signal change (PSC) within an individual for each stimulus category (Figure 2b). We then fit a linear mixed effects model using the lme4 package in R. We first estimated the response to each category as a function of a three-way interaction between otVWFA/ftVWFA, task-specific response, and stimulus category with a random intercept of participant, treating the otVWFA and the one-back BOLD response to text as references (See Table S3a). This was modeled with the following equation:

$$BOLD\ Response \sim Category\ \ *\ \ ROI\ Task\ \ *\ \ Map\ Task\ \ +\ \ (1 \mid Participant) \qquad (2)$$

This revealed, unsurprisingly, a negative main effect for each category ($p < 3.21e^{-12}$) confirming that response to text was higher than any other category within both tVWFAs and during both tasks. By splitting the data in half, we were able to cross validate the text-selectivity of VWFA across tasks. In other words, VWFAs defined with one-back data were still more responsive to text during a fixation task and VWFAs defined with fixation data were also more responsive to text during the one-back task compared to any other stimuli. We can therefore conclude that, even without perfect overlap (as seen with the dice similarity calculation) a text-selective region could be defined regardless of task demands.

3.3 Task Modulates VWFA Response

We next looked for within-participant differences in tVWFA properties. We first measured the sizes of each tVWFA to determine if the different tasks drove differences in the extent of cortical territory that was activated. Using a paired-samples t-test, we found that tVWFAs defined with one-back task data were slightly larger than tVWFAs defined with fixation task data ($t = 2.426$, $p = 0.018$; Table S2), in line with previous research regarding task-based modulations of response amplitude (Kay & Yeatman, 2017).

Next, we wanted to determine if the functional tuning properties of the tVWFAs were modulated by task. Again, using the estimates from equation 2, we found no main effect of tVWFA type ($\beta = 0.094$, $p = 0.105$), indicating that response to text during the one-back task does not differ within either the one-back or fixation tVWFA. We did however find an interaction effect between tVWFA type and task-specific BOLD response across all categories ($\beta = 0.174$, $p = 0.033$), suggesting that responses to all categories were slightly higher during the fixation task and within the fixation tVWFA compared to one-back task and tVWFA.

We next fit the same model separately for each category with only a two-way interaction between tVWFA type and BOLD response type. Effects differed slightly by category (See Table S3b for more details) but we found that response to text during the one-back task did not differ

between the one-back and fixation tVWFAs ($\beta = 0.088$, $p = 0.114$). We did however find that within the otVWFA, response to text was weaker during the fixation task when compared to the one-back task ($\beta = -0.254$, $p = 3.63e^{-6}$). We also found an interaction effect ($\beta = 0.174$, $p = 0.026$) such that response to text was strongest within the ftVWFA during the fixation task (See Figure 2b).

We then calculated an index to determine text selectivity in the individual by comparing a person's response to text relative to their response to other categories (Mitchell et al., 2025; Kubota et al., 2023; Figure 2c). We again fit a similar linear mixed effects model, this time estimating text selectivity as a function of the two-way interaction between tVWFA type and task-specific BOLD response (Table S4). Here, we found that text-selectivity during the one-back task was weaker in the ftVWFA compared to the otVWFA ($\beta = -0.081$, $p = 1.04e^{-5}$). We also found that within the otVWFA, text selectivity was weaker during the fixation task compared to one-back task ($\beta = -0.058$, $p = 0.001$). Finally, we found an interaction effect ($\beta = 0.137$, $p = 1.72e^{-7}$) which suggests, unsurprisingly, that text-selectivity is highest when looking at the response based on the same data on which the ROI was defined..

Together these results suggest task demand can influence the size and tuning properties of the VWFA despite the consistency seen in localization. In other words, we confirm that while this region is consistently responsive to the stimuli regardless of task as shown in the previous section, we also see that VWFA is also modulated by task demand. While this modulation does not affect whether a text-selective region can be defined, it does explain why variability in effect sizes may exist in literature when various tasks are used to define and evoke response within VWFA.

3.4 Group Averages in Template Space Obscure Text-Selective Responses

To determine if a text-selective region of cortex can be analyzed using a group level analysis in template space, we used a conventional approach to conducting statistics on the fsaverage template (Fischl, 2012; Friston et al., 1994; Nilearn contributors et al., 2024). This was done by first creating contrast maps for each individual and for each category and transforming them to the fsaverage template. We then ran a 1-sample t-test to determine the vertices that responded significantly more to each category compared to the others. As seen in Figure 3, the ventral surface displays strong negative t-statistics throughout VOTC in both the text > other and the pseudofonts > other group-level maps, indicating a greater response to other categories and no selectivity for text or text-like stimuli. Meanwhile, VOTC does show highly significant responses for the face > other, objects > other, and limbs > other contrasts. These maps reveal a distinct difference in the way group averaging of activation maps affects the inferences that will be drawn regarding different categories of visual stimuli: even though text-selective responses could be localized in each individual (and cross-validated across tasks) these small and anatomically variable regions get obscured in a group average.

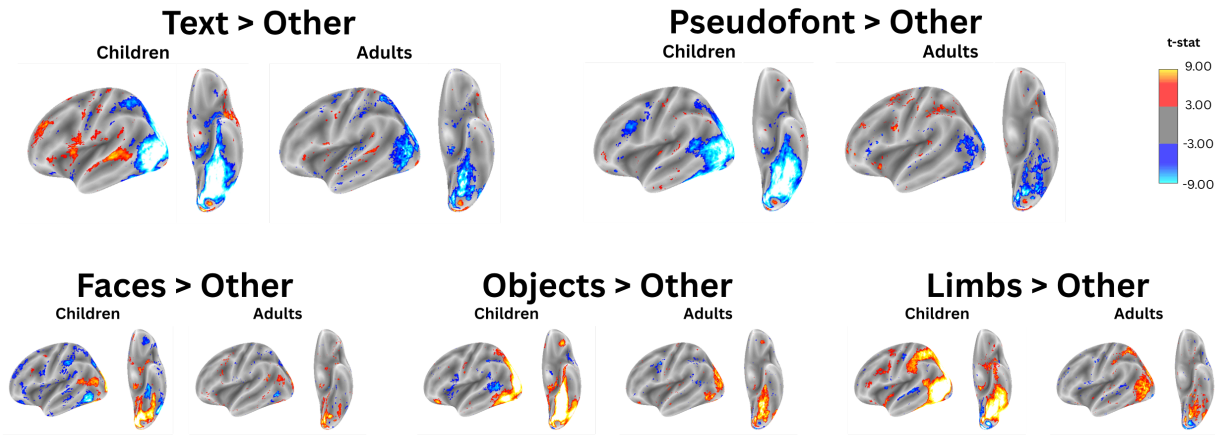

Figure 3 | Group Activations Obscure Text-Like Selectivity in High Level Visual Cortex
Group level activation maps in the left hemisphere in children (left) and adults (right) for each contrast category (text, pseudofonts, faces, objects, and limbs). A general linear model was fit to individual participant data to calculate contrast maps for each participant. Group level analysis was completed by performing a one-sample t-test on all contrast maps. Maps are thresholded at a t of +/- 3 with all positive values displaying selectivity for the named stimulus category (warm tones) and all negative values displaying selectivity of all other, unnamed stimulus categories (cool tones).

3.5 Group and Template VWFAs Obscure Text-Selective Results Due to Individual Differences

Next, we asked whether group-level ROIs, defined on a template, could capture the same response properties as natively drawn individual ROIs. In line with other studies that have defined and published template ROIs (Rosenke et al., 2021), we first created group average ROIs for both VWFA and FFA for children and adults. To this end, each participant's VWFA and FFA were projected to fsaverage (see Figure 4a). We then calculated a probability map representing the percentage of participants with an overlapping ROI at every fsaverage vertex in the left hemisphere. These heat maps, seen in Figure 4b, represent the areas of cortex a given ROI was found in any participant. These maps show that there was not a single vertex that contained the VWFA or FFA in all participants (100% probability). Thus, both FFA and VWFA vary significantly across participants. However, fewer participants shared VWFA vertices (maximum of 25.7% in children and 67.1% in adults) than FFA vertices (maximum of 46.2% in children and 78.6% in adults). This difference is even more pronounced in the child participants compared to the adult participants, as there is higher overlap for VWFA within adult participants than within child participants.

We next examined the variability in the spatial configuration of these regions with respect to each other. Figure 4c visualizes the overlap of text and face ROIs across all participants. This analysis revealed that the area of cortex that is text-selective in some participants, while also being face-selective in others, far outweighs the amount of cortex that is uniquely selective to each category separately. Moreover, the region of cortex that is only responsive to text is substantially smaller than the region of cortex that is only responsive to faces. The amount of cortex that is strictly text-selective is equivalent to 649 vertices in children and 962 vertices in

adults (equivalent to about 0.4% and 0.8% of the total 163,842 vertices in the fsaverage template), compared to the 1,369 vertices in children and 2,450 in adults (0.8% and 1.5%) that are strictly face-selective. Finally, 2,026 vertices in adults and 3,093 vertices in children (1.2% and 1.9%) belong to cortical territory that is text-selective in some participants but face-selective in others. Thus, nearly twice as many vertices are strictly face selective than are strictly text selective.

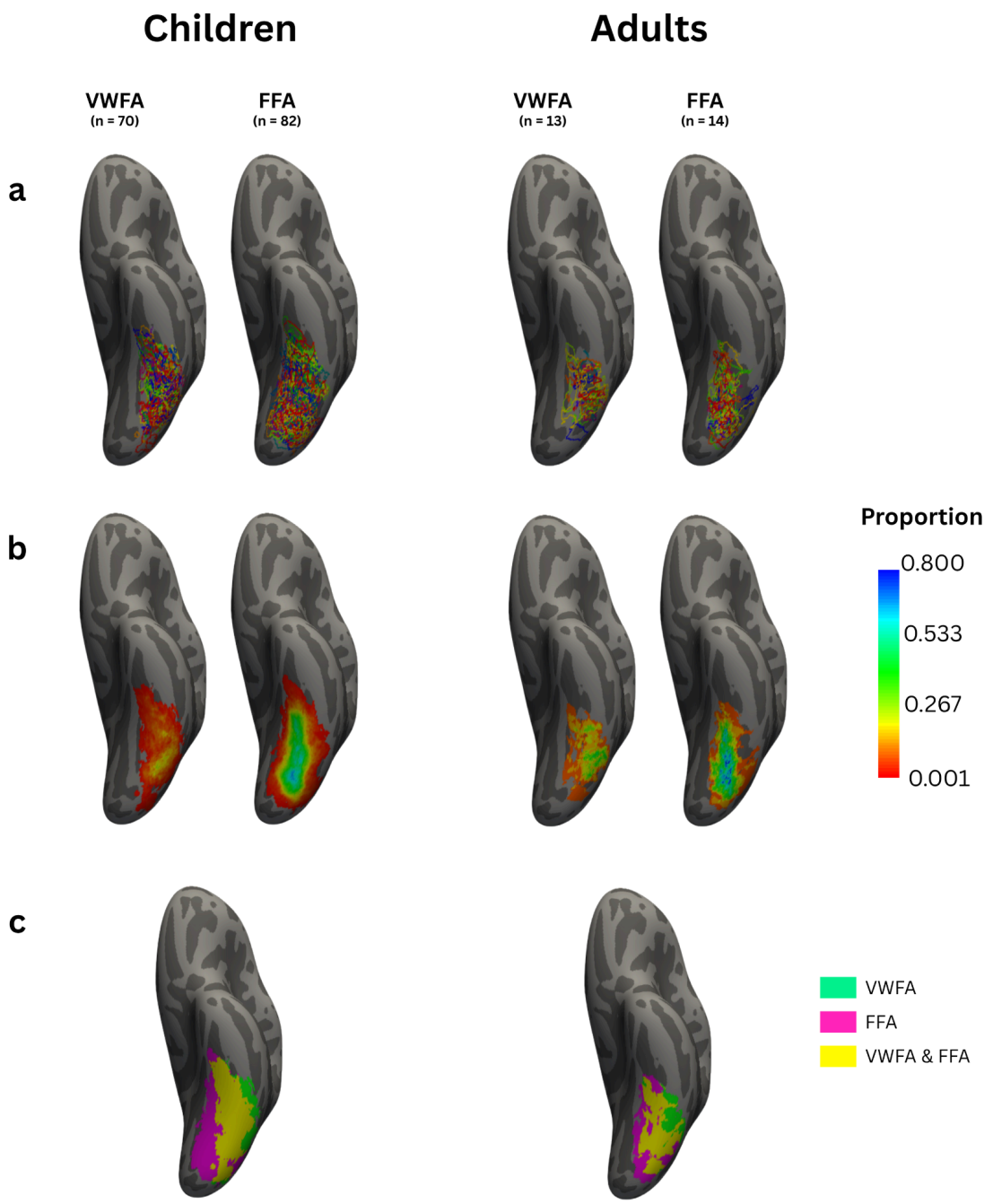

Figure 4 | Variability in Text and Face Selective Regions Across Individuals
a, native Visual Word Form Area labels (VWFA; left) and Fusiform Face Area labels (FFA; right) projected into average space and projected onto an average surface template for every participant in children (left) and adults (right). Different colors represent a label from a different participant. **b**, Heat maps displaying the proportion of participants who had a given region at every vertex. **c**, Indicator map displaying vertices where a VWFA (green) or an FFA (pink) was present in at least one participant in each group. Any vertex in which at least 1 participant had a VWFA and at least 1 other participant had a FFA is displayed in yellow.

Despite this large amount of overlap, we followed the standard approach to create group-average ROIs in order to determine if group-derived labels can accurately measure text-selective responses when applied to individual level activation maps. Specifically, ROIs were created by taking all vertices from the probability map that displayed at least 20% of participant overlap (E. Kubota et al., 2023; Rosenke et al., 2021). We created separate labels for child and adult participants in case any developmental differences exist between the two age groups. We also decided to use template ROIs from these recent studies so as to compare results across several examples of group-level ROI analytical choices. We then analyzed the response properties of these different ROIs to elucidate how these methodological choices affect the inferences that are made about the reading circuitry.

We first wanted to determine if the amplitude of the response to words is lower and the response to other categories was higher in group and template ROIs compared to native ROIs (Figure 5a). We fit a linear mixed effects model looking at BOLD responses as a function of ROI with a random intercept of participant (mean response for native space ROIs as the reference for each stimulus category; see Table S5). We found that most group and template VWFAs measured weaker response to text than the native VWFA ($p < 6.12e^{-4}$) with the exception of the Kubota et al. VWFA (kVWFA) which showed no difference in text response compared with the native VWFA ($\beta = -0.039$, $p = 0.431$). This confirmed our hypothesis that group and template ROIs won't capture peak mean response to text. For pseudofonts, we found mixed results: Compared with the native VWFA, the child group VWFA (cVWFA) showed no difference in response ($\beta = 0.085$, $p = 0.071$), the adult group VWFA (aVWFA) showed a lower response ($\beta = -0.085$, $p = 1.58e^{-4}$), both the Rosenke et al. VWFA (rVWFA) and kVWFA showed a higher response (rVWFA: $\beta = 0.157$, $p = 0.001$, kVWFA: $\beta = 0.261$, $p = 6.23e^{-8}$). As expected, all group and template VWFAs showed greater response magnitude to faces, objects, and limbs ($p < 0.0001$; see Table S5a), in line with the notion that text-responses get washed out when averaged across participants. Combined, this shows that response magnitude to every stimulus category varies greatly depending on the ROI used to analyze this form of functional tuning (see Table S5b for more information on FFAs).

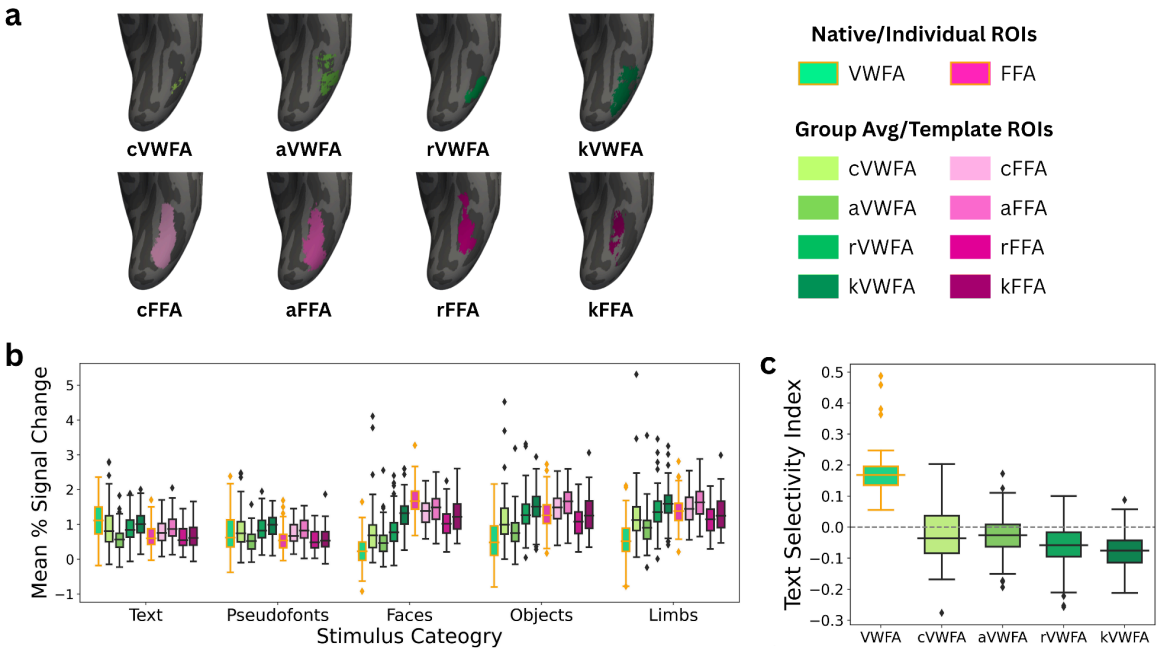

Figure 5 | Group Template ROIs Fail to Capture Individual-Level Text Response
a Each of the four group average and template Visual Word Form Areas (VWFA) and Fusiform Face Areas (FFA) displayed on an inflated ventral surface of the fsaverage anatomy. **b** Response profiles for each of the average and template VWFAs and FFAs displayed next to the response profile based on individually-defined native VWFAs and FFAs (highlighted in orange). Response is measured in terms of percent signal change from a blank condition. **c** Tuning properties of each of the group and template VWFAs calculated as a text-selectivity index displayed with the text selectivity index extracted from individually-defined native VWFAs. VWFA = values extracted from natively defined VWFA; FFA = values extracted from natively defined FFA; c = group average ROI of study child participants; a = group average ROI of study adult participants; r = template ROIs from Rosenke et al., 2018; k = template ROIs from Kubota et al., 2023

We then repeated the text-selectivity index calculation from earlier to determine if the group and template VWFAs accurately identified a patch of cortex that was indeed text selective in our participants. Figure 5b shows that while the individually drawn native ROIs are indeed text-selective (as expected given they were defined using a text-selective contrast map), all four of our group and literature-based template ROIs failed to capture text-selective cortex in our participants, showing a negative average selectivity index. We again fit a linear mixed effects model, this time looking at text selectivity as a function of ROI with a random intercept of participant, treating selectivity within the native individual VWFAs as the reference ROI (see Table S6). In doing so, we found a negative effect for each group and template VWFA ($p < 1.0e^{-77}$), indicating that none of these VWFAs were text selective.

The combination of the BOLD response analysis and the selectivity index analysis explains the lack of a text-selective response seen in Figure 3. When one patch of cortex is selective to one category in some individuals but is also selective to a different category in others, averaging responses across individuals will “wash out” category-specific responses.

4 Discussion

This study investigated how methodological choices affect the inferences that are drawn about text-selective responses in VOTC. We first used a functional localizer to confirm that we could localize the VWFA within the mosaic of category selective regions in VOTC. We manually defined ROIs for each category-selective region on the native cortical surface and confirmed that, within each individual's brain, a consistent patch of cortex selectively responded to words irrespective of the task. While we found that the cognitive task affected the size, response magnitude, and selectivity within the region, we also confirmed that VWFA was consistently localized within an individual across.

We then investigated how moving from an analysis of individual brains to an analysis in template space (fsaverage cortical surface) affected the inferences we might draw about text-selective responses in VOTC. We found that while faces, objects, and limbs evoked a response in a consistent region of cortex across participants, a text-selective cortex could not be identified in a group average despite its existence on an individual level. This suggests that anatomical and functional topography varies significantly across individuals. To understand this relationship better, we looked at how individual ROIs overlap and found that VWFAs across individuals vary in anatomical location. We found that patches of cortex that respond to text in some individuals respond to different visual categories in other individuals. This was even more pronounced in children. In a group level analysis, these conflicting ROIs that share cortical territory will neutralize text-selective responses when averaging across individuals. Additionally, we showed the importance of an individual approach to studying neural response to text by exploring nullifying results obtained from a group level ROI definition. From these findings, we conclude that, with the unique nature of smaller and more variable text-selective regions, individual approaches to analysis in the native space are crucial to obtain consistent and reproducible results. These conclusions are in line with Glezer & Riesenhuber (2013), further emphasizing the importance of an individual analysis approach and encouraging future researchers to follow this practice.

Previous research suggests that certain functionally-selective cortical patches for categories such as faces, limbs, and places in the VOTC are present even without access to visual experience (van den Hurk et al., 2017), and that this organization is the result of basic needs and the ecologically advantageous nature of efficient object discrimination (Rosch et al., 1976). As children grow, changes in their categorically-selective ROIs suggest a recycling model in which selectivity for text increases while selectivity for another category may decrease (Behrmann & Plaut, 2015; Dehaene et al., 2015; Nordt et al., 2021). This cortical recycling theory suggests that there is competition for cortical territory in the VOTC which could lead to the variability in VWFA localization observed in this study. Text recognition is a learned task that takes instruction and practice over years of development, and has only become common practice in the very recent evolutionary timeline compared to other categories of object recognition. Individual-specific experiences may shape the way this patch of cortex develops more than other known categorically selective regions because of its delayed emergence compared to other studied regions. Throughout our analyses we found that task-effects were more pronounced in children than adults, and that the extent of VWFA overlap across tasks was weaker in children relative to adults. This demonstrates the developing nature of the VWFA and

highlights the need for individually defined ROIs to properly capture the effects of this developing region. Further research is needed to understand how the VWFA emerges as people learn how to read, and the effect this development has on the VOTC and other contributing brain regions. Longitudinal studies allow us to measure the changes/emergence of VWFA in individuals as children gain reading skills. We recently showed that, using individually defined VWFA, we were able to capture growth in VWFA related to growth in reading ability and found a strong relationship between VWFA size and reading ability (Mitchell et al., 2025).

We argue that an accurate and comprehensive understanding of VWFA function necessitates its definition individually on native anatomy. Though this argument has been made previously (Glezer & Riesenhuber, 2013), we have yet to see consistent adherence to this guidance. Here we provide a deeper dive into sources of consistency within individuals and variability across individuals. VWFA exhibits consistent localization within individuals despite differences that may arise due to specific task demands utilized to selectively activate the region. Furthermore, group-level functional data and template-based ROIs are inherently inadequate for precisely identifying text-selective cortex within the ventral occipitotemporal cortex. This is particularly pronounced because text represents a unique case of functional specialization compared to other visual categories, such as faces, which have a longer evolutionary history of recognition. Our results suggest that a transition from template-based analyses to individually-localizing regions of interest will resolve ambiguities and controversies that persist in the literature. Particularly in the study of development, an individual approach will be critical for understanding the mechanisms of stability and change within high-level visual cortex.

5 Data and Code Availability

De-identified data has been made publicly available through the Stanford University Libraries Digital Repository and can be found at https://doi.org/10.25740/jp789mc0395. Code has been made publicly available through an online GitHub repository and can be found at https://github.com/jamielmitchell/Mitchell_VWFAmethods_2025.git.

6 Author Contributions

JLM: Conceptualization, Formal analysis, Investigation, Writing (original draft & review and editing), Visualization. **MFJ:** Investigation, Data curation, Project administration. **HLS:** Investigation, Data curation, Project administration. **MY:** Investigation, Supervision. **JDY:** Conceptualization, Resources, Writing (review and editing), Supervision, Funding acquisition.

7 Funding

This work was funded by NICHD R01-HD095861 to JDY.

8 Declaration of Competing Interests

The authors declare no competing interests.

9 Acknowledgements

We would like to thank the individuals and families who participated in this study and the research coordinators and assistants who contributed to data collection.

10 References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*. https://doi.org/10.3389/fninf.2014.00014

Behrmann, M., & Plaut, D. C. (2015). A vision of graded hemispheric specialization. *Annals of the New York Academy of Sciences*, *1359*, 30–46. https://doi.org/10.1111/nyas.12833

Ben-Shachar, M., Dougherty, R. F., Deutsch, G. K., & Wandell, B. A. (2011). The Development of Cortical Sensitivity to Visual Word Forms. *Journal of Cognitive Neuroscience*, *23*(9), 2387–2399. https://doi.org/10.1162/jocn.2011.21615

Burlingham, C. S., Ryoo, M., Roth, Z. N., Mirbagheri, S., Heeger, D. J., & Merriam, E. P. (2022). Task-related hemodynamic responses in human early visual cortex are modulated by task difficulty and behavioral performance. *eLife*, *11*, e73018. https://doi.org/10.7554/eLife.73018

Chauhan, V. S., McCook, K. C., & White, A. L. (2024). Reading Reshapes Stimulus Selectivity in the Visual Word Form Area. *Eneuro*, *11*(7), ENEURO.0228-24.2024. https://doi.org/10.1523/ENEURO.0228-24.2024

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. *NeuroImage*, *22*(1), 466–476. https://doi.org/10.1016/j.neuroimage.2003.12.049

Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Henaff, M.-A., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, *123*(2), 291–307.

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002).

Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, *125*(5), 1054–1069. https://doi.org/10.1093/brain/awf094

Dalski, A., Kular, H., Jorgensen, J. G., Grill-Spector, K., & Grotheer, M. (2023). *Both mOTS-words and pOTS-words prefer emoji stimuli over text stimuli during a reading task*. Cold Spring Harbor Laboratory. https://doi.org/10.1101/2023.11.07.565794

Dębska, A., Banfi, C., Chyl, K., Dzięgiel-Fivet, G., Kacprzak, A., Łuniewska, M., Plewko, J., Grabowska, A., Landerl, K., & Jednoróg, K. (2021). Neural patterns of word processing differ in children with dyslexia and isolated spelling deficit. *Brain Structure and Function*, *226*(5), 1467–1478. https://doi.org/10.1007/s00429-021-02255-2

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*, 1199366. https://doi.org/10.3389/fnhum.2023.1199366

Dehaene, S., & Cohen, L. (2007). Cultural Recycling of Cortical Maps. *Neuron*, *56*(2), 384–398. https://doi.org/10.1016/j.neuron.2007.10.004

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, *15*(6), 254–262. https://doi.org/10.1016/j.tics.2011.04.003

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. https://doi.org/10.1038/nrn3924

Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLOS Biology*, *16*(3), e2004103. https://doi.org/10.1371/journal.pbio.2004103

Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, *26*(3), 297–302.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., & Peters, T. M. (1993). 3D statistical neuroanatomical models from 305 MRI volumes. *1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, 1813–1817. https://doi.org/10.1109/nssmic.1993.373602

Feng, X., Monzalvo, K., Dehaene, S., & Dehaene-Lambertz, G. (2022). Evolution of reading and face circuits during the first three years of reading acquisition. *NeuroImage*, *259*, 119394. https://doi.org/10.1016/j.neuroimage.2022.119394

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Fischl, B., Sereno, M. I., Tootell, R. B. H., & Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, *8*(4), 272–284. https://doi.org/10.1002/(sici)1097-0193(1999)8:4<272::aid-hbm10>3.0.co;2-4

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. -P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*(4), 189–210. https://doi.org/10.1002/hbm.460020402

Glezer, L. S., & Riesenhuber, M. (2013). Individual Variability in Location Impacts Orthographic Selectivity in the “Visual Word Form Area.” *Journal of Neuroscience*, *33*(27), 11221–11226. https://doi.org/10.1523/JNEUROSCI.5002-12.2013

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. https://doi.org/10.1038/nrn3747

Grotheer, M., Jeska, B., & Grill-Spector, K. (2018). A preference for mathematical processing outweighs the selectivity for Arabic numbers in the inferior temporal gyrus. *NeuroImage*, *175*, 188–200. https://doi.org/10.1016/j.neuroimage.2018.03.064

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience*, *17*(11), 4302. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Kay, K. N., & Yeatman, J. D. (2017). Bottom-up and top-down computations in word- and face-selective cortex. *eLife*, *6*, e22341. https://doi.org/10.7554/eLife.22341

Kubota, E. C., Joo, S. J., Huber, E., & Yeatman, J. D. (2019). Word selectivity in high-level visual cortex and reading skill. *Developmental Cognitive Neuroscience*, *36*, 100593. https://doi.org/10.1016/j.dcn.2018.09.003

Kubota, E., Grill-Spector, K., & Nordt, M. (2024). Rethinking cortical recycling in ventral temporal cortex. *Trends in Cognitive Sciences*, *28*(1), 8–17. https://doi.org/10.1016/j.tics.2023.09.006

Kubota, E., Grotheer, M., Finzi, D., Natu, V. S., Gomez, J., & Grill-Spector, K. (2023). White matter connections of high-level visual areas predict cytoarchitecture better than category-selectivity in childhood, but not adulthood. *Cerebral Cortex*, *33*(6), 2485–2506. https://doi.org/10.1093/cercor/bhac221

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. https://doi.org/10.1016/s1364-6613(03)00134-7

Mitchell, J. L., Yablonski, M., Stone, H. L., Fuentes-Jimenez, M., Takada, M., Tang, K. A., Tran, J. E., Chou, C., & Yeatman, J. D. (2025). *Small or absent Visual Word Form Area is a*

trait of dyslexia. https://doi.org/10.1101/2025.01.14.632854

Nilearn contributors, Chamma, A., Frau-Pascual, A., Rothberg, A., Abadie, A., Abraham, A., Gramfort, A., Savio, A., Cionca, A., Sayal, A., Thual, A., Kodibagkar, A., Kanaan, A., Pinho, A. L., Joshi, A., Idrobo, A. H., Kieslinger, A.-S., Kumari, A., Rokem, A., … Nájera, Ó. (2024). *Nilearn* (Version 0.11.0) [Computer software]. Zenodo. https://zenodo.org/doi/10.5281/zenodo.8397156

Nordt, M., Gomez, J., Natu, V. S., Rezai, A. A., Finzi, D., Kular, H., & Grill-Spector, K. (2021). Cortical recycling in high-level visual cortex during childhood development. *Nature Human Behaviour*, *5*(12), 1686–1697. https://doi.org/10.1038/s41562-021-01141-5

Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *NeuroImage*, *19*(3), 473–481. https://doi.org/10.1016/S1053-8119(03)00084-3

Price, C. J., & Devlin, J. T. (2011). The Interactive Account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, *15*(6), 246–253. https://doi.org/10.1016/j.tics.2011.04.001

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. https://doi.org/10.1016/j.cub.2011.01.040

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Rosenke, M., van Hoof, R., van den Hurk, J., Grill-Spector, K., & Goebel, R. (2021). A Probabilistic Functional Atlas of Human Occipito-Temporal Visual Cortex. *Cerebral Cortex*, *31*(1), 603–619. https://doi.org/10.1093/cercor/bhaa246

Roth, Z. N., Ryoo, M., & Merriam, E. P. (2020). Task-related activity in human visual cortex. *PLoS Biology*, *18*(11), e3000921. https://doi.org/10.1371/journal.pbio.3000921

Sørensen, T. J. (1948). *A method of establishing groups of equal amplitude in plant sociology*

based on similarity of species content and its application to analyses of the vegetation on danish commons. I kommission hos E. Munksgaard.

Stone, H. L., Mitchell, J. L., Fuentes-Jimenez, M., Tran, J. E., Yeatman, J. D., & Yablonski, M. (2025). Anatomically distinct regions in the inferior frontal cortex are modulated by task and reading skill. *The Journal of Neuroscience*, e1767242025. https://doi.org/10.1523/JNEUROSCI.1767-24.2025

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. https://doi.org/10.1016/j.neuron.2012.08.026

van den Hurk, J., Van Baelen, M., & Op de Beeck, H. P. (2017). Development of visual category selectivity in ventral visual cortex does not require visual experience. *Proceedings of the National Academy of Sciences*, *114*(22), E4501–E4510. https://doi.org/10.1073/pnas.1612862114

Vogel, A. C., Church, J. A., Power, J. D., Miezin, F. M., Petersen, S. E., & Schlaggar, B. L. (2013). Functional network architecture of reading-related regions across development. *Brain and Language*, *125*(2), 231–243. https://doi.org/10.1016/j.bandl.2012.12.016

Vogel, A. C., Miezin, F. M., Petersen, S. E., & Schlaggar, B. L. (2012). The Putative Visual Word Form Area Is Functionally Connected to the Dorsal Attention Network. *Cerebral Cortex*, *22*(3), 537–549. https://doi.org/10.1093/cercor/bhr100

Vogel, A. C., Petersen, S. E., & Schlaggar, B. L. (2012). The Left Occipitotemporal Cortex Does Not Show Preferential Activity for Words. *Cerebral Cortex*, *22*(12), 2715–2732. https://doi.org/10.1093/cercor/bhr295

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*, S0960982223001823. https://doi.org/10.1016/j.cub.2023.02.042

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. https://doi.org/10.1073/pnas.1822137116

Yeatman, J. D., Rauschecker, A. M., & Wandell, B. A. (2013). Anatomy of the visual word form area: Adjacent cortical circuits and long-range white matter connections. *Brain and Language*, *125*(2), 146–155. https://doi.org/10.1016/j.bandl.2012.04.010

Yeatman, J. D., & White, A. L. (2021). Reading: The Confluence of Vision and Language. *Annual Review of Vision Science*, *7*(1), 487–517. https://doi.org/10.1146/annurev-vision-093019-113509