ORIGINAL ARTICLE

Learning to Read Increases the Informativeness of Distributed Ventral Temporal Responses

Marisa Nordt¹, Jesse Gomez², Vaidehi Natu³, Brianna Jeska³, Michael Barnett³ and Kalanit Grill-Spector^{2,3,4}

¹Department of Developmental Neuropsychology, Ruhr-Universität Bochum, 44801 Bochum, Germany,

²Neurosciences Program, Stanford University School of Medicine, Stanford, CA 94305, USA, ³Psychology

Department, Stanford University, Stanford, CA 94305, USA and ⁴Stanford Neurosciences Institute, Stanford University, Stanford, CA 94305, USA

Address correspondence to Kalanit Grill-Spector, Jordan Hall, Bldg. 420, Department of Psychology, Stanford Neurosciences Institute, Stanford, CA 94305-2130, USA. Email: kalanit@stanford.edu

Abstract

Becoming a proficient reader requires substantial learning over many years. However, it is unknown how learning to read affects development of distributed visual representations across human ventral temporal cortex (VTC). Using fMRI and a data-driven, computational approach, we quantified the development of distributed VTC responses to characters (pseudowords and numbers) versus other domains in children, preteens, and adults. Results reveal anatomical- and hemisphere-specific development. With development, distributed responses to words and characters became more distinctive and informative in lateral but not medial VTC, and in the left but not right hemisphere. While the development of voxels with both positive and negative preference to words affected distributed information, only development of voxels with positive preference to words (i.e., word-selective) was correlated with reading ability. These data show that developmental increases in informativeness of distributed left lateral VTC responses are related to proficient reading and have important implications for both developmental theories and for elucidating neural mechanisms of reading disabilities.

Key words: development, human ventral temporal cortex, MVPA, neuroimaging, reading

Introduction

Reading is a unique human ability that is learned. Each year, over a quarter of a billion children across the globe attend primary school (grades 1–6) whose central mission is literacy instruction. Reading entails visual processing of letters and words, associating these visual inputs with sounds and language, and extracting their meaning. Thus, reading involves a network of inferior parietal, temporal, and frontal brain regions involved in vision, audition, and language. Prior research has examined how learning to read affects properties of white matter tracts that connect regions of the reading network (Ben-Shachar et al. 2007; Schlaggar and McCandliss 2007; Carreiras et al. 2009; Wandell et al. 2012; Yeatman et al. 2012; Saygin et al.

2016). Additionally, the discovery of the visual word form area (VWFA, Cohen et al. 2000; Dehaene et al. 2002), a region in ventral temporal cortex (VTC) that responds more strongly to words and characters than other visual stimuli has led researchers to examine the development of this region during childhood (Brem et al. 2010; Ben-Shachar et al. 2011; Cantlon et al. 2011; Dehaene-Lambertz et al. 2018) and during literacy acquisition in adulthood (Dehaene et al. 2010). Indeed, several studies have shown that responses and selectivity to words in the VWFA increase with reading acquisition (Brem et al. 2010; Dehaene et al. 2015; Saygin et al. 2016; Dehaene-Lambertz et al. 2018).

Despite the role of the VWFA for reading, it is important to note that the VWFA is a region within a larger anatomical

expanse, VTC, that is involved in perceiving and recognizing not only characters but also shapes, objects, and visual categories (Cohen et al. 2000; Dehaene et al. 2002; Rauschecker et al. 2011; Grill-Spector and Weiner 2014; Hannagan et al. 2015; Dehaene-Lambertz et al. 2018). In particular, a large body of research has shown that distributed responses across VTC have a characteristic pattern that represents the category of the visual input (Haxby et al. 2001; Kriegeskorte 2008; Carlson et al. 2013; Jacques et al. 2016; Golarai et al. 2017). These patterns of response are reproducible across items of a category, and distinct for items from different categories. In fact, independent classifiers can decode from distributed VTC responses in a participant's brain what is the category of the stimulus they are observing (Haxby et al. 2001; Cox and Savoy 2003; Weiner and Grill-Spector 2010; Grill-Spector and Weiner 2014). However, it is unknown if distributed representations of words and characters across VTC develop as children learn to read, and if cortical developments of distributed responses have behavioral ramifications.

To address these gaps in knowledge, we examined 3 questions in school-aged children and young adults.

(1) *If and how does learning to read affect distributed responses across the VTC?* We considered two main developmental hypotheses. One hypothesis predicts that learning to read leads to the emergence of new distributed representations of words across VTC. In other words, this hypothesis suggests that extensive experience and learning of visually presented words during reading acquisition will lead to the emergence of a new representation to words in the form of a reliable distributed pattern of activity across VTC. This hypothesis predicts that distributed VTC responses to words and characters will become more distinct and informative from childhood to adulthood. A second hypothesis predicts that distributed representations for words and characters in VTC may not be different across children and adults for two reasons. First, both children and adults see characters in their natural environment. Thus, by the age of 5, distributed VTC representations to characters may be fully developed, as has been reported for other domains such as places and objects (Golarai et al. 2007, 2017). Second, development of reading ability in childhood may involve neural changes outside high-level visual cortex such as changes in phonologically-mediated processing and increased engagement of left prefrontal cortex (Kovelman et al. 2012).

(2) *Is there anatomical specificity to the development of distributed VTC responses to words and characters?* The theory of object form topography (Haxby et al. 2001) suggests that visual category information is obtained by distributed responses across the entire VTC (Haxby et al. 2001; Cox and Savoy 2003; Kriegeskorte 2008; Connolly et al. 2012; Carlson et al. 2013). This theory predicts no anatomical specificity within VTC to the development of distributed responses. In contrast, eccentricity bias theory (Levy et al. 2001; Hasson et al. 2002; Malach et al. 2002), suggests that reading requires fine-grain visual acuity afforded by foveal vision. This theory predicts that foveation on words during reading will lead to the development of word representations in cortical regions with a pre-existing foveal bias (higher responses to central than peripheral stimuli). In both children and adults regions lateral to the mid-fusiform sulcus (MFS, Weiner et al., 2014) are foveally-biased (Levy et al. 2001; Hasson et al. 2002; Weiner et al. 2014) and regions medial to the MFS are peripherally-biased (Levy et al. 2001; Hasson et al. 2002; Weiner et al. 2014). Thus, eccentricity bias predicts that learning to read will lead to development of distributed responses in foveally-biased lateral VTC, but not peripherally-biased medial VTC. A third theory of domain specificity suggests that visual

processing of words is accomplished by a specific region, the VWFA, as (i) it responds significantly more strongly to characters than other stimuli (Dehaene et al. 2002), (ii) it is causally involved in processing words (Gaillard et al. 2006), and (iii) it shows developmental increases in response amplitude to letters and (pseudo)words (Ben-Shachar et al. 2011; Cantlon et al. 2011). Thus, domain-specificity predicts that learning to read will induce an even more anatomically-specific development of distributed responses restricted to just word-selective voxels rather than the entire lateral VTC.

(3) *Does development of distributed VTC responses have behavioral ramifications?* One possibility is that the development of distributed VTC responses is linked to improvements in reading ability, predicting a positive correlation between reading ability and information in distributed VTC responses to words and characters. If such a correlation exists, it would be critical to determine if anatomically-specific compartments of VTC predict reading ability. Alternatively, development of reading ability may depend on white matter connections between VTC with downstream areas (Yeatman et al. 2012; Gullick and Booth 2015; Takeuchi et al. 2016) rather than distributed VTC responses. This alternative predicts no relationship between reading ability and development of distributed VTC responses.

To address these questions, we conducted fMRI in 3 age groups: 12 children (ages, 5–9 years; 11 female), 13 preteens (ages, 10–12 years; 6 female), and 26 adults (ages, 22–28 years; 10 female). During scanning, participants viewed images of characters (pseudowords and uncommon numbers) and items from 4 other domains, each consisting of two categories (Fig. 1a). Participants viewed both words and numbers to allow distinguishing if development is general to the domain of characters or is specific to words. We used pseudowords, which are enunciable but lack meaning, for two reasons: (i) to control for age-related differences in semantic knowledge and (ii) to control familiarity across domains, as items from other domains were also unfamiliar.

To test developmental hypotheses, we measured distributed responses to items from each category in each of the medial and lateral VTC compartments and examined: (i) if there are age-related differences in the information and discriminability of distributed VTC responses to words and characters, (ii) if development of distributed responses shows anatomical specificity, and (iii) if development of distributed VTC responses is linked to reading ability.

Materials and Methods

Participants

Sixty-six participants including 20 children (ages 5–9 years), 15 preteens (ages 10–12 years), and 31 adults (ages, 22–28 years) participated in this study. Data of 5 children, 1 preteen, and 5 adults were excluded because participants had motion values larger than 2.1 voxels in one or more of the 3 runs; data of 2 children were excluded because they completed less than 3 fMRI runs; and data from 1 child and 1 preteen were excluded because they fell asleep during fMRI. In total, we report data from 51 participants including 12 children (ages 5–9 years; 11 female), 13 preteens (ages 10–12 years; 6 female), and 26 adults (ages, 22–28 years; 10 female). Participants had normal or corrected-to normal vision and were healthy and neurotypical. The study protocol was approved by the Stanford Internal Review Board on Human Subjects Research. Adult participants and parents of the participating children gave written consent to study participation and children gave written assent.

Figure 1. Stimuli, anatomical ROI definitions, and between age-group controls. (a) Examples of stimuli corresponding to the 5 domains (in columns) with two categories per domain (rows). (b) Examples of lateral and medial ventral temporal cortex (VTC) on the inflated cortical surface of a representative 6-year-old child (left), 10-year-old child (middle), and 22-year-old adult (right). Yellow outline: medial VTC; green outline: lateral VTC. (c) Boxplots showing within-run motion (left) and between-run motion (right) for each of the 3 age groups. Light blue: children (5–9 year old, $n = 12$); blue: preteen (10–12 year old, $n = 13$); orange: adults (22–26 year old, $n = 26$). There were no significant differences across age groups. (d) Boxplot of the number of functional voxels in lateral (left) and medial (right) VTC averaged across hemispheres. Each functional voxel is 2.4 mm on a side. Same subjects as in (c). (e) Boxplot showing the number of word-selective voxels in lateral (left) and medial (right) VTC averaged across hemispheres. Number of subjects and voxel size same as (d). In (c)–(e) the box indicates 25–75% quartiles of the data; Horizontal lines in the box plots: median; whiskers: data range excluding outliers, encompassing 99.3% of the data; black dots: outliers, values that are more than 1.5 times the interquartile range away from the top or bottom of the box.

Participants underwent anatomical and functional MRIs. Prior to MRI, children were trained in a scanner simulator. Children were invited to MRI sessions only after having successfully completed the scanner simulator training. After completion of MRIs, subjects participated in behavioral tests outside the scanner. Different measurements were performed on different days.

MRI Data Acquisition

MRI data were collected at the Stanford Center for Cognitive and Neurobiological Imaging using a 3T Signa scanner

(GE Healthcare) and a custom-built phase-array 32-channel receive-only head coil.

Anatomical MRI

Whole-brain, high-resolution anatomical scans were acquired using T1-weighted quantitative MRI (qMRI, Mezer et al., 2013), using a spoiled gradient echo sequence with multiple flip angles ($\alpha = 4^\circ, 10^\circ, 20^\circ$, and 30° ; TR = 14 ms; TE = 2.4 ms). Voxel size = $0.8 \text{ mm} \times 0.8 \text{ mm} \times 1 \text{ mm}$, resampled to 1 mm^3 isotropic. Additionally, we acquired T1-calibration scans using spin-echo inversion recovery with an echo-planar imaging read-out,

spectral spatial fat suppression, and a slab inversion pulse (TR = 3 s, echo time=minimum full, 2× acceleration, inplane resolution = 2 mm²; slice thickness = 4 mm).

Functional MRI

fMRI data were obtained with a multi-slice EPI sequence (multiplexing factor=3; 48 slices oriented parallel to the parieto-occipital sulcus; TR = 1 s; TE = 30 ms; flip angle = 76°; FOV = 192 mm; 2.4 mm isotropic voxels; one-shot T2*-sensitive gradient echo sequence).

fMRI 5 Domain/10 Category Experiment

Participants completed 3 runs of the fMRI experiment. Each run lasted 5 min and 24 s. During fMRI participants viewed images from 5 domains, each consisting of two categories: characters (pseudowords and numbers), faces (adult and child faces), bodies (headless bodies and limbs), objects (cars and guitars), and places (houses and corridors) as in our prior studies (Stigliani et al. 2015). Pseudowords are the same as in (Glezer et al. 2009, 2015) and have similar bi-gram and trigram frequency as typical English words. Images were grayscale and contained a phase-scrambled background generated from randomly selected images (Fig. 1a). Images were presented at a rate of 2 Hz, in 4 s trials, and did not repeat. Image trials were intermixed with gray luminance screen baseline trials. Trials were counterbalanced across categories and baseline.

Task: Participants were instructed to view the images as they fixated on a central dot, and press a button when an image with only the phase-scrambled background appeared. These images appeared randomly 0, 1, or 2 times within a trial.

Data Analysis

Data were analyzed using MATLAB 2012b and mrVista (<http://github.com/vistalab>) as in previous publications (Stigliani et al. 2015; Natu et al. 2016; Gomez et al. 2017).

Anatomical Data Analysis

An artificial T1-weighted anatomy was generated from qMRI data using mrQ (<https://github.com/mezera/mrQ>). Brain anatomy was segmented into gray-white matter with FreeSurfer 5.3 (<https://surfer.nmr.mgh.harvard.edu/>), and manually-corrected to generate cortical surface reconstructions of each participant.

Anatomical Definition of Lateral and Medial Ventral Temporal Cortex

Lateral and medial ventral temporal cortex (VTC) were individually defined on each participant's inflated cortical surface in each hemisphere as in previous publications (Weiner and Grill-Spector 2010) (Fig. 1b). VTC definition: *anterior border*: anterior tip of the mid-fusiform sulcus (MFS) which aligned with the posterior end of the hippocampus; *posterior border*: posterior transverse collateral sulcus (ptCoS); *lateral VTC*: extended from the inferior temporal gyrus (ITG) to the MFS; *medial VTC*: extended from the MFS to the medial border of the collateral sulcus (CoS). VTC ROIs were drawn by BJ, divided into the lateral and medial part by MN, and checked by KGS.

fMRI Data Analysis

Functional data were aligned to each participant's native brain anatomy using rigid body transformation, no template was used. Data were not spatially smoothed and no slice-timing correction was performed. We motion-corrected data within a run and then across runs. Subjects with 3 runs with motion < 2.1 voxels were included in the analysis. After exclusion of 8 children, 2 preteens, and 5 adults, there were no significant differences in motion either within or between runs across age groups (Fig. 1c). The time courses of each voxel were transformed into percentage signal change by dividing each time point of each voxel's data by the average response across the entire run. To estimate the contribution of each of the 10 conditions a general linear model (GLM) was fit to each voxel by convolving the stimulus presentation design with the hemodynamic function (HRF) implemented in SPM (www.fil.ion.ucl.ac.uk/spm). Motion parameters were not incorporated into the GLM.

Multivoxel Pattern Analysis

In each anatomical ROI, multivoxel patterns (MVPs) for each category were represented as a vector of response amplitudes estimated from the GLM in each voxel. These values were transformed to z-scores by subtracting in each voxel its mean voxel response and dividing by: $\sqrt{\text{residual GLM variance}/df}$ (df = degrees of freedom). To evaluate similarity between MVPs we calculated all pair-wise correlations between MVP pairs resulting in a 10 × 10 cross-covariance matrix, referred to as representational similarity matrix (RSM, Kriegeskorte 2008). Each cell in the RSM reflects the average correlation among 3 permutations of MVP pairs (run 1&2, run 2&3, and run 1&3).

Winner-take-all Classifier

To evaluate category information in MVPs we used an independent winner-take-all (WTA) classifier. We implemented two versions of the WTA classifier. The first, evaluated domain information (characters/faces/bodies/objects/places, chance level 20%). The second, evaluated category information (pseudowords, numbers, adult faces, child faces, headless bodies, limbs, cars, guitars, corridors, buildings, change level 10%). The classifier was trained in each subject and ROI with data from one run and tested how well it predicted the stimulus of interest the subject viewed from MVPs from each of the other two runs. This resulted in 6 training and testing combinations per condition. We averaged across these combinations for each subject, and then averaged across subjects of each age group.

Information in Selective and Non-selective Voxels in Lateral VTC

We compared classification performance in selective and non-selective voxels using two analyses (1) lateral VTC character-selective (words+numbers > faces, bodies, objects, places, $t > 3$, voxel level) versus the remaining voxels, which we refer to as non-character-selective voxels, and (2) lateral VTC word-selective (words > numbers, faces, bodies, objects, places, $t > 3$, voxel level) versus the remaining voxels which we refer to as non-word-selective voxels.

Information in Systematically Increasing Proportions of Lateral VTC Voxels

We tested how the number of voxels within lateral VTC voxels affects classification in two analyses using different voxel sortings. Analyses were performed for both word classification and character classification as described above. Sorting 1: voxels were sorted by *selectivity* – from highest to lowest *t-value* for the relevant contrast. Sorting 2: voxels were sorted by *distinctiveness* – from highest to lowest *absolute t-value* for the relevant contrast. After sorting of voxels, analyses were identical: We calculated classification performance for increasing portions of lateral VTC voxels according to each sorting, starting with 10% of voxels, increasing by increments of 10%, up to all voxels.

To compare classification across age groups, we fitted each subject's classification performance as a function of number of voxels using a quadratic function, then determined its maximum. Statistical analyses determined if the maximal classification significantly varied across age groups using rmANOVAs on the estimated classification maxima. Due to numerical fitting, the estimated maximum could exceed 100% even as classification performance cannot exceed 100%. Thus, we performed a control analysis in which we rectified the maximum value to 100%. Results did not significantly differ from the original analysis.

Assessing Reading Ability

A subset of 7 children, 11 preteens, and 19 adults also completed the *word identification* and *word attack* tests from the Woodcock Reading Mastery Test (WRMT) outside the scanner. In the *word identification* task, participants were instructed to read a list of words as accurately as possible. In the *word attack* task, were instructed to read a list of pseudowords as accurately as possible. Tests do not have a time limit, but end when the participant makes 4 consecutive errors or has completed reading the list.

$$\text{Reading score: } 100 \frac{\text{words read correctly}}{\text{total number of words}}$$

Relating Reading Ability to Word and Character Information in VTC

We measured if there was a significant correlation between participant's reading ability as measured by the WRMT and information in lateral VTC. Correlations were measured between each reading test score (word identification/word attack test) and each classification (character/word) from lateral VTC. Significant correlations were followed with a subsequent analysis in which age was included as a factor. We report if correlations remained significant if age is partialled out of the correlation.

Since analyses of information in lateral VTC revealed that a subset of voxels in VTC contribute to either character or word information, we performed correlation analyses between reading and subsets of lateral VTC voxels: (1) entire lateral VTC, (2) 30% of most discriminative lateral VTC voxels, which yielded best word classification (Fig. 4), and (3) the remaining non-word-distinctive lateral VTC voxels (Fig. 4). The same analyses were done for 30% of lateral VTC voxels that were most discriminative for characters (Fig. 4, Fig. S4). We evaluated if the correlations in (2) and (3) were significantly different and

if information in most or least selective voxels correlated with reading ability.

Retinotopic Mapping Experiment

V1 was defined using data from a separate retinotopic mapping experiment (for a detailed description see, [Gomez et al. 2018](#)). A subset of participants comprising 8 children ages 5–9, 12 children ages 10–12, and 19 adults took part in a retinotopic mapping experiment using black and white checkerboard bars (width = 2° of visual angle, length = 14°) which changed contrast at a rate of 2 Hz, that swept across the screen. During retinotopic mapping subjects were instructed to fixate on a central stimulus and perform a color exchange task. Subjects' fixations were monitored with an eye tracker. Checkerboard bars swept the visual field in 8 different configurations in each run (4 orientations: 0°, 45°, 90°, 135°, each orientation was swept in 2 directions that were orthogonal to the bar). Same as [Dumoulin and Wandell \(2008\)](#); [Gomez et al. \(2018\)](#). We used checkerboard stimuli as they are the most ubiquitous stimuli that is used for population receptive field (pRF) mapping, and do not require cognitive processing that may differ across age groups. Subjects participated in 4 such runs, each run lasted 3 min and 24 s. Retinotopic data were collected on the same scanner as the main experiment and at the same resolution, with a 16-channel surface coil, acceleration factor × 2, and 28 slices.

Statistical Analyses

Unless otherwise noted, statistical analyses included the whole sample (5–9, *n* = 12; 10–12, *n* = 13; 22–28, *n* = 26). Statistical analyses were done using MATLAB 2015a. Outliers in boxplots are defined as values that were more than 1.5 times away from the interquartile range from the top or bottom of the box and are indicated by black dots.

Analyses Related to Figure 1

In analyses related to Figure 1c we tested if age groups differed significantly in the amount of motion during scanning using a repeated measures analysis of variance (rmANOVA) with factors of age group (5–9/10–12/22–28) and motion type (within-run motion, between-run motion). Similarly, in analyses related to Figure 1d we tested in each partition (lateral VTC, medial VTC) if there were statistically significant between group differences in the number of voxels using ANOVAs with the factor age group (5–9/10–12/22–28). The same procedure was applied for analyses related to Figure 1e, in which we tested the statistical significance of the number of word-selective voxels across groups.

Analyses Related to Figure 2

In analyses related to Figure 2 we first tested if classification performance for characters was significantly different from chance level (20%) in each age group using one-sample *t*-tests. Next, we tested for significant differences in classification performance using a 3-way rmANOVA with factors of age group (5–9/10–12/22–28), partition (lateral VTC/medial VTC), and hemisphere (left/right). Similar analyses were run for classification performance for character types, i.e., for pseudoword and number classification. Here, classification performance was tested against the chance level of 10%. Significant differences in decoding performance were tested using a 4-way rmANOVA with factors of age group (5–9/10–12/22–28), character type

(numbers/pseudowords), partition (lateral VTC/medial VTC), and hemisphere (left/right). To follow up on significant interactions between character type and age group we conducted separate 3-way rmANOVAs on number and pseudoword classification with factors of age group (5–9/10–12/22–28), partition (lateral VTC/medial VTC) and hemisphere (left/right). To further follow up on the significant effect of age, we directly compared classification performance for words across age groups using post-hoc t-tests.

Analyses related to Figure 3

In analyses related to Figure 3 we first tested if the mean within-domain correlations for pseudowords and numbers across partitions and hemispheres were significantly different from zero using one-sample t-tests. Next, we tested for significant differences of correlations using 3 rmANOVAs. The rmANOVAs tested correlations (i) within the domain of characters (words – words (w-w); and numbers–numbers (n-n)) with factors of age group (5–9/10–12/22–28), partition (lateral VTC/medial VTC), hemisphere (left/right), and character type (w-w/n-n), (ii) across character types (w-n), with factors of age group (5–9/10–12/22–28), partition (lateral VTC/medial VTC), and hemisphere (left/right) and (iii) between-domains (words – non-words, (w-nw); and numbers–non-numbers, (n-nn)) including the factors of age group (5–9/10–12/22–28), partition (lateral VTC/medial VTC), hemisphere (left/right), and character type (w-nw/n-nn). Furthermore, we tested for each character type in each partition and hemisphere if there was a significant effect of age group (1-way ANOVAs with the factor of age group (5–9/10–12/22–28)) to follow up on significant interactions revealed in rmANOVAs measured above. Significant effects of age are shown in Figure 3 with asterisks.

Analyses Related to Figure 4

In analyses related to Figure 4a we tested differences in word classification in lateral VTC in word-selective and non-selective voxels using a 3-way rmANOVA with factors of age group (5–9/10–12/22–28), hemisphere (left/right), and voxel type (selective/non-selective).

In analyses related to Figure 4b we compared the estimated maximal classification performance by populations of voxels sorted by selectivity using a 2-way rmANOVA with factors of age group (5–9/10–12/22–28) and hemisphere (left/right). Same analyses were conducted on classification performance on voxels sorted voxels by distinctiveness for the data shown in Figure 4c. Corresponding analyses were performed for character classification in character-selective and non-selective voxels in Figure 4d–f. We further tested if word information developed also in the remainder of non-selective/non-discriminative voxels. Thus, we ran an additional rmANOVA with factors of age group (5–9/10–12/22–28) and hemisphere (left/right) on word decoding in the remaining voxels (i.e., the lateral VTC voxels except the set of voxels which generated the maximal classification). In 4 subjects, classification performance was identical for all voxel set sizes exceeding 10% of the lateral VTC. For these subjects, we included the minimal set (10% of voxels) as the ones achieving maximal classification, and tested classification in the remaining 90% of voxels.

Analyses Related to Figure 5

In analyses related to Figure 5 we tested if reading scores (WRMT) are correlated with classification performance for words

and/or characters by calculating the Pearson correlation coefficient between these values; We report both the raw P-value and Bonferroni-corrected P-values. If correlations were found to be significant, we performed an additional partial correlation analysis to control for the effect of age. We also tested if correlations between reading scores and classification in different subsets of lateral VTC voxels (30% of most discriminative lateral VTC voxels vs. the remaining non-discriminative lateral VTC voxels) differed significantly using (<http://quantpsy.org/corrttest/corrttest2.htm>), which includes converting the correlation coefficients into z-scores using Fisher's r-to-z transformation.

Analyses Testing for an Association Between Face and Word Specialization (Discussion)

To find out if development of word (or character) information was related to development of face information we conducted two analyses: First, we tested if there was a significant correlation between word (or character) classification in the left hemisphere and face classification in the right hemisphere. Second, we examined if there was a correlation between laterality of face information and laterality of word information. To obtain measures of laterality we used the following calculations:

- Face classification laterality index =

$$\frac{[\text{face classification (right)} - \text{face classification (left)}]}{[\text{face classification (right)} + \text{face classification (left)}]}$$

- Word classification laterality index =

$$\frac{[\text{word classification (left)} - \text{word classification (right)}]}{[\text{word classification (left)} + \text{word classification (right)}]}$$

These analyses were performed both on the level of categories (word information and child face information) and on the level of domains (character information and face information). For both types of analyses we calculated correlations and partial correlations that were corrected for the influence of age (r_p). These results revealed non-significant results, which we refer to in the discussion.

Analysis of V1 MVPs

After fitting the pRF model (Dumoulin and Wandell 2008) with a compressive spatial summation (Kay et al. 2013) in each voxel, maps of pRF phase and eccentricity were projected onto the inflated cortical surface of each subject's brain. V1 was defined in each hemisphere as the visual field map containing a hemi-field representation in which the horizontal meridian occupied the calcarine sulcus, the lower vertical meridian occupied the upper lip of the calcarine and the upper visual meridian occupied the lower lip of the calcarine.

After defining V1 we performed MVPA and classification analyses on V1 of each hemisphere. This analysis tested whether category information in VTC was beyond low-level information that could be extracted from V1 MVPs. Results shown in Figure S1b indicate that the domain information in VTC is significantly higher than V1 (main effect of area, $F(2,72) = 142.1$, $P < 0.001$; lateral VTC > V1, $t(38) = 16.32$, $P < 0.001$; medial VTC > V1, $t(38) = 10.13$, $P < 0.001$).

Data and Software Availability

All code relevant to data analysis for the main findings (Figures 1–5) will be available on github.com/VPNL/Word_Development. Any source data relevant to these analyses will also be made available upon request. The majority of the code used in this study was derived from scripts and functions available through the open-source vistasoftware code library: <https://github.com/vistalab/vistasoft>.

Results

There were no significant differences across age groups in (i) motion during fMRI (Fig. 1c, $F(2,48) = 2.67$, $P = 0.08$), (ii) the number of voxels in anatomical partitions of VTC (Fig. 1d, all $F_s \leq 0.86$, $P_s > 0.42$, analysis of variance (ANOVA) with the factor age group) and (iii) the number of word-selective voxels (Fig. 1e, all $F_s \leq 0.71$, $P_s > 0.49$, ANOVAs with factor age group). These analyses show that data quality and the anatomical size of VTC is similar across age groups.

Does Information of Distributed VTC Responses for Characters, Words, and Numbers Develop After Age 5?

We used a decoding approach to quantify if there are developmental changes either in decoding characters, words, or numbers from distributed VTC responses. Thus, we quantified using a winner-take-all classifier three types of information in

distributed VTC responses: (i) characters (pseudowords+numbers) versus other domains (faces, places, objects, and body parts), (ii) pseudowords versus the other 9 categories (adult faces, child faces, houses, corridors, cars, guitars, bodies, limbs, Fig. 1a), and (iii) numbers versus the other 9 categories.

Results show that in all age groups, VTC partitions, and hemispheres, decoding character information was significantly higher than the 20% chance level (Fig. 2a, all $P_s \leq 0.005$). Notably, decoding character information significantly increased from age 5 to adulthood (Fig. 2a, main effect of age group, $F(2,48) = 5.74$, $P = 0.006$, 3-way rmANOVA with factors of age group, VTC partition, hemisphere).

Critically, this development was anatomically specific: character information was better decoded from multivoxel patterns (MVPs) across lateral VTC compared to medial VTC (main effect of VTC partition, $F(1,48) = 113.58$, $P < 0.001$), and also better decoded from left than right hemisphere MVPs (main effect of hemisphere, $F(1,48) = 40.17$, $P < 0.001$, no significant interactions, all $P_s > 0.066$). The largest development was observed in left lateral VTC in which decoding of character information improved on average by 25% from age 5 to 25. Specifically, decoding characters versus other stimuli from left lateral VTC yielded an accuracy of $65 \pm 6\%$ (mean $\pm$ SEM) in children, but reached a $90 \pm 3\%$ accuracy in adults (significantly higher than 5–9 year-olds, post-hoc t-test, $t(36) = 4.2$, $P < 0.001$, Fig. 2a).

In contrast to the development of character information, we did not find a significant development of information for the

Figure 2. Differential development of word and number information after age 5. (a) Character classification performance in lateral VTC (top) and medial VTC (bottom) for the left (black) and right (gray) hemispheres across age groups. Classification performance was quantified with a winner-take-all (WTA) classifier. Data show mean performance for children (5–9-year-olds, $n = 12$), preteens (10–12-year-olds, $n = 13$), and adults (22–28-year-olds, $n = 26$). Error bars: standard error of the mean (SEM). Chance level is 20% (horizontal gray line). (b) WTA classification of character type (either pseudowords or numbers) in lateral VTC (top) and medial VTC (bottom). Chance level is 10%. Same conventions as in (a). See also Figures S1 and S2.

domains of bodies, objects, and places in VTC (Fig. S1a). However, we found a significant development of face information in VTC, consistent with prior research (Golarai et al. 2017). Critically, decoding of domain information was not due to low-level features, as decoding domain information from VTC was significantly higher than from V1 (Fig. S1b, main effect of ROI, $F(2,72) = 142.1$, $P < 0.001$, rmANOVA with the factors group and ROI (lateral VTC, medial VTC, V1) in a subset of subjects with V1 defined retinotopically (5–9 year-olds, $n = 8$, 10–12 year-olds, $n = 12$, 22–28 year-olds, $n = 19$). Together, these analyses suggest that information about some domains is adult-like in VTC in 5–9 year olds, even as character information continues to develop.

Surprisingly, analysis of pseudoword and number decoding revealed a differential development of word and number information across VTC partitions (Fig. 2b, age group $\times$ character type $\times$ VTC partition interaction, $F(2,48) = 3.93$, $P = 0.03$, 4-way rmANOVA with factors of age group, character type, partition, hemisphere), as well as a differential development of word and number information across hemispheres (age group $\times$ character type $\times$ hemisphere interaction, $F(2,48) = 5.10$, $P = 0.01$).

That is, pseudoword information specifically developed in the left lateral VTC, even as decoding pseudowords was significantly higher than chance in all age groups (all P s < 0.003). Notably, decoding of pseudowords from left lateral VTC progressively increased from $40 \pm 7\%$ in 5–9 year-olds (significantly lower than adults, post-hoc t -test, $t(36) = -4.47$, $P < .001$), to $55 \pm 8\%$ in 10–12 year-olds (significantly lower than adults, post-hoc t -test, $t(37) = -2.70$, $P = 0.01$), to $78 \pm 5\%$ in 22–28 year-olds (Fig. 2b – top left). This reflects an almost 2-fold improvement in decoding word information from left lateral VTC from age 5 to adulthood. In contrast, in the right lateral VTC decoding pseudowords was not significantly different across age groups and overall lower than left lateral VTC, averaging at an accuracy of $31 \pm 4\%$ (not significantly different across age groups, all $P > 0.06$, Fig. 2b – top left). Additionally, there were no significant differences across age groups in decoding pseudowords from medial VTC (all P s > 0.14 , Fig. 2b – bottom left), and performance was around 30% in the left hemisphere and only around 13% in the right hemisphere.

In contrast to the development of word information, there was no significant development of number information in any partition or hemisphere (no significant effects of age group or interactions with age group, all F s < 1.03 , P s > 0.36 , 3-way rmANOVA on number classification with the factors age group, partition, hemisphere). Indeed, number decoding averaged about $30 \pm 4\%$ across age groups, partitions, and hemispheres (Fig. 2b – right). Additionally, there was no significant development of information for other categories in VTC, except for corridors (Fig. S2). Thus, information in VTC for 8 out of 10 categories remained stable from childhood to adulthood. This provides evidence that the increased information for pseudowords from lateral VTC responses is specific, and does not reflect a general developmental increase in category information across VTC.

Taken together, these analyses reveal two important findings. First, we find evidence for development of both character and word information in VTC. Second, the development of character information occurs across the lateral VTC in both hemispheres, but development of word information is largely restricted to left lateral VTC.

Is the Developmental Increase in Word and Character Information Driven by Changes in Within-domain Similarity or Between-domain Distinctiveness?

We next sought to determine which changes drive the development of word and character information. We considered three

hypotheses: First, the development might be due to an increase in the similarity between MVPs within the character domain. Second, the development might be due to a decrease in the dissimilarity (increase in distinctiveness) between MVPs of characters versus other domains. Or third, the development might be due to increases in both within-character-domain similarity and between-domain distinctiveness. Similarity was estimated by computing the Pearson correlation coefficient between MVPs to different items from different runs.

Results reveal three main findings. First, across age groups and VTC partitions, within-character-domain correlations between MVPs to pseudowords (w–w) and numbers (n–n) were positive (Fig. 3, P s < 0.001), indicating that MVPs generalize to other items within the category. Second, crucially, within-character-domain similarity of word MVPs (w–w) and number MVPs (n–n) systematically increased from age 5, to age 12, to adulthood (main effect of age, $F(2,48) = 4.16$, $P = 0.02$, 4-way rmANOVA with factors age group, hemisphere, VTC partition, character type). Third, development of the reliability of pseudoword and number MVPs was anatomically heterogeneous. Development significantly varied across hemispheres and VTC partitions (significant age group $\times$ hemisphere $\times$ VTC partition interaction, $F(2,48) = 4.21$, $P = 0.02$), as well as across hemispheres and character types (significant age group $\times$ character-type $\times$ hemisphere interaction, $F(2,48) = 3.52$, $P < 0.04$, Fig. 3, w–w/n–n).

The largest development of within-character-domain similarity of MVPs occurred for pseudowords in left lateral VTC, as compared to pseudowords or numbers in medial VTC or the right hemisphere. Indeed, similarity between pseudoword MVPs in left lateral VTC increased from a median value of 0.19 ± 0.18 (median $\pm$ interquartile range) in 5–9 year-olds to 0.44 ± 0.26 in adults (Fig. 3a, w–w). In contrast, similarity between pseudoword MVPs in left medial VTC only increased from 0.07 ± 0.20 to 0.13 ± 0.13 from age 5 to adulthood (Fig. 3c, w–w), and the similarity of distributed responses to numbers in left lateral VTC, increased from a median of 0.12 ± 0.20 to 0.23 ± 0.20 from age 5 to adulthood (Fig. 3a, n–n). In lateral VTC, pseudoword MVPs also became more similar to number MVPs from age 5 to adulthood (Fig. 3, w–n, significant age group $\times$ VTC partition interaction, $F(2,48) = 9.33$, $P < 0.001$, 4-way rmANOVA).

Evaluation of between-domain correlations revealed that in all age groups words and numbers generate MVPs that were distinct from items of other domains (Fig. S3 shows the entire representational similarity matrix). Indeed the correlation between MVPs to words (or numbers) versus items of other domains were negative (Fig. 3, w–nc/n–nc) and were significantly lower than the within-character-domain correlations (main effect of domain type, lateral VTC: $F(1,48) = 330.4$, $P < 0.001$; medial VTC: $F(1,48) = 157$, $P < 0.001$, rmANOVA with factors age group, hemisphere, domain type (within-domain/between-domain)).

Notably, MVPs to both words and numbers became significantly less correlated (or more dissimilar) from MVPs to other domains from age 5 to adulthood (Fig. 3, w–nc/n–nc, main effect of age, $F(2,48) = 4.57$, $P < 0.02$, 4-way rmANOVA with factors age group, VTC partition, hemisphere, character type). This developmental decrease of between-domain similarity of character MVPs versus other domains varied across VTC partitions (significant age group $\times$ VTC partition interaction, $F(2,48) = 9.79$, $P < 0.001$) as well as hemispheres (significant age group $\times$ VTC partition $\times$ hemisphere interaction, $F(2,48) = 3.96$, $P = 0.026$). Between-domain correlations significantly decreased in left lateral VTC, for both pseudowords and numbers (Fig. 3a, w–nc/n–

Figure 3. Within-domain similarity and between-domain distinctiveness of distributed responses to characters increase from age 5 to adulthood. Boxplots depict the Pearson correlation coefficient between multivoxel patterns (MVPs) of distributed responses within the domain of characters (w-w: pseudowords–pseudowords; n-n: numbers–numbers; w-n: pseudowords–numbers) and between domains (w-nc: pseudowords–non-characters (all stimuli except words and numbers); n-nc: numbers–non-characters). (a) Left lateral VTC; all comparisons showed a significant main effect of age as indicated by the asterisk ($F(2,48) > 3.34$; $P < 0.04$). (b) Right lateral VTC; w-n and n-nc showed a significant effect of age ($F(2,48) > 4.05$; $P < 0.02$). (c) Left medial VTC. (d) Right medial VTC. Boxplots are colored by age; light blue: 5–9-year-olds, $n = 12$; bright blue: 10–12-year-olds, $n = 13$, orange: 22–28-year-olds, $n = 26$. In the boxplots, horizontal lines indicate the median, whiskers correspond to approximately ± 2.7 standard deviations which captures 99.3% of the data, and extend to the most extreme value that is not an outlier. Asterisks: Significant main effect of age. See also Figure S3.

nc, both $F_s \geq 6.07$, $P_s \leq 0.005$) and in right lateral VTC for numbers (Fig. 3b, n-nc, $F(2,48) = 4.05$, $P = 0.02$).

Together, these analyses reveal that the increase in word and character information in lateral VTC appears to be driven by both increases in within-character-domain similarity as well as increases in dissimilarity (distinctiveness) across domains.

Which Voxels Drive Development of Distributed Responses for Words and Characters in Lateral VTC?

We observed that the largest development of word and character information is in the left lateral VTC. As the lateral VTC contains the visual word form area (VWFA), this raises the question whether the development of the VWFA, which responds more strongly to characters and words versus the other domains, is what drives the development of word and character information in lateral VTC. Alternatively, it may be that development of the entire lateral VTC including non-selective voxels increases word and character information.

To test these hypotheses, we examined classification performance separately for selective and non-selective voxels for pseudowords and characters within the lateral VTC, respectively. The first hypothesis predicts that information in selective voxels, but not non-selective voxels will increase across development. The second predicts that information in both types of voxels will increase. Selective voxels were defined in

each subject and hemisphere as in prior studies (Weiner and Grill-Spector 2010; Golarai et al. 2010, 2017) as voxels that responded more strongly to the preferred stimulus compared to other stimuli. In our analyses, we separately considered voxels in lateral VTC that were word-selective (pseudowords > non-words, $t > 3$, voxel-level), and character-selective (pseudowords + numbers > non-characters, $t > 3$, voxel level). Non-selective voxels were defined as the remaining lateral VTC voxels. We then examined if the WTA classifier can (i) classify word and character information from the selective and non-selective voxels and (ii) if there is differential development of information across the two voxel types.

Results are consistent with the second hypothesis. First, significant development of word information occurred in both selective and non-selective voxels (main effect of age, $F(2,42) = 5.39$, $P = 0.008$, 3-way rmANOVA with factors of age group, hemisphere, voxel type (selective/non-selective), and no age group $\times$ voxel type interaction, $F(2,42) = 0.26$, $P = 0.77$) (left hemisphere: Figure 4a; right hemisphere: Fig. S4a). Second, similar results were obtained for character classification (Fig. 4d, Fig. S4d, main effect of age, $F(2,47) = 12.79$, $P < 0.001$ 3-way rmANOVA with factors of age group, hemisphere, voxel type (selective/non-selective), and no age group $\times$ voxel type interaction, $F(2,47) = 2.13$, $P = 0.13$). However, conclusions from this analysis need to be considered with caution, as: (1) results may depend on the threshold used to define selectivity, (2) selective

Figure 4. Development of word and character information in selective, non-selective, and discriminative voxels. (a) Dark blue bars: Mean WTA classification performance of left lateral VTC word-selective voxels ($t > 3$). Light blue bars: Mean WTA classification performance of left lateral VTC voxels excluding the word-selective voxels ($t \leq 3$). Horizontal lines: mean (black) and SEM (gray) WTA classification performance from the entire left lateral VTC. Legend: number of voxels included in each analysis. Only subjects that had word-selective voxels are included (5–9-year-olds: $n = 9$; 10–12-year-olds: $n = 11$; 22–28-year-olds: $n = 25$). (b) Word classification performance as a function of percentage of left lateral VTC voxels sorted from most to least word-selective (descending t -value for the contrast pseudowords > non-words). Lines: mean performance; Shaded areas: SEM. (c) Word classification performance as a function of percentage of left lateral VTC voxels sorted by descending absolute t -value ($|t|$ -value) for the contrast pseudowords > non-words. (d–f) Same as a–c but for the contrast characters > non-characters. In d–e only subjects that had character-selective voxels are included (5–9-year-olds: $n = 11$; 10–12-year-olds: $n = 13$; 22–28-year-olds: $n = 26$). See also Figures S4 and S5.

and non-selective voxel subsets are of vastly different set sizes (Fig. 4a,d – legend), and (3) classification performance from these subsets is substantially lower than from all lateral VTC (Fig. 4a,d, gray lines). Therefore, we sought to conduct a principled analysis that compares information in a systematic manner across threshold levels and percentage of lateral VTC voxels.

First, we tested if selective voxels drive decoding performance using a flexible approach in which we systematically varied the threshold. Thus, we sorted each subject's lateral VTC voxels based on their selectivity to words (descending t -value for the contrast pseudowords > non-words) and tested classification performance as a function of number of voxels. Second, we tested if discriminative rather than selective voxels drive

decoding performance. We reasoned that not only voxels with positive selectivity to words, but also voxels with negative selectivity to words may be informative. Thus, we sorted each subject's lateral VTC voxels based on their absolute t -value (i.e., either positive or negative preference to words) from the greatest to least in magnitude.

Analysis by word selectivity revealed that in all age groups, word classification performance gradually increased as more voxels with progressively lower word-selectivity were added. Additionally, at all thresholds, classification in adults was higher than in children. Performance in both hemispheres was maximal when all lateral VTC voxels were included (Fig. 4b, Fig. S4b). Maximal decoding of pseudowords was significantly higher in adults compared to children (Fig. 4b, Fig. S4b, main

effect of age group, $F(2,48) = 5.34$, $P = 0.008$, 2-way rmANOVA with factors of age group, hemisphere on estimated maximum decoding, voxels sorted by selectivity) in the left (Fig. 4b), but not in the right hemisphere (Fig. S4b, age group $\times$ hemisphere interaction, $F(2,48) = 6.55$, $P = 0.003$).

Analysis by discriminability revealed that word classification from discriminative voxels yielded maximal performance using just ~35% of lateral VTC voxels (Fig. 4c, Fig. S4c). Across age groups and hemispheres performance plateaued for a range of 30–60% of voxels. Including additional voxels reduced performance.

Surprisingly, pseudoword classification from the entire lateral VTC was substantially lower than the maximal classification from the discriminative subset of voxels (all P s ≤ 0.005 , Fig. 4c, Fig. S4c). For example, in 5–9 year olds, maximal classification using $36 \pm 31\%$ of discriminative voxels was $77 \pm 6\%$, but performance dropped to $40 \pm 7\%$ using the entire left lateral VTC. The difference was even more pronounced in the right hemisphere: in 5–9 year olds maximal classification using $36 \pm 21\%$ of discriminative voxels was $64 \pm 8\%$, but performance dropped to $32 \pm 8\%$ for the entire right lateral VTC. The threshold of defining discriminative voxels (absolute t -value corresponding to the percentage of voxels that yielded maximal classification) did not differ significantly across age groups (no significant effect of age group, $F(2,44) = 0.15$, $P = 0.86$, 2-way rmANOVA on threshold absolute t -values with factors age group, hemisphere).

Critically, classification of word information from discriminative voxels developed: decoding was significantly higher in adults compared to children (Fig. 4c, Fig. S4c main effect of age, $F(2,48) = 7.46$, $P = 0.002$, 2-way rmANOVA on maximum word classification with factors age group, hemisphere). However, there was no significant development of word information in the remainder of lateral VTC voxels after the discriminative voxels that achieved highest classification were removed (no significant effect of age group, $F(2,44) = 0.48$, $P = 0.62$, rmANOVA with factor age group, hemisphere on non-discriminative voxels). Maximum classification from discriminative voxels was higher in adults compared to both 5–9 and 10–12 year-old children in both hemispheres (post-hoc t -tests, all $P < 0.02$). The number of discriminative voxels yielding maximal word classification differed across hemispheres in adults, but not in children (age group $\times$ hemisphere interaction, $F(2,37) = 5.05$, $P = 0.01$, 2-way rmANOVA). In adults, highest classification was achieved using $51 \pm 29\%$ of discriminative left lateral VTC voxels versus $24 \pm 17\%$ of right lateral VTC voxels.

We observed similar patterns of results when varying the percentage of character-selective (Fig. 4e, Fig. S4e) and character-discriminative voxels from lateral VTC (Fig. 4f, Fig. S4f). In contrast, there were no significant differences across age groups in decoding number information from lateral VTC using either number-selective or number-discriminative voxels (all F s < 1.5 , P s > 0.23 , Fig. S5).

These analyses lead to a surprising insight: development of word information results from developmental increases in the distinctiveness of distributed responses to pseudowords as compared to other stimuli in lateral VTC. This suggests that both voxels with the strongest positive preference to pseudowords (and characters) and those with the most negative preference contribute to word information. However, voxels with no preference (either positive or negative to pseudowords) do not contribute to classification and in fact adding many of them decreases information.

Is Word and Character Information in Lateral VTC Related to Reading Ability?

We reasoned that if VTC development is related to reading ability, participants with more informative representations (i.e., those whose VTC produced better classification) would also read better. Thus, we examined if there is a correlation between reading performance (assessed with the Woodcock Reading Mastery Test, WRMT) outside the scanner with word and character classification performance from distributed lateral VTC responses. The WRMT tests reading accuracy by scoring how many words or pseudowords of increasing difficulty subjects can read until they make 4 consecutive errors.

Consistent with our hypothesis, WRMT performance was significantly correlated with both character classification from left lateral VTC ($r = 0.60$, $P < 0.001$; after partialling out age, $r = 0.36$, $P < 0.03$, Fig. 5a; the former is also significant for a Bonferroni-corrected P -value: $0.05/3 = 0.016$) and word classification from left lateral VTC ($r = 0.58$, $P < 0.001$; after partialling out age $r = 0.22$, $P = 0.19$). In contrast, there was no significant correlation between WRMT performance and either word or character classification from medial VTC in either hemisphere (r s ≤ 0.27 , P s > 0.1). There was also no significant correlation between WRMT performance and word classification in the right lateral VTC ($r = 0.19$, $P = 0.45$), and the correlation between WRMT performance and character classification in right lateral VTC ($r = 0.33$, $P < 0.05$), was not significant after factoring out age ($r = -0.08$, $P = 0.64$). Together, these results suggest that development of character and word information in left lateral VTC correlates with reading ability.

As our analyses of information in VTC reveal that discriminative voxels drive the development of character and word information in left lateral VTC, we next tested if discriminative voxels better correlate with reading performance than the remaining lateral VTC voxels. Results reveal that reading ability was significantly correlated with classification of characters based on the top 30% discriminative voxels ($r = 0.67$, $P < 0.001$, also significant when partialling out age $r = 0.52$, $P < 0.001$, and for a Bonferroni-corrected P -value of $0.05/3 = 0.016$, Fig. 5b). In contrast, there was no significant correlation between WRMT and character classification of the non-discriminative lateral VTC voxels ($r = 0.27$, $P = 0.11$, Fig. 5c). Furthermore, the former correlation was significantly higher than the latter (Fisher transform, $P = 0.03$). Likewise, WRMT was correlated with word classification based on the 30% left lateral VTC voxels which were most discriminative for pseudowords ($r = 0.57$, $P < 0.001$; after partialling out age $r = 0.28$, $P = 0.1$), but there was no significant correlation between WRMT and the left lateral VTC excluding these discriminative voxels ($r = 0.23$, $P = 0.16$).

Finally, we determined which of discriminative voxels, those with positive or those with negative preference, were correlated with reading performance. Results indicate that WMRT performance was significantly correlated with classification from voxels with positive preference to characters ($r = 0.67$, $P < 0.001$; significant after partialling out age, $r = 0.44$, $P = 0.007$, Fig. 5d; both correlations are significant after Bonferroni-correction: Bonferroni-corrected P -value: $0.05/2 = 0.025$), but not with those with negative preference ($r = 0.14$, $P = 0.44$, Fig. 5e). Similarly, WRMT performance was correlated with classification based on voxels with positive preference to words ($r = 0.54$, $P < 0.001$, after partialling out age, $r = 0.31$, $P = 0.07$), but not with those with negative preference ($r = 0.25$, $P = 0.15$). In other words, reading ability is linked to distributed information from voxels with positive preference (that is, are selective) to characters and words

Figure 5. Character classification in discriminative voxels predicts reading performance. (a–c) Scatterplots of the correlation between reading ability measured by performance in the WRMT and character classification from distributed left VTC responses. Graphs differ in the brain classification data. (a) All left lateral VTC voxels. (b) 30% most discriminative voxels (see Figure 4c). (c) Left lateral VTC voxels except the 30% most discriminative voxels. Voxel data from (b) was further split into two sets, those with a positive preference to words (d) and those with a negative preference to words (e). In all plots, correlations are indicated by lines and numbers in the bottom left. Solid black lines; significant correlations; Dashed gray: non-significant correlations. Bonferroni-corrected thresholds are applied: top row: $0.05/3 = 0.016$; bottom row: $0.05/2 = 0.025$. Only participants who completed the WRMT are shown here (5–9-year-olds, $n = 7$; 10–12-year-olds, $n = 11$; 22–28-year-olds, $n = 19$).

within the top 30% discriminative voxels. Critically, this relationship dissolves when these selective voxels are excluded.

In sum, these analyses reveal that development of a subset of voxels rather than the entire left lateral VTC, best correlates with reading performance. This suggests that development of reading ability is guided by neural development that is both anatomically specific (left lateral VTC) and functionally specific (discriminative voxels), and is largely driven by the word- and character-selective voxels.

Discussion

Reading is a complex process requiring years of practice until it is mastered. The present study investigated how learning to

read during childhood is related to the development of distributed representations of characters and words across VTC. Our study is the first to show that distributed responses in left lateral VTC become more distinctive and informative after age 5, and crucially that this development is linked to reading ability. This conclusion is supported by 4 observations: First, character and word information in distributed VTC responses increase from age 5 to adulthood in an anatomical and hemispheric specific manner. The most prominent development occurs in left lateral VTC compared to right lateral VTC or medial VTC, bilaterally. Second, development of character information occurs even as information for the domains of bodies, objects, and places do not develop significantly, indicating a differential development of domain information in VTC (Golarai et al. 2007,

2010, 2017). Third, developmental increases in information regarding words/characters are due to development of distributed responses across the subset of discriminative word/character voxels within lateral VTC. Fourth, even as information in VTC develops in both voxels with positive and negative preference to characters/words, it is the development of selective voxels with positive preference that is linked to reading ability. These findings have important implications for both developmental theories and understanding the neural bases of reading disabilities, which we detail below.

Notably, the anatomical specificity of development of word/character information in distributed VTC responses is consistent with the predictions of the eccentricity bias (Malach et al. 2002) and graded hemispheric specialization theories (Behrmann and Plaut 2015). The former is supported by data showing that word information develops in lateral VTC, which shows a foveal bias in both children and adults (Weiner et al. 2014), but not in medial VTC, which shows a peripheral bias. The latter is supported by data revealing a more prominent development of word/character information in the left than right lateral VTC. According to this theory, when children learn to read, processing of letters and words depends more on the left hemisphere because of left lateralization of language areas in other parts of the brain which is present in infancy (Dehaene-Lambertz et al. 2002). In our data, the 5–9 year olds do now show lateralization of information for characters, words, or numbers in VTC (Fig. 2), and lateralization increases with age. This suggests the possibility that left-lateralized top-down influences of language areas (Dehaene-Lambertz et al. 2002; Behrmann and Plaut 2015), as well as white matter changes across development (Yeatman et al. 2012; Takeuchi et al. 2016), generate lateralized distributed representations within VTC in adulthood. Future research can also test if the increased lateralization in word information is accompanied by de-correlation or reduced cortical synchronization of lateral VTC across development.

Another novel aspect of our study is its data-driven, computational approach, which enabled us to quantify which voxels are informative and contribute to behavior. This approach revealed that (i) development of a subset of discriminative voxels within lateral VTC is what increases both word and character information from age 5 to adulthood, and (ii) discriminative voxels contain more information than the entire lateral VTC (Fig. 4). These findings show for the first time that development of literacy affects a larger set of neuronal populations within VTC, not only just the VWFA (Dehaene et al. 2010; Ben-Shachar et al. 2011; Cantlon et al. 2011). This finding has important implications for assessment of the neural bases of reading disabilities as it suggests that future investigations should consider examining functional differences and atypical development across VTC beyond the VWFA.

Crucially, our results show that reading performance is linked to the amount of character and word information in lateral VTC. First, reading ability was correlated with patterns of activity across discriminative voxels in lateral VTC, but not the rest of lateral VTC. Second, reading ability was linked to information distributed across lateral VTC voxels with positive preference (i.e., selective voxels) rather than negative preference to words and characters.

These findings have two important implications.

First, our data advance understanding of the neural basis of reading ability by providing a neural mechanism underlying its development. Notably, we found that neural development is driven both by increased within-domain similarity of distributed responses as well as increased between-domain distinctiveness (Fig. 3). Together these developments lead to increased

information about words in the left lateral VTC, which, in turn, are correlated with reading ability.

Second, our innovative approach provides a new computational and data-driven method to assess which voxels (features) within distributed patterns contribute to behavior. By combining brain and behavioral measurements we show that distributed information across the subset of word/character selective voxels is correlated with reading ability even as distributed responses across a broader set of discriminative lateral VTC voxels shows substantial neural development. This observation highlights that information in distributed neural responses does not guarantee its behavioral relevance. Importantly, our new approach can be applied broadly across the brain to evaluate the contribution of distributed responses to behavior, and resolves outstanding debates about the utility of MVPA for fMRI data (Norman et al. 2006; Dubois et al. 2015). Together, our data underscore that it is necessary to combine brain and behavioral measurements to understand the impact of distributed information on behavior, and provides empirical support for the domain-specific view that the left VWFA is critical for reading (Gaillard et al. 2006).

We acknowledge that our finding of a correlation between word and character information and reading ability does not provide evidence that this relationship is causal. Further, we recognize that the amount of word and character information in lateral VTC does not explain all the variance in reading ability data. This suggests the possibility that additional processes involved in reading, which were not investigated in the present study, such as the association of visual inputs with sounds and language, may contribute to the remaining variance in reading ability. Future research investigating additional cortical regions (such as those in prefrontal and parietal cortex) as well as longitudinal measurements in a larger sample of children (Dehaene-Lambertz et al. 2018), may provide higher sensitivity compared the present study and would be instrumental to test these hypotheses.

Our data also generate new questions for future research. *First, why does word information develop more than number information in lateral VTC even as both types of stimuli are learned during school years?* We propose several possible hypotheses for the stronger development for words in comparison to numbers observed in the present study. One possibility is that differences in holistic processing of words versus numbers underlie differences in the development of distributed responses to these stimuli. For example, adults (native English speakers) find it difficult not to read the entire pseudoword “Sib” in Figure 1a; in contrast, in order to infer the numerical quantity represented by the number ‘1453’ in Figure 1a, it is necessary to process every digit separately. The former, holistic processing of words, may be accomplished via spatial integration by large and foveal population receptive fields (pRFs), in word-selective voxels in left lateral VTC, which continue to develop after age 5 (Gomez et al. 2018). However, if processing of numbers does not involve the same spatial integration as words, number pRFs, and in turn, distributed representations to numbers may show lesser development. Another possibility is that number-selective regions may be smaller, more heterogeneous, and located more laterally than word-selective regions, thereby making it more difficult to measure fMRI signals from these regions (Daitch et al. 2016; Grotheer et al. 2018).

Second, *what features of word/character information develop?* Does learning to read affect the tuning to characters, orthographic information, lexical information, and/or the tuning to the statistics of the language (e.g., bigram frequency, Binder

et al. 2006; Vinckier et al. 2007; Glezer et al. 2009; Taylor et al. 2014)? This question can be addressed in the future using fMRI-adaptation (Grill-Spector et al. 1999; Natu et al. 2016; Nordt et al. 2016) and longitudinal measurements in children as they learn to read (Dehaene-Lambertz et al. 2018).

Third, how does the development of word/character information affect development of information to other categories? Our data show that as domain information for characters develops in VTC, there is no development of information for the domains of bodies, objects, and places. However, consistent with prior research (Golarai et al. 2017) we found developmental increases in domain information to faces.

One possibility is that development of face and word information occur in tandem, but are independent from each other. For example, research in developmental and acquired prosopagnosia suggests that impairments in face recognition can occur even as visual processing of words is normal (Susilo et al. 2015; Rubino et al. 2016; Burns et al. 2017). In contrast, the “many-to-many” view, which suggests shared neural circuits for word and face processing, predicts that developmental deficits for faces and words will co-occur (Behrmann and Plaut 2013; Collins et al. 2017). Additionally, developmental theories (Behrmann and Plaut 2015; Dehaene et al. 2015) have also suggested that the interactive development of face and word processing is due to competition on foveal resources, resulting in the left lateralization of word information and the right lateralization of face information (Behrmann and Plaut 2015; Dehaene et al. 2015; Gomez et al. 2018). In our data, we found no significant correlation between word classification in left lateral VTC and face classification in right lateral VTC ($-0.099 < r < 0.044$, $P_s > 0.49$). Additionally, there was no significant correlation between the laterality index for characters (or words) in the left lateral VTC and faces (or child faces) in the right lateral VTC ($-0.062 < r_s < 0.12$, $P_s > 0.42$). Thus, in our data development of distributed responses to faces and words are largely independent from each other. Future longitudinal research can examine if and how competition between processing of faces and words may shape VTC responses.

Finally, is there a critical period in which development of distributed patterns in VTC can occur, or is this flexibility maintained throughout the lifespan? Studies of people who gained literacy in adulthood show lesser changes in VWFA response amplitudes with literacy as compared with people who gained literacy as children (Dehaene et al. 2010, 2015). Additionally, research in non-human primates reveal that extensive symbol training leads to development of regions selective to trained symbols in juvenile, but not adult macaque monkeys (Srihasam et al. 2012). These data suggest that distributed representations to characters and symbols may be more malleable in childhood than adulthood. This hypothesis can be examined in future longitudinal investigations both in children and in illiterate adults as they gain reading proficiency. Additionally, longitudinal studies in people who gain literacy in adulthood can also reveal how the effects of literacy acquisition can be distinguished from general developmental effects occurring during childhood.

In sum, our data show that not only the amount of word information in VTC increases as children learn to read in an anatomical- and hemisphere-specific way, but also that this development is correlated with reading ability. These findings suggest that development of distributed responses to words and characters may be influenced both by foveal biases and interactions with left-lateralized language areas. Consequently, they have important implications for both developmental theories and for understanding reading disabilities.

Supplementary Material

Supplementary material is available at *Cerebral Cortex* online.

Authors' Contributions

M.N. developed the analysis pipeline, analyzed the data, and wrote the manuscript. V.N. and J.G. contributed to experimental design, collection and analysis of data, and preparation of the manuscript. B.J., M.B. contributed to collection of data, and preparation of the manuscript. K.G.-S. designed the experimental design, developed the analysis pipeline, contributed to the data analysis, and wrote the manuscript.

Notes

This work was supported by a scholarship of the German National Academic Foundation and by the Ruhr University Research School PLUS, funded by Germany's Excellence Initiative (DFG GSC 98/3) awarded to M.N.; NIH grants (Grant numbers 1R01EY02231801A1, 1R01EY02391501A1 to K.G.S.); a training grant (Grant number 5T32EY020485 to V.N.); the NSF Graduate Research Development Program (Grant number DGE-114747 to J.G.) and Ruth L. Kirschstein National Research Service Award (Grant number F31EY027201 to J.G.). *Conflict of interest:* None declared.

References

- Behrmann M, Plaut DC. 2013. Distributed circuits, not circumscribed centers, mediate visual recognition. *Trends Cogn Sci*. 17:210–219.
- Behrmann M, Plaut DC. 2015. A vision of graded hemispheric specialization. *Ann N Y Acad Sci*. 1359:30–46.
- Ben-Shachar M, Dougherty RF, Deutsch GK, Wandell BA. 2011. The development of cortical sensitivity to visual word forms. *J Cogn Neurosci*. 23:2387–2399.
- Ben-Shachar M, Dougherty RF, Wandell BA. 2007. White matter pathways in reading. *Curr Opin Neurobiol*. 17:258–270.
- Binder JR, Medler DA, Westbury CF, Liebenthal E, Buchanan L. 2006. Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage*. 33:739–748.
- Brem S, Bach S, Kucian K, Kujala JV, Guttorm TK, Martin E, Lyytinen H, Brandeis D, Richardson U. 2010. Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proc Natl Acad Sci*. 107:7939–7944.
- Burns EJ, Bennetts RJ, Bate S, Wright VC, Weidemann CT, Tree JJ. 2017. Intact word processing in developmental prosopagnosia. *Sci Rep*. 7:1–12.
- Cantlon JF, Pineda P, Dehaene S, Pelphrey KA. 2011. Cortical representations of symbols, objects, and faces are pruned back during early childhood. *Cereb Cortex*. 21:191–199.
- Carlson T, Tovar DA, Alink A, Kriegeskorte N. 2013. Representational dynamics of object vision: the first 1000 ms. *J Vis*. 13:1.
- Carreiras M, Seghier ML, Baquero S, Estévez A, Lozano A, Devlin JT, Price CJ. 2009. An anatomical signature for literacy. *Nature*. 461:983–986.
- Cohen L, Dehaene S, Naccache L, Lehéicy S, Dehaene-Lambertz G, Hénaff M-A, Michel F. 2000. The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*. 123:291–307.
- Collins E, Dundas E, Gabay Y, Plaut DC, Behrmann M. 2017. Hemispheric organization in disorders of development. *Vis cogn*. 25:416–429.

- Connolly AC, Guntupalli JS, Gors J, Hanke M, Halchenko YO, Wu YC, Abdi H, Haxby JV. 2012. The representation of biological classes in the human brain. *J Neurosci.* 32:2608–2618.
- Cox DD, Savoy RL. 2003. Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage.* 19:261–270.
- Daitch AL, Foster BL, Schrouff J, Rangarajan V. 2016. Mapping human temporal and parietal neuronal population activity and functional coupling during mathematical cognition. *Proc Natl Acad Sci U S A.* 113:E7277–E7286.
- Dehaene S, Cohen L, Morais J, Kolinsky R. 2015. Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nat Rev Neurosci.* 16:234–244.
- Dehaene S, Le Clec’h G, Poline J-B, Le Bihan D, Cohen L. 2002. The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport.* 13:321–325.
- Dehaene S, Pegado F, Braga LW, Ventura P, Nunes Filho G, Jobert A, Dehaene-Lambertz G, Kolinsky R, Morais J, Cohen L. 2010. How learning to read changes the cortical networks for vision and language. *Science.* 330:1359–1364.
- Dehaene-Lambertz G, Dehaene S, Hertz-Pannier L. 2002. Functional neuroimaging of speech perception in infants. *Science.* 298:2013–2015.
- 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 Biol.* 16:e2004103.
- Dubois J, de Berker AO, Tsao DY. 2015. Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *J Neurosci.* 35:2791–2802.
- Dumoulin SO, Wandell BA. 2008. Population receptive field estimates in human visual cortex. *Neuroimage.* 39:647–660.
- Gaillard R, Naccache L, Pinel P, Clémenceau S, Volle E, Hasboun D, Dupont S, Baulac M, Dehaene S, Adam C, et al. 2006. Direct intracranial, fMRI, and lesion evidence for the causal role of left inferotemporal cortex in reading. *Neuron.* 50:191–204.
- Glezer LS, Jiang X, Riesenhuber M. 2009. Evidence for highly selective neuronal tuning to whole words in the “Visual Word Form Area”. *Neuron.* 62:199–204.
- Glezer LS, Kim J, Rule J, Jiang X, Riesenhuber M. 2015. Adding words to the brain’s visual dictionary: novel word learning selectively sharpens orthographic representations in the VWFA. *J Neurosci.* 35:4965–4972.
- Golarai G, Ghahremani DG, Whitfield-Gabrieli S, Reiss A, Eberhardt JL, Gabrieli JD, Grill-Spector K. 2007. Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nat Neurosci.* 10:512–522.
- Golarai G, Liberman A, Grill-Spector K. 2017. Experience shapes the development of neural substrates of face processing in human ventral temporal cortex. *Cereb Cortex.* 27:1229–1244.
- Golarai G, Liberman A, Yoon JM, Grill-Spector K. 2010. Differential development of the ventral visual cortex extends through adolescence. *Front Hum Neurosci.* 3:80.
- Gomez J, Barnett MA, Natu V, Mezer A, Palomero-Gallagher N, Weiner KS, Amunts K, Zilles K, Grill-Spector K. 2017. Microstructural proliferation in human cortex is coupled with the development of face processing. *Science.* 355:68–71.
- Gomez J, Natu V, Jeska B, Barnett M, Grill-Spector K. 2018. Development differentially sculpts receptive fields across early and high-level human visual cortex. *Nat Commun.* 9:788.
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, Malach R. 1999. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron.* 24:187–203.
- Grill-Spector K, Weiner KS. 2014. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat Rev Neurosci.* 15:536–548.
- 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.
- Gullick MM, Booth JR. 2015. The direct segment of the arcuate fasciculus is predictive of longitudinal reading change. *Dev Cogn Neurosci.* 13:68–74.
- Hannagan T, Amedi A, Cohen L, Dehaene-Lambertz G, Dehaene S. 2015. Origins of the specialization for letters and numbers in ventral occipitotemporal cortex. *Trends Cogn Sci.* 19:374–382.
- Hasson U, Levy I, Behrmann M, Hendler T, Malach R. 2002. Eccentricity bias as an organizing principle for human high-order object areas. *Neuron.* 34:479–490.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science.* 293:2425–2430.
- Jacques C, Witthoft N, Weiner KS, Foster BL, Rangarajan V, Hermes D, Miller KJ, Parvizi J, Grill-Spector K. 2016. Corresponding ECoG and fMRI category-selective signals in human ventral temporal cortex. *Neuropsychologia.* 83:14–28.
- Kay KN, Winawer J, Mezer A, Wandell BA. 2013. Compressive spatial summation in human visual cortex. *J Neurophysiol.* 110:481–494.
- Kovelman I, Norton ES, Christodoulou JA, Gaab N, Lieberman DA, Triantafyllou C, Wolf M, Whitfield-gabrieli S, Gabrieli JDE. 2012. Brain basis of phonological awareness for spoken language in children and its disruption in dyslexia. *Cereb Cortex.* 22:754–64.
- Kriegeskorte N. 2008. Representational similarity analysis – connecting the branches of systems neuroscience. *Front Syst Neurosci.* 2:4.
- Levy I, Hasson U, Avidan G, Hendler T, Malach R. 2001. Center-periphery organization of human object areas. *Nat Neurosci.* 4:533–539.
- Malach R, Levy I, Hasson U. 2002. The topography of high-order human object areas. *Trends Cogn Sci.* 6:176–184.
- Mezer A, Yeatman JD, Stikov N, Kay KN, Cho NJ, Dougherty RF, Perry ML, Parvizi J, Hua le H, Butts-Pauly K, et al. 2013. Quantifying the local tissue volume and composition in individual brains with magnetic resonance imaging. *Nat Med.* 19:1667–1672.
- Natu VS, Barnett MA, Hartley J, Gomez J, Stigliani A, Grill-Spector K. 2016. Development of neural sensitivity to face identity correlates with perceptual discriminability. *J Neurosci.* 36:10893–10907.
- Nordt M, Hoehl S, Weigelt S. 2016. The use of repetition suppression paradigms in developmental cognitive neuroscience. *Cortex.* 80:61–75.
- Norman KA, Polyn SM, Detre GJ, Haxby JV. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci.* 10:424–430.
- Rauschecker AM, Bowen RF, Perry LM, Kevan AM, Dougherty RF, Wandell BA. 2011. Visual feature-tolerance in the reading network. *Neuron.* 71:941–953.
- Rubino C, Corrow SL, Corrow JC, Duchaine B, Barton JJS. 2016. Word and text processing in developmental prosopagnosia. *Cogn Neuropsychol.* 33:315–328.

- Saygin ZM, Osher DE, Norton ES, Youssoufian DA, Beach SD, Feather J, Gaab N, Gabrieli JD, Kanwisher N. 2016. Connectivity precedes function in the development of the visual word form area. *Nat Neurosci.* 19:1250–1255.
- Schlaggar BL, McCandliss BD. 2007. Development of neural systems for reading. *Annu Rev Neurosci.* 30:475–503.
- Srihasam K, Mandeville JB, Morocz IA, Sullivan KJ, Livingstone MS. 2012. Behavioral and anatomical consequences of early versus late symbol training in Macaques. *Neuron.* 73:608–619.
- Stigliani A, Weiner KS, Grill-Spector K. 2015. Temporal processing capacity in high-level visual cortex is domain specific. *J Neurosci.* 35:12412–12424.
- Susilo T, Wright V, Tree JJ, Duchaine B. 2015. Acquired prosopagnosia without word recognition deficits. *Cogn Neuropsychol.* 32:321–339.
- Takeuchi H, Taki Y, Hashizume H, Asano K, Asano M, Sassa Y, Yokota S, Kotozaki Y, Nouchi R, Kawashima R. 2016. Impact of reading habit on white matter structure: cross-sectional and longitudinal analyses. *Neuroimage.* 133:378–389.
- Taylor JSH, Rastle K, Davis MH. 2014. Distinct neural specializations for learning to read words and name objects. *J Cogn Neurosci.* 26:2128–2154.
- Vinckier F, Dehaene S, Jobert A, Dubus JP, Sigman M, Cohen L. 2007. Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. *Neuron.* 55:143–156.
- Wandell BA, Rauschecker AM, Yeatman JD. 2012. Learning to see words. *Annu Rev Psychol.* 63:31–53.
- Weiner KS, Golarai G, Caspers J, Chuapoco MR, Mohlberg H, Zilles K, Amunts K, Grill-Spector K. 2014. The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *Neuroimage.* 84:453–465.
- Weiner KS, Grill-Spector K. 2010. Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage.* 52:1559–1573.
- Yeatman JD, Dougherty RF, Ben-Shachar M, Wandell BA. 2012. Development of white matter and reading skills. *Proc Natl Acad Sci U S A.* 109:E3045–E3053.