Learning to read increases the informativeness of distributed ventral temporal responses

Marisa Nordt[1], Jesse Gomez[2], Vaidehi Natu[3], Brianna Jeska[3], Michael Barnett[3], and Kalanit Grill-Spector[2,3,4*]

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

[2]Neurosciences Program, Stanford University School of Medicine, Stanford, CA 94305, USA.

[3]Psychology Department, Stanford University, Stanford, CA 94305, USA.

[4]Stanford Neurosciences Institute, Stanford University, Stanford, CA 94305, USA.

*Corresponding author:

Kalanit Grill-Spector

Jordan Hall, Bldg. 420

Department of Psychology

Stanford Neurosciences Institute

Stanford, California 94305-2130

Tel: 650-269-9605

Email: kalanit@stanford.edu

Summary

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) vs. 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 development of voxels with both positive (that is, word-selective) and negative preference to words affected distributed information, only development of word-selective voxels predicted reading ability. These data show that developmental increases in informativeness of distributed left lateral VTC responses enable proficient reading and have important implications for both developmental theories and for elucidating neural mechanisms of reading disabilities.

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 and associating these visual inputs with sounds and language. Thus, reading involves a network of brain regions involved in vision, audition, and language. As such, prior research has examined how learning to read affects white matter tracts of the reading network (Ben-Shachar et al., 2007; Carreiras et al., 2009; Schlaggar and McCandliss, 2007; Wandell et al., 2012; Yeatman et al., 2012) and the emergence of the visual word form area (VWFA, Ben-Shachar et al., 2011; Cantlon et al., 2011). The VWFA (Cohen et al., 2000; Dehaene et al., 2002) is located within ventral temporal cortex (VTC), an anatomical expanse that processes high-level visual information (Cohen et al., 2000; Dehaene et al., 2002; Grill-Spector and Weiner, 2014; Hannagan et al., 2015; Rauschecker et al., 2011). A separate body of research has shown that distributed responses across VTC have a characteristic pattern that represents the category of the visual input (Carlson et al., 2013; Golarai et al., 2017; Haxby et al., 2001; Kriegeskorte, 2008). Indeed, independent classifiers can decode from distributed VTC responses the category of the viewed stimulus (Cox and Savoy, 2003; Grill-Spector and Weiner, 2014; Weiner and Grill-Spector, 2010). However, it is unknown how learning to read during childhood affects the development of distributed representations of words and characters across VTC and if cortical development has behavioral ramifications.

To address these gaps in knowledge, we examined three 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. 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 because both groups 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., 2010, 2017).

(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 (Carlson et al., 2013; Connolly et al., 2012; Cox and Savoy, 2003; Haxby et al., 2001; Kriegeskorte, 2008). This theory predicts no anatomical specificity within VTC to the development of distributed responses. In contrast, eccentricity bias theory (Hasson et al., 2002; Levy et al., 2001; 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 (Hasson et al., 2002; Levy et al., 2001; Weiner et al., 2014) and regions medial to the MFS are peripherally-biased (Hasson et al., 2002; Levy et al., 2001; 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 improves 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 (Gullick and Booth, 2015; Takeuchi et al., 2016; Yeatman et al., 2012) 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 three 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 four other domains, each consisting of two categories (**Figure 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 delineated in each subject their lateral and medial VTC (**Figure 1b**), measured distributed responses to items from each category, and examined: (i) if there are age-related differences in reliability, distinctiveness, or informativeness of distributed VTC responses to words and characters, (ii) the anatomical specificity of functional development, and (iii) the relationship between distributed VTC patterns and reading ability across development.

Results

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

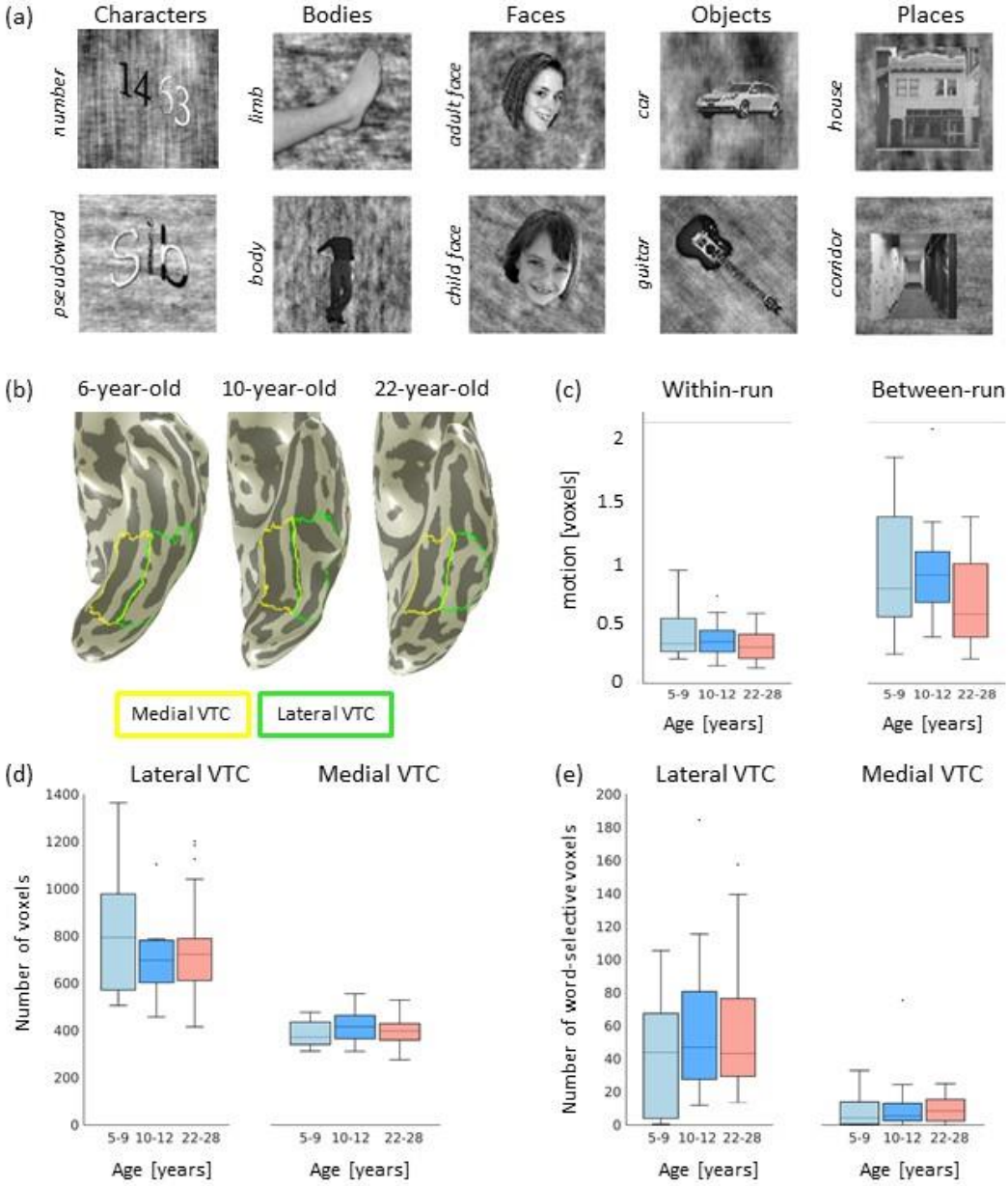

Figure 1. Stimuli, anatomical ROI definitions, and between age-group controls (a) Examples of stimuli corresponding to the five domains (in columns) with two categories per domain (rows). (b) Examples of lateral and medial ventral temporal cortex (VTC) on the inflated cortical surface of a representative 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 three 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.

Do distributed VTC responses to characters, words, and numbers develop after age five?

In each age group, we measured the similarity (Pearson correlation coefficient) between multivoxel patterns (MVPs) of responses to different items within the character

domain in different runs (pseudowords-pseudowords (w-w); numbers-numbers (n-n); pseudowords-numbers (w-n)). We reasoned that reliable representations should generate positive correlations between MVPs of characters and developmental increases in correlations would indicate increased reliability of MVPs across items within the character domain.

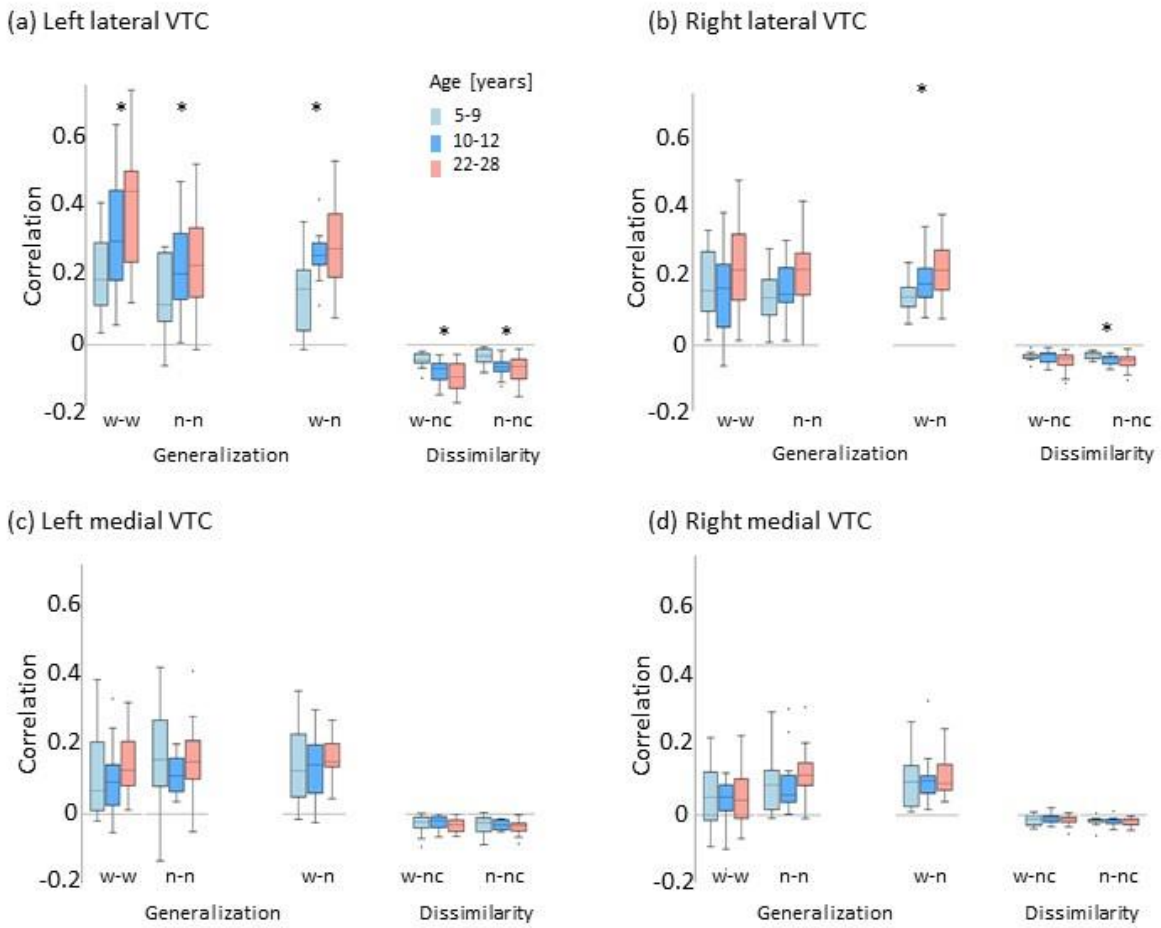

Figure 2. Within-domain generalization and between-domain dissimilarity of distributed responses to characters increase from age 5 to adulthood. Boxplots depict the Pearson correlation coefficient between multivoxel patterns (MVPs) of distributed responses within the domain of characters (w–w: pseudowords–pseudowords; 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. The correlations between MVPs are shown for (a) Left lateral VTC; all comparisons showed a significant main effect of age as indicated by the asterisk (Fs(2,48) > 3.34; ps< 0.04). (b) Right lateral VTC; w–n and n–nn showed a significant effect of age (Fs(2,48) > 4.05; ps< 0.02). (c) Left medial VTC. (d) Right medial VTC. Boxplots are colored by age; *light blue*: 5-9-year-old, 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 for a comparison, p<0.05. *See also Figure S1*.

Results reveal three main findings. First, across age groups and VTC partitions correlations between MVPs to pseudowords (w-w) and numbers (n-n) were positive (**Figure 2,** *p*s<0.001) with left lateral VTC exhibiting the most reliable distributed responses to pseudowords **(Figure 2a).** Indeed, correlations between pseudoword and number MVPs varied across hemispheres and VTC partitions (significant 3-way interaction between hemisphere, partition, and character type, $F(1,48)=10.86$, $p=0.002$, 4-way repeated measures

analysis of variance (rmANOVA) with factors of age group (5-9/10-12/22-28), hemisphere (left/right), VTC partition (lateral/medial), and character type (w-w/n-n)). Second, correlations between pseudoword MVPs and number MVPs 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 (**Figure 2,w-w/n-n**): (i) development significantly varied across hemispheres and VTC partitions (significant age group × hemisphere × VTC partition interaction, $F(2,48)=4.21$, $p=0.02$), and (ii) development significantly varied across hemispheres and character types (significant age group × character-type × hemisphere interaction, $F(2,48)=3.52$, $p<0.04$).

The largest development of MVP reliability occurred for pseudowords in left lateral VTC, as compared to pseudowords or numbers in medial VTC or the right hemisphere. Indeed, similarity between MVPs to pseudowords in left lateral VTC increased from a median value of .19±.18 (median±interquartile range) in 5-9 year-olds to .44±.26 in adults (**Figure 2a, w-w**). In contrast, similarity among distributed responses to pseudowords in left medial VTC, only increased from .07±.20 to .13±.13 from age 5 to adulthood (**Figure 2c, w-w**), and the similarity of distributed responses to numbers in left lateral VTC, increased from a median of .12±.20 to .23±.20 from age 5 to adulthood (**Figure 2a, n-n**). At the same time, in lateral VTC, pseudoword MVPs became more similar to number MVPs from age 5 to adulthood (**Figure 2, w-n,** significant age group × VTC partition interaction, $F(2,48)=9.33$, $p<0.001$, 4-way rmANOVA). This indicates that across development, MVPs to pseudowords and numbers in VTC become more reliable and better generalize across items, especially so for pseudowords in left lateral VTC.

Next, we assessed whether there were developmental changes in distinctiveness, or dissimilarity between distributed responses to characters versus other domains. Thus, we measured the correlation between MVPs to pseudowords (or numbers) and MVPs to items of each of the other domains (**Figure S1** shows the entire representational similarity matrix). We reasoned that if characters elicited distinct responses from other stimuli, between-domain correlations among MVPs would be lower than within-domain correlations and development would result in decreases in between-domain correlations.

Results show that across all age groups, characters generated distributed response patterns across the VTC that were dissimilar from the other domains, as illustrated by the negative correlation between distributed responses to pseudowords and numbers vs. non-character stimuli (**Figure 2, w-nc/n-nc**). Between-domain correlations among MVPs 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 characters became significantly less correlated (or more dissimilar) from MVPs to other domains from age 5 to adulthood (**Figure 2, 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 the similarity between MVPs to characters vs. MVPs to other domains varied across VTC partitions (significant age group × VTC partition interaction, $F(2,48)=9.79$, $p<0.001$) as well as hemispheres (significant age group × VTC partition × hemisphere interaction, $F(2,48)=3.96$, $p=0.026$). The between-domain correlations significantly decreased in left lateral VTC, for both pseudowords and numbers (**Figure 2a, w-nc/n-nc**, both $Fs\geq6.07$, $ps\leq0.005$) and in right lateral VTC for numbers

(**Figure 2b, n-nc,** $F(2,48)=4.05$, $p=0.01$). These analyses indicate that across development, distributed responses to pseudowords and numbers in lateral VTC became more dissimilar from distributed responses to other domains and especially so in the left hemisphere.

Does information content in distributed VTC responses for characters, words, and numbers develop after age 5?

Motivated by these findings we used a decoding approach to quantify developmental changes in both character domain information and character type information in VTC. Thus, we quantified three types of information in distributed VTC responses: (i) characters (pseudowords+numbers) vs. other domains (faces, places, objects, and body parts), (ii) pseudowords vs. the other nine categories (adult faces, child faces, houses, corridors, cars, guitars, bodies, limbs, **Figure 1a**), and (iii) numbers vs. the other nine categories. To quantify information, we used a winner-take-all (WTA) classifier to evaluate how well it could decode information from MVPs.

Results show that in all age groups, VTC partitions, and hemispheres, decoding character information was significantly higher than the 20% chance level (**Figure 3a**, all $ps \leq 0.005$). Notably, decoding character information significantly increased from age 5 to adulthood (**Figure 3a,** main effect of age group, $F(2,48)=5.74$, $p=0.006$, 3-way rmANOVA with factors of age group, VTC partition, hemisphere). Consistent with the prior analyses, character information was better decoded from 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 $ps>0.066$).

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 vs. other stimuli from left lateral VTC yielded an accuracy of 65%±6% (mean±SEM) in children, but reached a 90%±3% accuracy in adults (significantly higher than 5-9 year-olds, post-hoc *t*-test, $t(36)=4.2$, $p<0.001$, **Figure 3a)**. In contrast to the development of character information, we did not find a significant development of information for bodies, objects, and places in VTC (**Figure S2a**). 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 (**Figure S2b,** 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.

Next, we examined if character type (pseudowords, numbers) information develops in VTC. Different than development of character domain information, we found differential development of word and number information across VTC partitions (**Figure 3b**, age group × character type × 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 × character type × hemisphere interaction, $F(2,48)=5.10$, $p=0.01$).

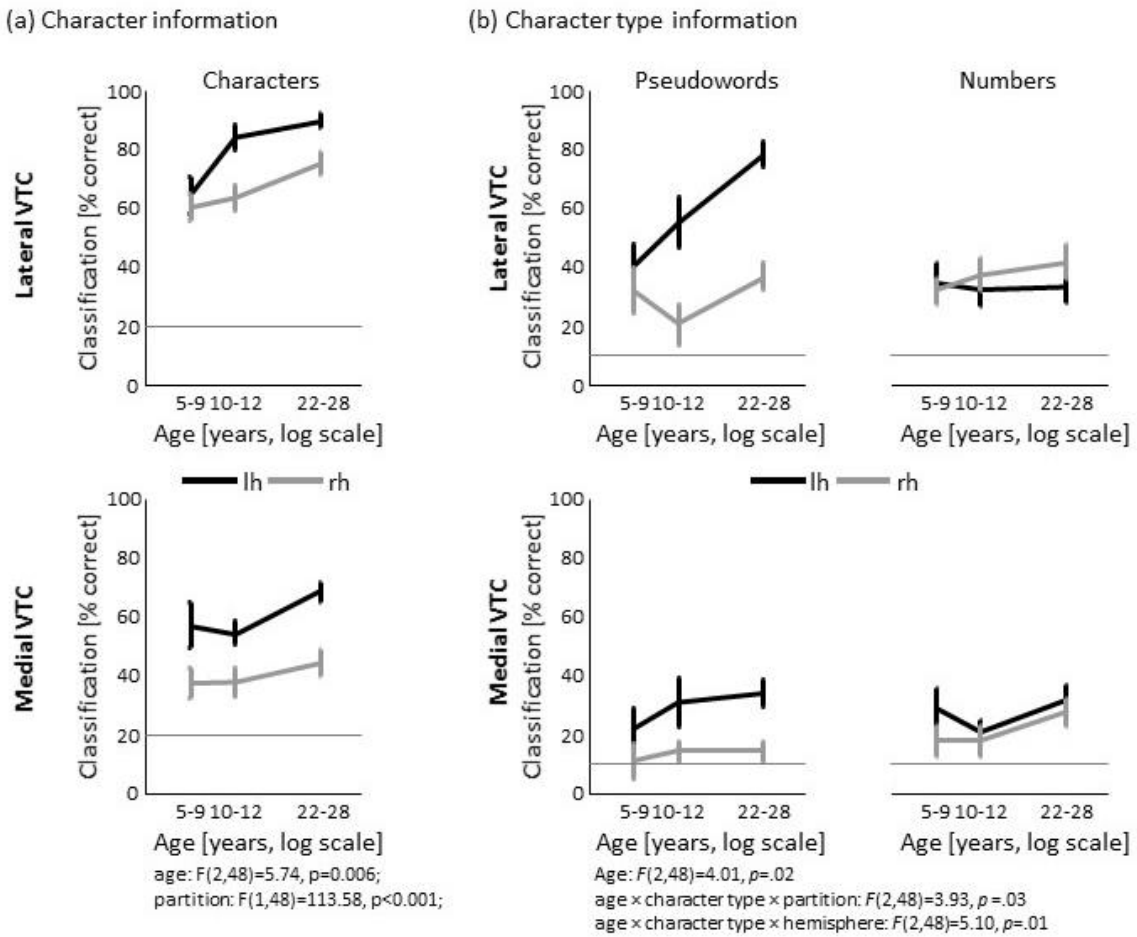

Figure 3. Differential development of word and number information after age 5. (a) Character classification performance in lateral VTC (top) and medial VTC (bottom) for the left (black) and right (gray) hemispheres across age groups. Classification performance was quantified with a winner-take-all (WTA) 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 S2 and S3.*

Notably, even as decoding character type from left lateral VTC was significantly higher than chance in all age groups and for both character types (all *p*s<0.003), decoding of pseudowords from left lateral VTC progressively increased from 40%±7% in 5-9 year-olds (significantly lower than adults, post-hoc *t*-test, $t(36)=-2.95$, $p=0.006$), to 55%±8% in 10-12 year-olds (significantly lower than adults, post-hoc *t*-test, $t(37)=-2.72$, $p=0.01$), to 78%±5% in 22-28 year-olds (**Figure 3b - top left**). This reflects an almost two-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%±4% (not significantly different across age groups, all *ps*>0.15, **Figure 3b-top left**). Likewise, there were no significant differences across age groups in decoding pseudowords from medial VTC (all *ps*>0.19, **Figure 3b-bottom left**) and performance was around 30% in the left hemisphere and around 13% in the right hemisphere.

In contrast to the development of decoding word information, there was no significant development of decoding number information in any partition or hemisphere (no significant

effects of age group or interactions with age group, all *Fs*<1.03, *ps*>0.36, 3-way rmANOVA on number classification with the factors age group, partition, hemisphere). Indeed, number decoding averaged about 30%±4% across age groups, partitions, and hemispheres (**Figure 3b-right**). Additionally, there was no significant development of information for other categories in VTC, except for corridors (**Figure S3**). 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 seems to be largely restricted to left lateral VTC.

Which voxels drive development of distributed responses for words and characters in lateral VTC?

One possibility is that development of word- and character-selective voxels (akin to the VWFA) drives the development of word and character information in lateral VTC, respectively. Alternatively, it may be that development of the entire lateral VTC including development of non-selective voxels increases word and character information, respectively.

To test these hypotheses, we examined classification performance for selective and non-selective voxels for pseudowords and characters, respectively. Selective voxels were defined in each subject and hemisphere as in prior studies (Golarai et al., 2010, 2017; Weiner and Grill-Spector, 2010) as voxels that responded more strongly to the preferred stimulus compared to other stimuli. In our analyses, we 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). The 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 reveal two main findings (left hemisphere: **Figure 4a;** right hemisphere: **Figure S4a**). First, as in previous analyses, word classification improved with age (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). Second, the development occurred in both selective and non-selective voxels (no age group × voxel type interaction, $F(2,42)=0.26$, $p=0.77$). Similar results were obtained for character classification (**Figure 4d, Figure S4d**). These analyses suggest that both selective and non-selective voxels in lateral VTC contribute to the development of word and character information. However, this interpretation should be considered with the following cautionary notes: (1) it is based on a particular choice of threshold to define selectivity, (2) the selective and non-selective voxels subsets are of vastly different set sizes (**Figure 4a,d**-legend), and (3) classification from these subsets is substantially lower than from all lateral VTC. Therefore, we sought to conduct a more principled analysis that compares information in a systematic manner across threshold levels and percentage of lateral VTC voxels.

Thus, we conducted two additional analyses. The first 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. The second analysis 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 could 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.

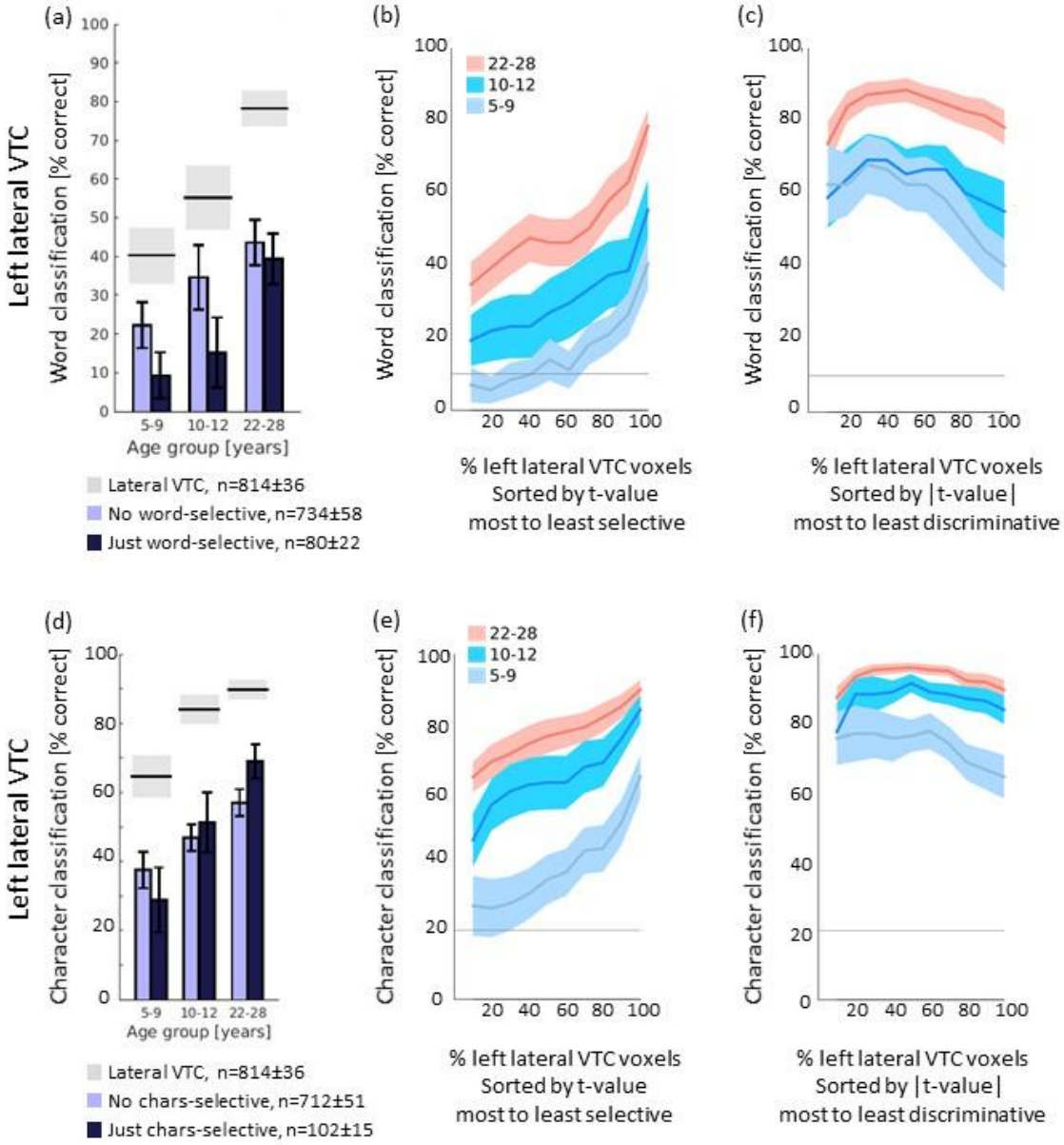

Figure 4. Development of word and character information in selective, non-selective, and discriminative voxels. (a) *Dark blue bars:* Mean WTA classification performance of left lateral VTC word-selective voxels (t>3). *Light blue bars:* Mean WTA classification performance of left lateral VTC voxels excluding the word-selective voxels (t ≤ 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,e,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.*

In all age groups, word classification performance gradually increased as more voxels with progressively lower word-selectivity were used for classification. Performance in both hemispheres was maximal when all lateral VTC voxels were included (**Figure 4b, Figure S4b**). Consistent with the prior analysis (**Figure 3**), maximal decoding of pseudowords was significantly higher in adults compared to children (**Figure 4b**, **Figure 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 (**Figure 4b**), but not in the right hemisphere (**Figure S4b,** age group × hemisphere interaction, $F(2,48)=6.55$, $p=0.003$).

In contrast, across all age groups, word classification from discriminative rather than selective voxels, yielded maximal performance using just ~35% of lateral VTC voxels. Performance plateaued for a range of 30-60% of voxels (**Figure 4c, Figure S4c**). Surprisingly, in all age groups and in both hemispheres, including additional voxels subsequently reduced performance. In fact, pseudoword classification from the entire lateral VTC was substantially lower than the maximal classification of the discriminative subset of voxels (all ps≤0.005, **Figure 4c, Figure S4c**). For example, in 5-9 year olds, maximal classification from the discriminative voxels was 77%±6% (using 36%±31% of voxels), but decoding dropped to 40%±7% using the entire left lateral VTC. The difference was even more pronounced in the right hemisphere: in 5-9 year-olds maximal classification from discriminative voxels was 64%±8% (using 36%±21% of voxels), but dropped to 32%± 8% for the entire right lateral VTC. The threshold of the 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 (**Figure 4c, Figure 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). Maximum classification 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 × hemisphere interaction, $F(2,37)=5.05$, $p=0.01$, 2-way rmANOVA). In adults, highest classification was achieved using 51%±29 of discriminative left lateral VTC voxels vs. 24%±17% of right lateral VTC voxels. Notably, no development of word information was observed in the remainder of the voxels after the lateral VTC 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). Similar patterns of results were observed for character classification (left hemisphere: **Figure 4f,** right hemisphere: **Figure S4f**).

In contrast to the development of both word and character information, complementary analyses of number decoding showed no development. There were no significant differences across age groups in decoding number information from lateral VTC using either number-selective or number-discriminative voxels (all Fs<1.5, ps>0.23, **Figure 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 the 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.

Does word and character information in lateral VTC predict reading ability?

We reasoned that if VTC development affects reading ability, participants with more informative representations (i.e. those whose VTC produced better classification) would also read better. Thus, we examined the correlation between reading performance (assessed with the Woodcock Reading Mastery Test, WRMT) outside the scanner with word and character classification performance from lateral VTC. 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=.60, p<0.001; also significant after partialling out age, r=.36, p<0.03, **Figure 5a**) and word classification from left lateral VTC (r=.58, p<.001; after partialling out age r=.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 (rs≤.27, ps>0.1). There was also no significant correlation between WRMT performance and word classification in the right lateral VTC (r=.19, p=0.45), and the correlation between WRMT performance and character classification in right lateral VTC (r=.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 to a lesser extent word information in left lateral VTC correlates with reading ability.

As our prior 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 predict 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=.67, p<0.001, also significant when partialling out age r=.52, p< 0.001, **Figure 5b**). In contrast, there was no significant correlation between WRMT and character classification of the non-discriminative lateral VTC voxels (r=.27, p=0.11, **Figure 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= .57, p<0.001; after partialling out age r=.28, p=0.1), but there was no significant correlation between WRMT and the left lateral VTC excluding these discriminative voxels (r=.27, p=0.11).

Finally, we determined if it is the voxels with positive or negative preference which predict reading performance. Results indicate that WMRT performance was significantly correlated with classification from voxels with positive preference to characters (r=.67, p<0.001; significant after partialling out age, r=.44, p=0.007, **Figure 5d**) but not with those

with negative preference (r=.14, p=0.44, **Figure 5e**). Similarly, WRMT performance was correlated with classification based on voxels with positive preference to words (r=.54, p<0.001, after partialling out age, r=.31, p=0.07), but not with those with negative preference

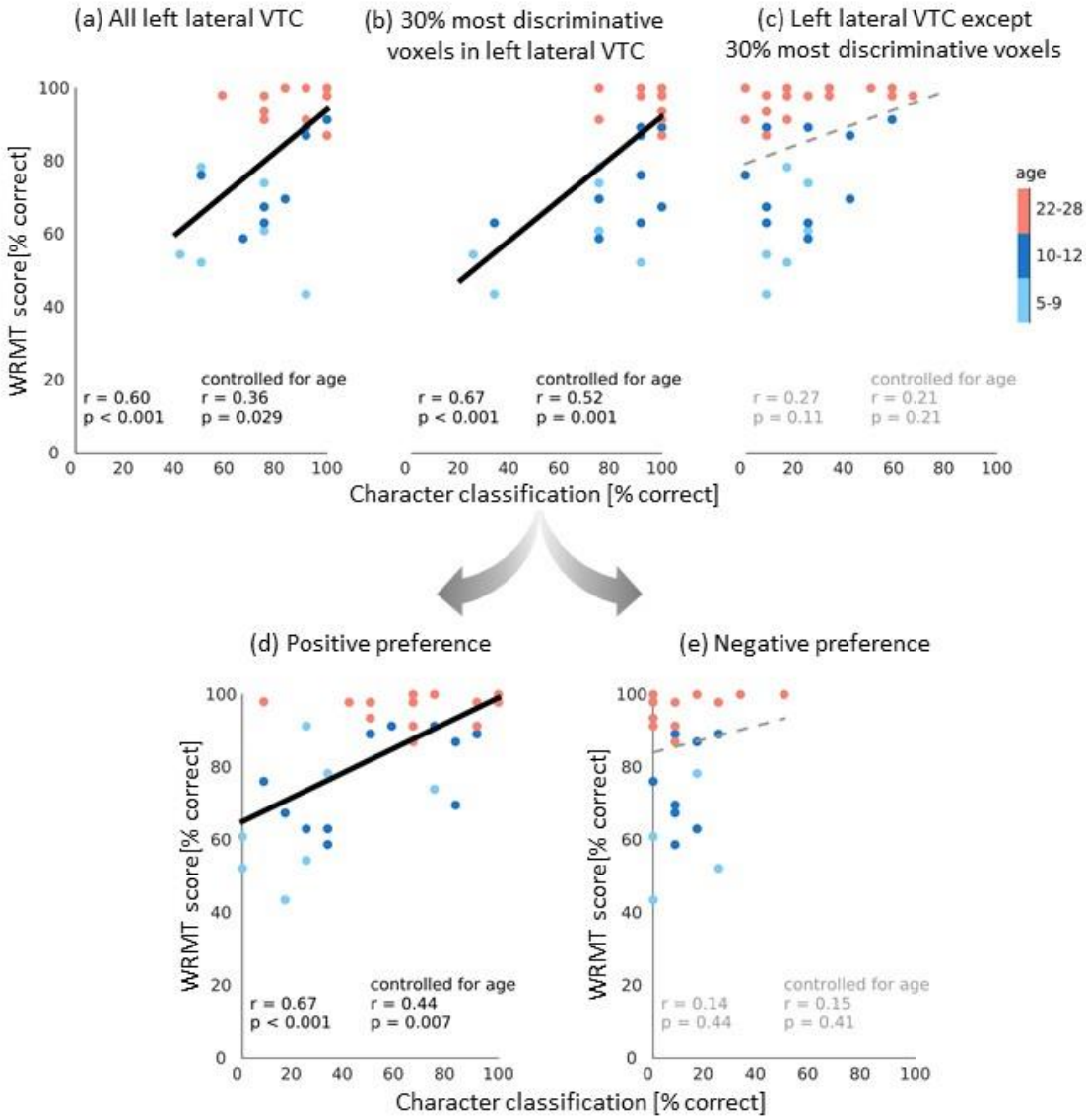

Figure 5. Character classification in discriminative voxels predicts reading performance. (a-c). Scatterplots of the correlation between reading ability measured by performance in the WRMT and character classification from distributed left VTC responses. Graphs differ in the brain classification data. (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. 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).

(r=.25, p=0.15). In other words, reading ability can be predicted from voxels that show positive preference (or selectivity) to characters and words within the top 30% discriminative voxels and critically, this prediction breaks down when these voxels are excluded.

In sum, these analyses reveal that development of a subset of voxels, which generate distinctive responses to characters and pseudowords in left lateral VTC (rather than the entire left lateral VTC), correlates with reading performance. This suggests that development of reading ability shows anatomical specificity (left lateral VTC) and functional specificity (discriminative voxels) but it is largely driven by 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 affects the development of distributed representations of characters and words across VTC, reporting four main findings. First, character and word information in distributed VTC responses increases 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, this development occurs even as information for the domains of bodies, objects, and places do not show significant development, 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, development of distributed responses across selective voxels with positive preference best predicts reading ability. Thus, our study is the first to show that distributed responses in left lateral VTC become more distinctive after age 5, and crucially that this development 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 both the predictions of the eccentricity bias and graded hemispheric specialization theories. The finding that word information develops in lateral VTC, which shows a foveal bias in both children and adults (Weiner et al., 2014), but not medial VTC is consistent with the predictions of eccentricity bias theory. The finding of a more prominent development of word/character information the left than right lateral VTC also supports the predictions of the graded hemispheric specialization theory (Behrmann and Plaut, 2015). According to this theory, when children learn to read, processing of letters and words depends more on the left hemisphere compared to the right because of left lateralization of language areas in other parts of the brain, which be left-lateralized already 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 (**Figure 3**), and lateralization increases with age. This suggests the possibility that left-lateralized top-down influences of language areas (Behrmann and Plaut, 2015; Dehaene-Lambertz et al., 2002), as well as white matter changes across development (Takeuchi et al., 2016; Yeatman et al., 2012), generate lateralized distributed representations within VTC in adulthood. The order of

developments can be assessed in future longitudinal studies in children, ideally starting with pre-readers.

A novel aspect of our study is its data-driven, computational approach, which enabled us to quantify which are the informative voxels. 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 (**Figure 4**). These findings show for the first time that not only word-selective voxels (that constitute the visual word form area, VWFA) change across development (Ben-Shachar et al., 2011; Cantlon et al., 2011) and with reading experience (Dehaene et al., 2010), but development affects a larger set of neuronal populations within VTC, which increase the distinctiveness of distributed responses. 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 the amount of character and word information in lateral VTC predicts reading performance. Specifically, we found that that the development of information across a subset of lateral VTC voxels is what predicts reading ability. First, within lateral VTC, discriminative voxels, but not the rest of lateral VTC, predict reading ability. Second, within the set of discriminative voxels, information distributed across voxels with positive preference (i.e., selective voxels) rather than negative preference predicts reading ability.

These findings have two important implications. First, our data suggest that development of word and character information across word- and character-selective voxels in left lateral VTC is related to reading proficiency. These findings complement prior reports of increased responses to letters and words in the VWFA from childhood to adulthood (Ben-Shachar et al., 2011; Brem et al., 2010; Cantlon et al., 2011; Olulade et al., 2013), as well as increased VWFA responses to letters and words with literacy in adults (Dehaene et al., 2010). Our data provide for the first time a neural mechanism explaining why reading ability improves, filling an important gap in knowledge. Specifically, we show that with development (and reading experience) representations to words become both more reliable, as well as more distinguishable from other stimuli (**Figure 2**). This leads to increased information about words in the left lateral VTC, which, in turn, improves reading ability.

Second, by combing brain and behavioral measurements we show that distributed information across the subset of word/character selective voxels predicts reading ability even as distributed responses across a broader set of discriminative lateral VTC voxels shows substantial development. This observation highlights that information in distributed neural responses does guarantee that it is behaviorally relevant. Thus, our innovative approach provides a new computational and data-driven method to assess which voxels (features) within distributed patterns contribute to behavior. This method 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 (Dubois et al., 2015; Norman et al., 2006). Together, our data underscores 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).

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?* One possibility is that in adults pseudowords are processed holistically, but unfamiliar numbers are not, and differences in holistic processing underlie differences in development of distributed responses. For example, as an adult, it is difficult for you not to read the entire pseudoword *'Sib'* in **Figure 1a**; in contrast, you likely have processed separately each digit of *'1453'* in **Figure 1a** to infer the quantity it represents. Holistic processing of words may be accomplished via spatial integration by large and foveal population receptive fields (pRFs, Kay, Weiner and Grill-Spector, 2015; Le et al., 2017) in word-selective voxels in left lateral VTC, which have been shown to develop from age 5 to adulthood (Gomez et al., in press). This suggests the possibility that as children learn to read, pRFs in lateral VTC develop, in turn affecting both spatial processing and distributed representations. However, because processing of number symbols does not require the same type of spatial integration as words, number pRFs, and in turn, distributed representations to numbers show lesser development. Alternatively, familiarity may play a role in the development of distributed responses (Connolly et al., 2012; Golarai et al., 2017). This predicts that distributed responses to familiar numbers and words may show a similar development. These possibilities can be addressed in future research.

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, Jiang and Riesenhuber, 2009; Taylor, Rastle and Davis, 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).

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. Another possibility, suggested by the eccentricity bias (Hasson et al., 2002; Levy et al., 2001; Malach et al., 2002) and graded hemispheric lateralization (Behrmann and Plaut, 2015; Dehaene et al., 2015) theories is that competition on foveal resources for faces and words, produces an interactive development across domains, resulting in the left lateralization of word information and the right lateralization of face information (Behrmann and Plaut, 2015; Dehaene et al., 2015). These hypotheses may be tested in future longitudinal research.

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 more generally may be more malleable in childhood than adulthood. Future longitudinal investigations in children and illiterate adults as they obtain literacy can resolve this outstanding question.

In sum, the present 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 data suggest that development of distributed responses to words and characters may be influenced both by foveal biases and interactions with left-lateralized language areas and have important implications for both developmental theories and for understanding reading disabilities.

Author 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.

Acknowledgements

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 MN, NIH grants 1RO1EY02231801A1, 1RO1EY02391501A1 to KGS, training grant 5T32EY020485 supporting VN, and the NSF Graduate Research Development Program (grant DGE-114747) and Ruth L. Kirschstein National Research Service Award (grant F31EY027201) to JG.

References

Behrmann, M., and Plaut, D.C. (2015). A vision of graded hemispheric specialization. Ann. N. Y. Acad. Sci. *1359*, 30–46.

Ben-Shachar, M., Dougherty, R.F., and Wandell, B.A. (2007). White matter pathways in reading. Curr. Opin. Neurobiol. *17*, 258–270.

Ben-Shachar, M., Dougherty, R.F., Deutsch, G.K., and Wandell, B.A. (2011). The development of cortical sensitivity to visual word forms. J. Cogn. Neurosci. *23*, 2387–2399.

Binder, J.R., Medler, D.A., Westbury, C.F., Liebenthal, E., and 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, J. V., Guttorm, T.K., Martin, E., Lyytinen, H., Brandeis, D., and Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. Proc. Natl. Acad. Sci. *107*, 7939–7944.

Cantlon, J.F., Pinel, P., Dehaene, S., and Pelphrey, K.A. (2011). Cortical representations of symbols, objects, and faces are pruned back during early childhood. Cereb. Cortex *21*, 191–199.

Carlson, T., Tovar, D.A., Alink, A., and Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. J. Vis. *13*, 1.

Carreiras, M., Seghier, M.L., Baquero, S., Estévez, A., Lozano, A., Devlin, J.T., and Price, C.J. (2009). An anatomical signature for literacy. Nature *461*, 983–986.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.-A., and Michel, F. (2000). The visual word form areaSpatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain *123*, 291–307.

Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.C., Abdi, H., and Haxby, J. V (2012). The representation of biological classes in the human brain. J. Neurosci. *32*, 2608–2618.

Cox, D.D., and Savoy, R.L. (2003). Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage *19*, 261–270.

Dehaene, S., Le Clec’H, G., Poline, J.-B., Le Bihan, D., and 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, L.W., Ventura, P., Nunes Filho, G., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., and Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. Science (80-.). *330*, 1359–1364.

Dehaene, S., Cohen, L., Morais, J., and Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. Nat. Rev. Neurosci. *16*, 234–244.

Dehaene-Lambertz, G., Dehaene, S., and Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. Science (80-.). *298*, 2013–2015.

Dubois, J., de Berker, A.O., and Tsao, D.Y. (2015). Single-Unit Recordings in the Macaque

Face Patch System Reveal Limitations of fMRI MVPA. J. Neurosci. *35*, 2791 LP-2802.

Dumoulin, S.O., and Wandell, B.A. (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, L.S., Jiang, X., and Riesenhuber, M. (2009). Evidence for Highly Selective Neuronal Tuning to Whole Words in the "Visual Word Form Area." Neuron *62*, 199–204.

Glezer, L.S., Kim, J., Rule, J., Jiang, X., and 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, D.G., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J.L., Gabrieli, J.D., and 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., Yoon, J.M., and Grill-Spector, K. (2010). Differential development of the ventral visual cortex extends through adolescence. Front. Hum. Neurosci. *3*, 80.

Golarai, G., Liberman, A., and Grill-Spector, K. (2017). Experience Shapes the Development of Neural Substrates of Face Processing in Human Ventral Temporal Cortex. Cereb. Cortex bhv314.

Gomez, J., Natu, V., Jeska, B., Barnett, M., and Grill-Spector, K. (in press). Development differentially sculpts receptive fields across human visual cortex. Nat. Commun.

Gomez, J., Barnett, M.A., Natu, V., Mezer, A., Palomero-Gallagher, N., Weiner, K.S., Amunts, K., Zilles, K., and Grill-Spector, K. (2017b). Microstructural proliferation in human cortex is coupled with the development of face processing. Science (80-.). *355*, 68–71.

Grill-Spector, K., and Weiner, K.S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. Nat. Rev. Neurosci. *15*, 536–548.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron *24*, 187–203.

Gullick, M.M., and Booth, J.R. (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., and 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., and Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. Neuron *34*, 479–490.

Haxby, J. V, Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science (80-.). *293*, 2425–2430.

Kay, K.N., Winawer, J., Mezer, A., and Wandell, B.A. (2013). Compressive spatial summation in human visual cortex. J. Neurophysiol. *110*, 481–494.

Kay, K.N., Weiner, K.S., and Grill-Spector, K. (2015). Attention reduces spatial uncertainty in human ventral temporal cortex. Curr. Biol. *25*, 595–600.

Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. Front. Syst. Neurosci.

Le, R., Witthoft, N., Ben-Shachar, M., and Wandell, B. (2017). The field of view available to the ventral occipito-temporal reading circuitry. J. Vis. *17*, 6.

Levy, I., Hasson, U., Avidan, G., Hendler, T., and Malach, R. (2001). Center-periphery organization of human object areas. Nat Neurosci *4*, 533–539.

Malach, R., Levy, I., and Hasson, U. (2002). The topography of high-order human object areas. Trends Cogn Sci *6*, 176–184.

Mezer, A., Yeatman, J.D., Stikov, N., Kay, K.N., Cho, N.J., Dougherty, R.F., Perry, M.L., 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, V.S., Barnett, M.A., Hartley, J., Gomez, J., Stigliani, A., and Grill-Spector, K. (2016). Development of Neural Sensitivity to Face Identity Correlates with Perceptual Discriminability. J. Neurosci. *36*, 10893–10907.

Nordt, M., Hoehl, S., and Weigelt, S. (2016). The use of repetition suppression paradigms in developmental cognitive neuroscience. Cortex *80*, 61–75.

Norman, K.A., Polyn, S.M., Detre, G.J., and Haxby, J. V (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. *10*, 424–430.

Olulade, O.A., Flowers, D.L., Napoliello, E.M., and Eden, G.F. (2013). Developmental differences for word processing in the ventral stream. Brain Lang. *125*, 134–145.

Rauschecker, A.M., Bowen, R.F., Perry, L.M., Kevan, A.M., Dougherty, R.F., and Wandell, B.A. (2011). Visual feature-tolerance in the reading network. Neuron *71*, 941–953.

Schlaggar, B.L., and McCandliss, B.D. (2007). Development of Neural Systems for Reading. Annu. Rev. Neurosci. *30*, 475–503.

Srihasam, K., Mandeville, J.B., Morocz, I.A., Sullivan, K.J., and Livingstone, M.S. (2012). Behavioral and Anatomical Consequences of Early versus Late Symbol Training in Macaques. Neuron *73*, 608–619.

Stigliani, A., Weiner, K.S., and Grill-Spector, K. (2015). Temporal Processing Capacity in High-Level Visual Cortex Is Domain Specific. J. Neurosci. *35*, 12412–12424.

Takeuchi, H., Taki, Y., Hashizume, H., Asano, K., Asano, M., Sassa, Y., Yokota, S., Kotozaki, Y., Nouchi, R., and Kawashima, R. (2016). Impact of reading habit on white matter structure: Cross-sectional and longitudinal analyses. Neuroimage *133*, 378–389.

Taylor, J.S.H., Rastle, K., and Davis, M.H. (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, J.P., Sigman, M., and 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, B.A., Rauschecker, A.M., and Yeatman, J.D. (2012). Learning to see words. Annu Rev Psychol *63*, 31–53.

Weiner, K.S., and Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. Neuroimage *52*, 1559–1573.

Weiner, K.S., and Grill-Spector, K. (2011). Not one extrastriate body area: Using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. Neuroimage *56*, 2183–2199.

Weiner, K.S., Golarai, G., Caspers, J., Chuapoco, M.R., Mohlberg, H., Zilles, K., Amunts, K., and 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.

Yeatman, J.D., Dougherty, R.F., Ben-Shachar, M., and Wandell, B.A. (2012). Development of white matter and reading skills. Proc. Natl. Acad. Sci. U. S. A. *109*, E3045-53.

Figure Legends

Figure 1. Stimuli, anatomical ROI definitions, and between age-group controls (a) Examples of stimuli corresponding to the five 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 three 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.

Figure 2. Within-domain generalization and between-domain dissimilarity 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. The correlations between MVPs are shown for (a) Left lateral VTC; all comparisons showed a significant main effect of age as indicated by the asterisk ($Fs(2,48) > 3.34$; $ps < 0.04$). (b) Right lateral VTC; w–n and n–nn showed a significant effect of age ($Fs(2,48) > 4.05$; $ps < 0.02$). (c) Left medial VTC. (d) Right medial VTC. Boxplots are colored by age; *light blue*: 5-9-year-old, 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 for a comparison, $p < 0.05$. *See also Figure S1*.

Figure 3. **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 S2 and S3.*

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,e,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*.

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. 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).

STAR Methods

CONTACT FOR RESOURCE SHARING

Further information and request for resources should be directed to and will be fulfilled by the Lead Contact, Kalanit Grill-Spector (kalanit@stanford.edu).

EXPERIMENTAL MODEL AND SUBJECTS DETAILS

66 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 three 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 typical. 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.

Participants underwent anatomical and functional MRIs. Prior to MRI, children were trained in a scanner stimulator. 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.

METHODS DETAILS

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 (α = 4, 10, 20, and 30°; TR=14 ms; TE=2.4 ms. Voxel size= 0.8mm × 0.8mm × 1mm, resampled to 1mm^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 mm^2; 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=1s; TE=30ms; flip angle=76°; FOV = 192 mm; 2.4mm isotropic voxels; one-shot T2*-sensitive gradient echo sequence).

fMRI 5 domain/10 category localizer

Participants completed 3 runs of the fMRI experiment. Each run lasted 5 min and 24 s. During fMRI participants viewed images from five 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 (**Figure 1a**). Images were presented at a rate of 2Hz, in 4s 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.

QUANTIFICATION AND STATISTICAL ANALYSIS

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

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) (**Figure 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

Data were processed aligned to each participant's native brain anatomy and were not spatially smoothed. 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 (**Figure 1c**). The time courses of each voxel were transformed into percentage signal change and a general linear model (GLM) was fit to each voxel to estimate the contribution of each of the 10 conditions.

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: $\sqrt{residual\ GLM\ variance/_{df}}$ (df=degrees of freedom). To evaluate similarity between MVPs we calculated all pair-wise correlations between MVP pairs resulting in a 10x10 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 six 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) vs. 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) vs. 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 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 four consecutive errors or has completed reading the list.

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

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 (**Figure 4**), and (3) the remaining nonword-distinctive lateral VTC voxels (**Figure 4**). The same analyses were done for 30% of lateral VTC voxels that where most discriminative for characters (**Figure 4, Figure S4**). We evaluated if the correlations in (2) and (3) were significantly different and within (2) if information in most or least selective voxels correlated with reading ability.

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.

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.

In analyses related to **Figure 2** 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 (w-w; b-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 (w-nw; 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 2** with asterisks.

In analyses related to **Figure 3** 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.

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 four 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.

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 and their significance. 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/corrtest/corrtest2.htm), which includes converting the correlation coefficients into z-scores using Fisher's r-to-z transformation.

Analysis of V1 MVPs

V1 was defined using data from a separate retinotopic mapping experiment. 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 2Hz, 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; Weiner and Grill-Spector, 2011). 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 minutes and 24 seconds. 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 x 2, and 28 slices. 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 an the inflated cortical surface of each subject's brain. V1 was defined in each hemisphere as the visual field map containing a hemifield 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 S2b** indicate that the domain information in VTC is significantly higher than V1 (main effect of area, $F(2,72) = 142.1$, $p < .001$; lateral VTC > V1, $t(38) = 16.32$, $p < .001$; medial VTC > V1, $t(38) = 10.13$, $p < .001$).

DATA AND SOFTWARE AVAILABILITY

All code relevant to data analysis for the main findings (Figures 1-5) will be available on github.com/VPNL. 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 vistasoft code library: https://github.com/vistalab/vistasoft