

Longitudinal development of category representations in ventral temporal cortex predicts word and face recognition

Marisa Nordt^{1,2*}, Jesse Gomez³, Vaidehi S. Natu¹, Alex A. Rezai¹, Dawn Finzi¹, Holly Kular¹, Kalanit Grill-Spector^{1,4,5}

¹ Department of Psychology, Stanford University, Stanford, CA, US.

² University Hospital Aachen, RWTH Aachen, Aachen, Germany.

³ Princeton Neuroscience Institute, Princeton University, Princeton, NJ, US.

⁴ Neurosciences Program, Stanford University, Stanford, CA, US.

⁵ Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, US.

Abstract

Regions in ventral temporal cortex (VTC) that are involved in visual recognition of categories like words and faces, undergo differential development during childhood. However, categories are also represented in distributed responses across VTC. How do distributed category representations develop and relate to behavioral changes in recognition? Here, we used fMRI to longitudinally measure the development of distributed responses across VTC to 10 categories in school-age children over several years. Our results reveal both strengthening and weakening of category representations with age, which was mainly driven by changes across category-selective voxels. Representations became particularly more distinct for words in the left hemisphere and for faces bilaterally. Critically, distinctiveness for words and faces across category-selective voxels in left and right lateral VTC, respectively, predicted individual children's word and face recognition performance. These results suggest that the development of distributed VTC representations has behavioral ramifications and advance our understanding of prolonged cortical development during childhood.

Recognizing faces and written words is important for our everyday life. While extensive experience with faces starts from birth¹ extensive experience with written words typically begins around the age of 5-6 with the onset of formal reading education. Nonetheless, both skills continue to improve throughout childhood and adolescence^{2,3}. Visual recognition is thought to involve the end stage of the human ventral visual processing stream – that is, ventral temporal cortex (VTC). Indeed, VTC contains both (i) clustered regions that are selective to and causally involved in the perception of ecological relevant

categories^{4–6} such as faces⁷, limbs^{8,9}, places¹⁰, and words¹¹, as well as (ii) distributed representations that are reproducibly unique to each category^{12–14}. That is, stimuli from different categories generate distinct patterns of response across VTC, even stimuli that are not associated with clustered regions and when excluding category-selective regions altogether. However, it is unknown if and how distributed category representations in VTC develop longitudinally during childhood and if these developments are linked with behavioral improvements in visual recognition.

Understanding the longitudinal development of distributed representations during childhood is important for three main reasons. First, to date, examinations of the development of distributed VTC responses has been largely confined to cross-sectional studies which compared representations in children to adults using a limited number of categories with inconsistent results^{15–18}. Where some studies have reported that distributed responses to faces, objects and places in children are similar to adults^{15,17}, other studies have reported that own-age face representations¹⁷ and word representations¹⁸ are enhanced from childhood to adulthood. Thus, to understand how distributed VTC representations change during childhood it is necessary to measure the longitudinal development of distributed representations within the same children and with many categories. Second, much of the prior developmental research has examined the development of category selectivity within clustered regions selective to ecologically-relevant categories^{16,19–25}. However, the development of distributed responses may be different than the development of category representations in clustered regions as different signals contribute to distributed compared to clustered responses. That is, the distributed pattern of response is determined by all values of responses (both high and low) across VTC, whereas in clustered regions the selectivity is driven by the highest responses²⁶. Third, examining the relation between development of distributed VTC responses and development of face recognition and reading abilities will address a key debate in the field. Researchers argue whether the entire distributed pattern of response across VTC¹³ or the specific distributed responses over the selective voxels^{18,27–29} give rise to developmental improvements in visual recognition behavior.

How may distributed category representations in VTC develop during childhood?

Prior research has found that clustered regions selective to faces and words become larger and more selective to their respective categories from childhood (4-7 year old) to adolescence (13-17 year old) to adulthood (>18)^{16,19-25}. Recent longitudinal research in children has surprisingly found that these increases in face- and word-selectivity in lateral VTC are coupled with decreases in limb-selectivity and, in fact, regions that are limb-selective earlier in childhood become selective to faces and words by adolescence²³. These findings suggest that changes in category selectivity may affect the nature of distributed VTC representations. One possibility is that developmental increases in category-selectivity would lead to increases in the distinctiveness of distributed responses for categories associated with a clustered region in VTC. For instance, developmental increases in word- and face-selectivity would increase the distinctiveness of distributed responses to words and faces, making distributed responses both more consistent across items of their respective category and more different from distributed responses to items of other categories. This hypothesis also predicts that developmental decreases in limb-selectivity would lead to decreases in the distinctiveness of distributed responses to limbs. A second possibility is that distributed responses in VTC develop for many categories beyond faces, limbs, and words, and also across voxels that don't have a strong preference for any category. This hypothesis predicts that developmental changes will be observed also for categories, such as numbers, for which there is no clustered region in VTC and also across voxels that have no specific selectivity to any category. A third possibility is that because clustered regions selective to faces, words, and limbs only constitute a minority of VTC voxels¹⁸, changes in these regions will have little effect on large scale distributed patterns. Thus, this hypothesis predicts no significant development of distributed VTC responses during childhood, consistent with cross-sectional data reporting adult-like distributed responses to faces, objects, and scenes by age 7^{15,17}.

These hypotheses also make different behavioral predictions. The first hypothesis predicts that if distributed representations for faces and words become more distinct over childhood development, then face recognition and reading performance will concurrently improve. This hypothesis also predicts that behavioral changes in face recognition and

reading will be coupled with specific increases in the distinctiveness of distributed responses over the category-selective voxels. The second hypothesis predicts that developmental increases in distinctiveness of distributed responses across the entire VTC rather than over category-selective voxels will be coupled with behavioral improvements in face recognition and reading, and this relationship may hold even when voxels of category-selective regions are excluded from the distributed response across VTC. In contrast, the third hypothesis does not predict a relationship between development of face recognition and reading performance and their respective distributed representations in VTC as the latter are predicted to stay stable during childhood development.

Here, we tested these predictions using longitudinal measurements of distributed responses to many visual categories as well as behavioral assessments of face recognition and reading in the same school-age children over several years. These longitudinal measurements are crucial not only for tracking the development of brain and behavior within the same child over several years, but also for evaluating the rate of development of distributed responses both for categories with and without clustered regions of strong selectivity in VTC. Thus, we collected longitudinal functional magnetic resonance imaging (fMRI) and behavioral data in 29 school-age children over a span of 1 to 5 years (mean \pm SD: 3.75 \pm 1.5 years, 4.4 \pm 1.92 sessions per child) totaling 128 fMRI sessions and 146 behavioral datasets (Fig. S1A). During the fMRI experiment, children viewed overall 1440 images from 10 categories spanning 5 domains of ecological relevance (Fig. S1B). These included characters (pseudowords, numbers), faces (adult faces, child faces), body parts (headless bodies, limbs), objects (cars, string instruments), and places (houses, corridors). We measured in each session distributed responses for each of the 10 categories across VTC and examined if they changed as children aged. To relate brain development to visual recognition performance, we measured in the same children their face recognition and reading performance outside the scanner (face recognition: 29/29 children, 2.83 \pm 1.0 sessions per child collected over 3.38 \pm 1.5 years, reading: 26/29 children, 2.21 \pm 1.1 sessions per child collected over 2.86 \pm 1.19 years). Then, we tested if there is a relationship between behavioral and brain development.

How do category representations in VTC develop longitudinally?

To assess the nature of distributed category representations in children, we computed the distributed pattern of responses for each category. As face-, limb-, and word-selective regions are located in lateral aspect of VTC^{7,8,11,30} and development of word- and face-selectivity varies across hemispheres^{16,18,21,31}, we measured in each child and session distributed responses across lateral VTC (Fig. 1B, see Methods) to each of the 10 categories, separately for the right and left hemispheres. Vectors of distributed responses to each category – also called multivoxel patterns (MVPs) – were computed independently for each of the two functional runs in each session in which participants viewed different images. We then calculated correlations between all pairs of MVPs (run-1 to run-2), resulting in a 10x10 representational similarity matrix (RSM¹⁴, for each child and session (Fig. 1A). On-diagonal values in the RSM illustrate how similar distributed responses are across different images of the same category, and off-diagonals indicate how similar are distributed responses are to images of different categories. Examining individual RSMs revealed that even in young children there is category structure in distributed VTC responses as on-diagonal values are positive and higher than off-diagonal values (Fig.1A).

To quantify category information, we used the RSM to measure the distinctiveness of distributed responses for each category. The distinctiveness of a category is defined as the difference between within-category similarity and between-category similarity (Fig. 1A, gray box); distinctiveness ranges from 0-2. Higher distinctiveness indicates that distributed responses to a category are highly similar across different images of the category and highly dissimilar from distributed responses to other categories. We measured distinctiveness separately for each category and hemisphere in each of the 128 sessions. Then, we tested if category distinctiveness develops over childhood using linear mixed models (LMMs) relating category distinctiveness and age, with participant as a random factor (Fig. 1C).

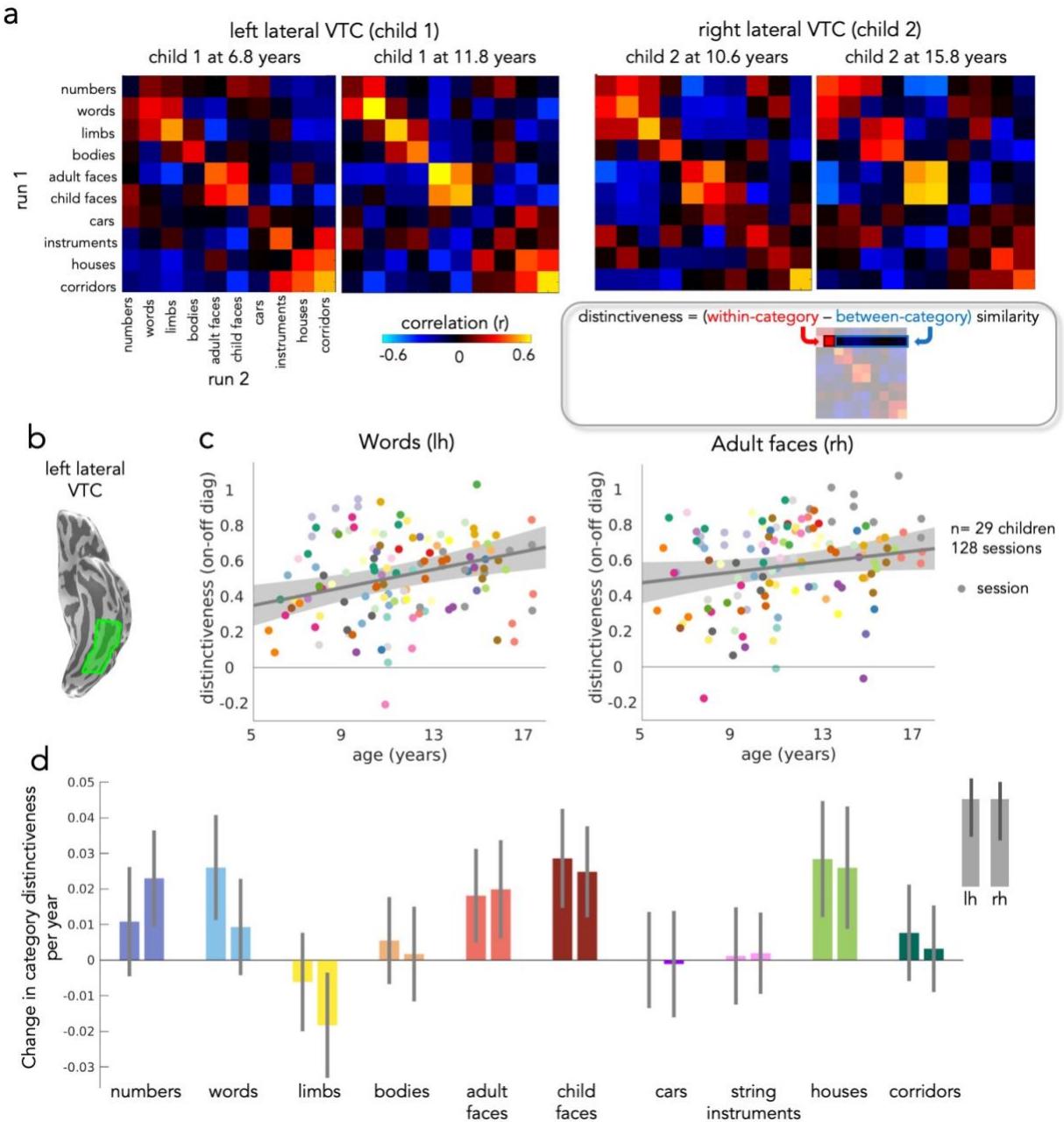


Fig. 1. Differential longitudinal development of category representations in children's lateral ventral temporal cortex (VTC). (A) Representational similarity matrices (RSM) of left and right lateral VTC in individual sessions of two children at different ages. Gray box: Schematic illustrating how distinctiveness is computed for each category. (B) Left lateral VTC on the ventral inflated surface of an example participant. (C) Scatter plots illustrate the relationship between distinctiveness and age. Gray line: Linear mixed model (LMM) prediction of distinctiveness by age (random intercept model with subject as a random effect). Shaded gray: 95% confidence interval (CI). Participants are coded by color. Each dot is a session. Note that the statistics in the text and the data in (D) report developmental results from an LMM that includes two factors: age and tSNR. (D) LMM slopes indicating change in distinctiveness per year (LMM relating distinctiveness to age, with tSNR as an independent factor, and participant as random effect, $n=128$ sessions, 29 children) for each category; for space, we refer to pseudowords as words. Error bars: 95% confidence interval (CI). If the CI does not cross the $y=0$ line, the change in distinctiveness is significantly different than 0.

To ensure that developmental effects are not driven by differences in scan quality across age, we first tested whether motion during scanning and timeseries signal-to-noise ratio (tSNR) contribute to measures of distinctiveness. Adding motion as a predictor to the LMM did not significantly contribute to the model fit except for distinctiveness for string instruments (no significant contribution of age to distinctiveness of string instruments with or without adding motion). Adding tSNR as a predictor to the LMM contributed to the model independent from age for several categories. Thus, we include tSNR as an additional predictor in the LMM (Table 1) and we report age-related changes in distinctiveness that are independent from tSNR.

Example scatter plots show the distinctiveness for pseudowords (Fig. 1C-left) and adult faces (Fig. 1C-right) as a function of a child's age. As you can see in these examples, distinctiveness for pseudowords in the left lateral VTC and distinctiveness for adult faces in the right lateral VTC steadily increases from age 5 to 17. The slope of the LMM summarizes the development of category distinctiveness with age. A positive slope indicates that category distinctiveness increases from age 5 to 17 and a negative slope indicates that category distinctiveness decreases across childhood.

Results reveal differential development of category distinctiveness in lateral VTC that varied by category and hemisphere (Fig. 1D) and that was not limited to categories with a clustered region in lateral VTC. Consistent with the first hypothesis, we find increases in distinctiveness for pseudowords and faces and decreases in distinctiveness for limbs: Distinctiveness for pseudowords increased significantly with age in the left lateral VTC ($\beta_{age}=0.026$, $t(125)=3.49$, $p_{FDR}=0.0038$, Fig.1C,D), but not right lateral VTC ($\beta_{age}=0.009$, $t(125)=1.36$, $p_{FDR}=0.32$). Distinctiveness for both adult faces (left: $\beta_{age}=0.018$, $t(125)=2.72$, $p_{FDR}=0.019$; right: $\beta_{age}=0.02$, $t(125)=2.85$, $p_{FDR}=0.015$) and child faces (left: $\beta_{age}=0.029$, $t(125)=4.06$, $p_{FDR}=0.0017$; right: $\beta_{age}=0.025$, $t(125)=3.84$, $p_{FDR}=0.002$) increased significantly with age in both hemispheres (Fig.1C,D). In contrast to these developmental increases, distinctiveness for limbs decreased significantly with age in the right hemisphere ($\beta_{age}=-0.018$, $t(125)=-2.45$, $p_{FDR}=0.035$) while distinctiveness for bodies remained largely unchanged from age 5 to 17. Consistent with the second hypothesis,

we find increases in distinctiveness even for categories that do not have a clustered region in lateral VTC. Specifically, (i) distinctiveness for numbers increased significantly in the right lateral VTC ($\beta_{\text{age}}=0.023$, $t(125)=3.36$, $p_{\text{FDR}}=0.004$); while there was no significant development in left lateral VTC ($\beta_{\text{age}}=0.01$, $t(125)=1.39$, $p_{\text{FDR}}=0.32$) and (ii) distinctiveness for houses increased significantly bilaterally (left: $\beta_{\text{age}}=0.028$, $t(125)=3.45$, $p_{\text{FDR}}=0.0038$, right: $\beta_{\text{age}}=0.026$, $t(125)=2.98$, $p_{\text{FDR}}=0.01$). We find no evidence for other significant age-related changes (Table 1). Overall, these analyses reveal differential development of distributed category representation in lateral VTC from age 5 to 17.

Which voxels drive the development of distributed representations?

We next asked which voxels in lateral VTC drive the development of category representation: Is the development of distinctiveness driven by category-selective voxels, non-selective voxels, or both? We reasoned that if changes in distributed responses are driven by the development of category selectivity, then the development of category distinctiveness will be evident in the selective but not in the non-selective voxels of lateral VTC. Alternatively, voxels that are selective may be already developed, predicting that the non-selective voxels are driving the observed development of category distinctiveness. A third possibility is that category information is carried by the relative response across the entire neural population and in fact there is nothing special about the selective voxels (Haxby et al., 2001). This hypothesis predicts that development of category distinctiveness is driven by all voxels of lateral VTC, including both the selective and non-selective voxels.

To test these predictions, we separated lateral VTC into two sets of voxels – those which were category selective, and the rest that were not selective to any category. That is, for each session, we first identified lateral VTC voxels which were selective ($t>3$, voxel-level, Methods) for each of the 10 categories, then took the union of these voxels across the 10 categories to generate the set of all category-selective voxels. Non-selective voxels were defined as the remainder of lateral VTC voxels and were not selective to any of these categories. We next computed LMMs relating distinctiveness to age, with tSNR as an independent factor, in these two subsets of voxels (see Tables 2 & 3 for all model

parameters). We report the slope of the LMM indicating the effect of age for each category in Fig. 2. Although there are overall fewer selective voxels (left: 39.24% \pm 11; right: 34% \pm 11%) than non-selective voxels (left: 60.76% \pm 11; right: 66% \pm 11%) we find a significant development of category distinctiveness across the union of selective voxels of lateral VTC (Fig. 2-maroon & pink bars), but no evidence for significant development of distinctiveness across the non-selective voxels of lateral VTC (Fig. 2-gray bars). These findings were not due to higher tSNR in the union of selective voxels compared to the non-selective voxels (no significant effect of voxel subset in left lateral VTC: $\beta_{\text{subset_Selective}}=1.8$, $t(254)=1.34$, $p=0.18$ and higher tSNR in non-selective vs. selective voxels in right lateral VTC: $\beta_{\text{subset_Selective}}=-3.4$, $t(254)=-2.47$, $p=0.01$, LMM with binary predictor voxel subset).

In fact, the development of category distinctiveness in the subset of selective voxels largely replicates the findings when considering all lateral VTC voxels (compare Figs. 1D and 2). That is, in the union of selective voxels in the left hemisphere, we find a significant increase in the distinctiveness for pseudowords ($\beta_{\text{age}}=0.026$, $t(125)=3.99$, $p_{\text{FDR}}=0.002$), faces (adult: $\beta_{\text{age}}=0.026$, $t(125)=3.88$, $p_{\text{FDR}}=0.002$; child: $\beta_{\text{age}}=0.034$, $t(125)=4.39$, $p_{\text{FDR}}=0.0009$), and houses ($\beta_{\text{age}}=0.036$, $t(125)=3.69$, $p_{\text{FDR}}=0.003$). In the right hemisphere, we find a significant increase in the distinctiveness for numbers ($\beta_{\text{age}}=0.027$, $t(125)=3.08$, $p_{\text{FDR}}=0.017$), faces (adult: $\beta_{\text{age}}=0.020$, $t(125)=2.95$, $p_{\text{FDR}}=0.02$; child: $\beta_{\text{age}}=0.017$, $t(125)=2.75$, $p_{\text{FDR}}=0.03$), and houses ($\beta_{\text{age}}=0.034$, $t(125)=3.49$, $p_{\text{FDR}}=0.005$). The decrease in the distinctiveness for limbs was not significant in this subset of voxels ($\beta_{\text{age}}=-0.01$, $t(125)=-1.67$, $p_{\text{FDR}}=0.33$). In contrast, we find no evidence for significant changes in distinctiveness for any category in the non-selective voxels of lateral VTC (Fig. 2-gray bars, Table 3).

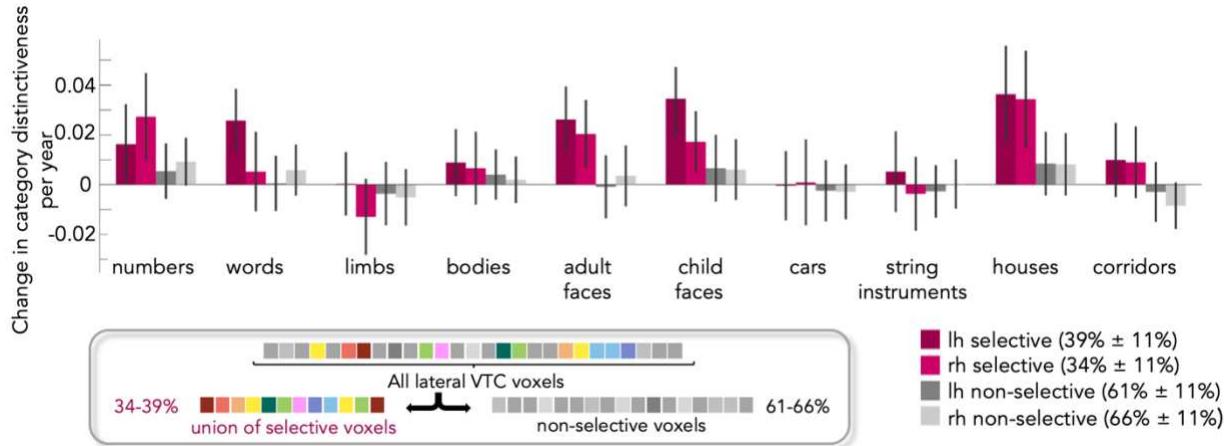


Fig. 2. Development of distributed representation only in the selective voxels in lateral VTC. Bars indicate the change in category distinctiveness per year (LMM relating distinctiveness to age and tSNR, with participant as a random effect, n=128 sessions, 29 children) in different subsets of voxels. *Maroon bars*: the union of voxels in lateral VTC that were selective to one of the 10 categories. Category-selectivity was computed by contrasting responses to a category vs. all other categories except the other category from the same domain (e.g., numbers vs. all other categories except words). A voxel was defined as selective to a category when $t > 3$. Overall ~36.5% of lateral VTC voxels were selective to one of the categories (see schematic in box). *Gray bars*: the remainder, the non-selective voxels of lateral VTC that were not selective to any of these categories. *Darker colors*: left hemisphere. *Lighter colors*: right hemisphere. *Error bars*: 95% CI. If the CI does not cross the y=0 line, the change in distinctiveness is significantly different than 0.

How does the nature of the internal representational space change from childhood to adolescence?

As there is a heterogeneous development of category distinctiveness both across categories and subsets of VTC voxels, it is interesting to consider how this relates to the internal representational space of these ten categories. To visualize the representational space and how it changes with age, we computed mean RSMs for 5-9-year-olds and 13-17-year-olds (Fig. S2A). We then used multidimensional scaling (MDS) to visualize the representational space in 2D.

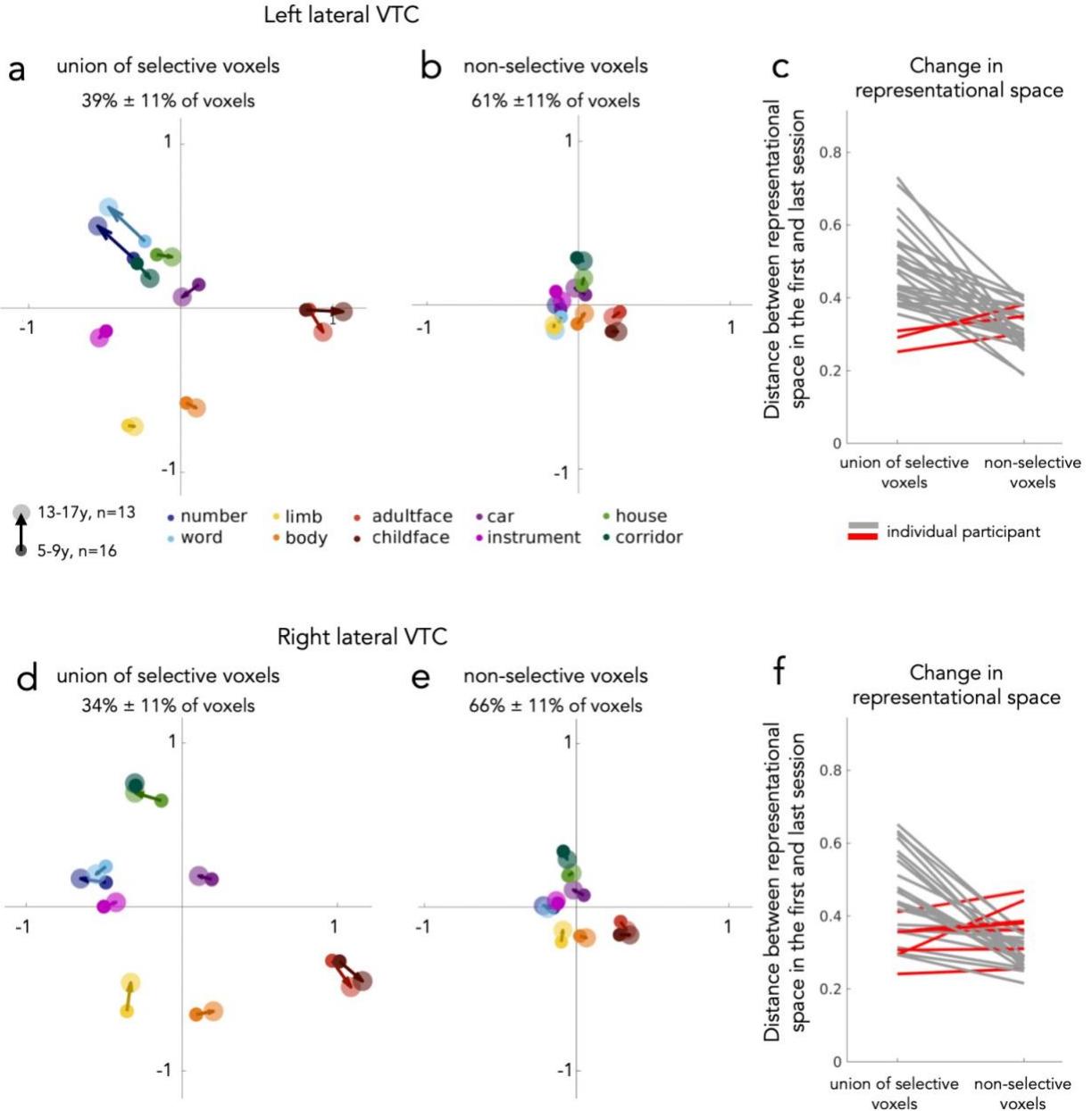


Fig. 3 Development of the representational space of the 10 categories in lateral VTC. Multidimensional scaling (MDS) embeddings for the category representation in different subsets of voxels for two age groups: 5-9-year-olds (n=16 participants, small circles) and 13-17-year-olds (n=13 participants, larger circles). One session per child is included per MDS of each age group. (A,D) MDS embedding of the representational space across the union of selective voxels in left (A) and right (D) lateral VTC. (B,E) MDS embedding of the representational space of the remainder, non-selective voxels of left (B) and right (E) lateral VTC. (C,F) Line plots depicting the change in representational spaces in individual children in left (C) and right (F) lateral VTC across the selective and non-selective voxels. The change in representation is the mean Euclidian distance between category positions in the MDS embedding of a child's first session vs their last session. Each line is a participant (n=29); Gray: larger distances in the selective voxels; Red: larger distances in the non-selective voxels.

This visualization illustrates two key findings.

First, distributed representations across both the union of selective voxels (Fig. 3,A,D, Supplemental movie 1) and non-selective voxels (Fig. 3B,E) reveal a categorical structure, however, the categorical representation over nonselective voxels is strongly diminished compared to that over the union of selective voxels (Fig. 3, compare A&D to B&E). That is, MDS embeddings over both the union of selective voxels (Fig. 3A,D) and over the non-selective voxels (Fig. 3B,E) reveal that representations of animate stimuli (faces, bodies) are largely separate from those of inanimate stimuli (objects, places, characters) in both children and adolescents. Yet, comparing the embeddings of the union of selective voxels (Fig. 3A,D) to that of the non-selective voxels (Fig. 3B,E) reveals that the representation of category information is much clearer and enhanced over the selective voxels.

Second, examining the representational structure over the union of selective voxels from childhood (small circles) to adolescence (large circles) reveals developmental changes. In contrast the representational structure of the non-selective voxels remains largely unchanged. Examining the MDS embeddings over the union of the selective voxels reveals development of the categorical structure in several ways: (i) in both hemispheres representations of faces strengthen and become more separable from other categories from childhood to the teens (Fig. 3A,D, red arrows moving outward, see developmental trajectory in Supplemental Movie 1), (ii) representations of both pseudowords and numbers strengthen from age 5 to age 17, and this development is particularly pronounced in the left hemisphere (Fig. 3A,D, blue arrows moving outward, Supplemental Movie 1), and (iii) representations of limbs in the right hemisphere weaken from childhood to the teens (Fig. 3D, Supplemental movie 1, yellow arrow moving inward). In contrast, comparing representations over the non-selective voxels in children and adolescents suggests no substantial development.

We quantified the development of the representational structure in both selective and non-selective voxels in each child (Fig. 3C,F, Methods). Thus, for each set of voxels we aligned the MDS embedding of each child's first session to that of their last session and

measured the mean distance between the coordinates of each category in this shared embedded space. This analysis reveals that there is a significantly larger developmental change in the representation over the union of selective voxels than the non-selective voxels in both hemispheres (left: $t(28)=6.40$, $p<0.001$, t-test Fig. 3C, and right: $t(28)=4.92$, $p<0.001$, Fig. 3F). Importantly, this effect is visible in most individual children illustrating within-child development of distributed responses (Fig. 3C,F). In sum, these analyses reveal that while there is categorical structure in both sets of voxels (union of selective, non-selective), the representation in the union of selective voxels (i) is enhanced compared to that of the non-selective voxels, and (ii) undergoes stronger development.

Is development of distributed responses linked to improvements in behavior?

So far, accumulating evidence reveals the enhancement of distributed category representations in lateral VTC from age 5 to 17. Given that face recognition and reading also improve from age 5 to 17, we measured face recognition and reading ability in our longitudinal sample (Fig. S3) and tested if these developments are linked. We hypothesized that developmental increases in distinctiveness for faces and pseudowords may enhance recognition performance for these categories. We tested this hypothesis using LMMs relating behavioral performance to category distinctiveness, with participant as a random factor.

We found a significant and positive relationship between reading performance of pseudowords and distinctiveness for pseudowords over the union of the selective voxels of left lateral VTC (Fig. 4A, $\beta_{\text{distinctiveness}}=40.19$, $t(62)=2.94$, $p=0.005$, LMM, random slope and intercept across participants). That is, better reading scores were associated with higher values of distinctiveness for pseudowords. The effect of distinctiveness predicting reading performance remained significant when age was added to the LMM ($\beta_{\text{distinctiveness}}=28.92$, $t(61)=2.4$, $p=0.02$; $\beta_{\text{age}}=2.45$, $t(61)=3.44$, $p=0.001$), showing that the effect of distinctiveness was independent from the effect of age. Importantly, this link between reading scores and distinctiveness for pseudowords in left lateral VTC was specific, as no significant link was found between reading performance and (i) pseudoword distinctiveness over the non-selective voxels of left lateral VTC

($\beta_{\text{distinctiveness}}=14.96$, $t(62)=1.29$, $p=0.20$), (ii) pseudoword distinctiveness over the selective voxels in right lateral VTC ($\beta_{\text{distinctiveness}}=-6.90$, $t(62)=-0.50$, $p=0.62$), or (iii) face distinctiveness over the selective voxels in either hemisphere (left: $\beta_{\text{distinctiveness}}=6.0$, $t(62)=0.51$, $p=0.61$; right: $\beta_{\text{distinctiveness}}=0.72$, $t(62)=0.05$, $p=0.96$).

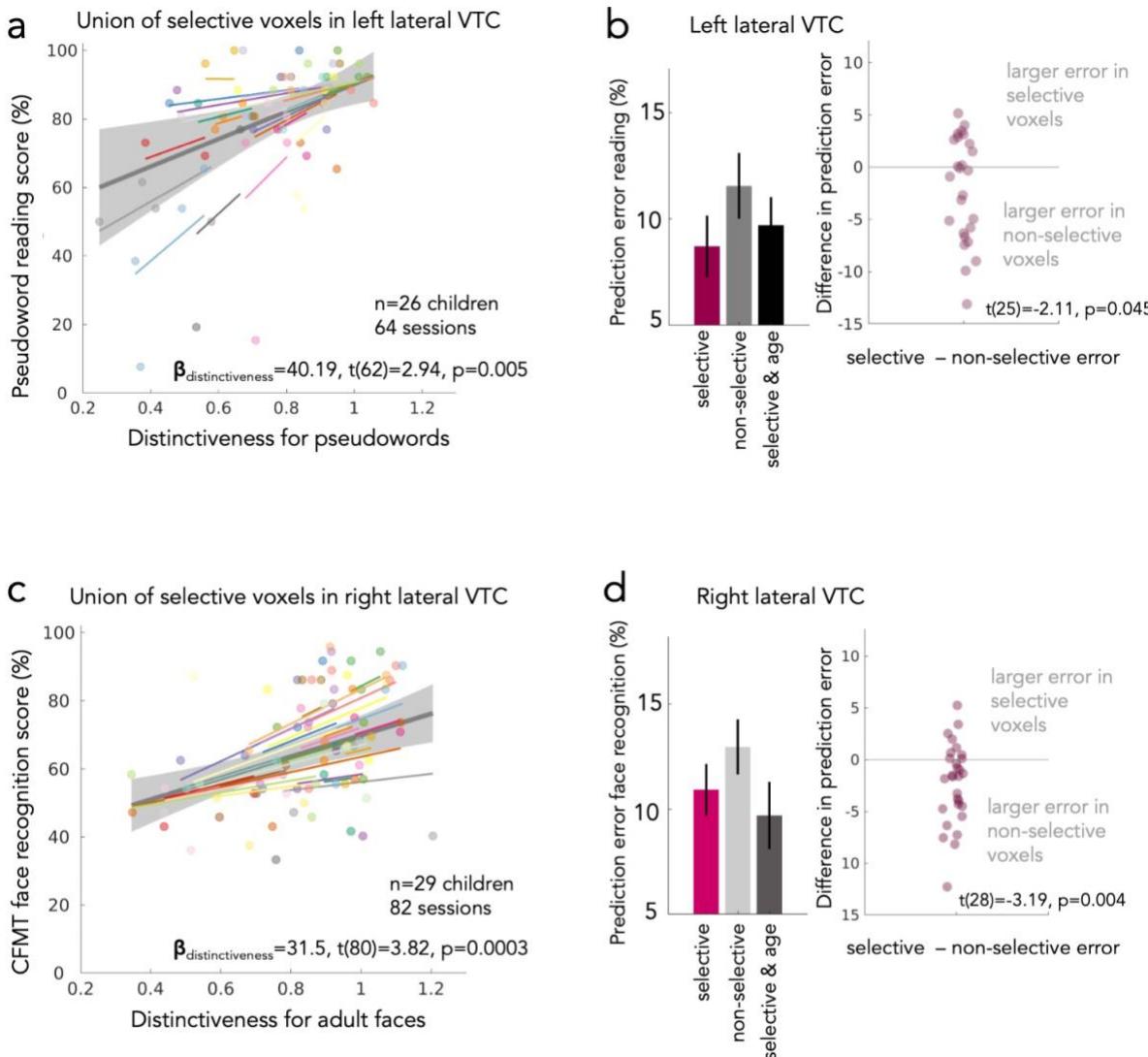


Fig. 4. Distinctiveness for words and faces in left and right lateral VTC, respectively, predict reading and face recognition performance in individual children. (A) Linear mixed model (LMM) with random slopes and intercepts relating reading performance of pseudowords (Woodcock Reading Mastery Test, WRMT) to pseudoword distinctiveness over the union of selective voxels of left lateral VTC. Model parameters indicated in the bottom. Each dot is a session; Dots are colored by participant. Colored lines: individual slopes and intercepts. Thick gray line: LMM prediction. Shaded gray: 95% CI. (B) Left: Median error in predicting the reading performance of a left-out subject from their distinctiveness for pseudowords in left lateral VTC using the parameters derived from the LMM of the rest of the subjects (29-fold leave-one-out-cross validation). Higher values indicate worse model prediction. The bars show the prediction error for three different models: *Selective*: LMM predicting behavior from distinctiveness over the union of selective voxels. *Non-selective voxels*: LMM predicting behavior from distinctiveness over the non-selective voxels. *Selective & age*: LMM predicting behavior from distinctiveness over the union of selective voxels with age as an additional factor. *Error bars*: standard error across participants. Right: Swarm plots showing the difference between the

prediction error for selective vs. non-selective voxels. Each dot is a participant. Statistics (at bottom) test if the difference in error is significantly different than zero. (C) Same as (A) but for LMM parameters for a model relating face recognition performance (Cambridge face recognition memory test (CFMT), adult faces) and distinctiveness for adult faces over the selective voxels of right lateral VTC. (D) Same as (B) but for face recognition performance and distinctiveness for adult faces in right lateral VTC.

We further reasoned that if distinctiveness predicts behavior, then this model can be used to predict reading performance in new subjects just from the distinctiveness of their lateral VTC selective voxels. We tested this prediction using a leave-one participant-out cross validation (LOOCV) approach. Results reveal that distinctiveness for pseudowords in the union of selective voxels in left lateral VTC successfully predicts reading performance in left out subjects with a median prediction error of 8.7% (Fig. 4B, purple bar). Adding age to the model did not further reduce the prediction error (Fig. 3B, black bar; no significant difference in performance, $t(25)=0.47$, $p=0.64$). In contrast, predicting reading performance from distinctiveness for pseudowords in non-selective voxels revealed a larger median prediction error of 11.5% (Fig. 4B, gray bar). Indeed, for most subjects the prediction error was larger for a model based on distinctiveness of the non-selective voxels than a model based on distinctiveness of the selective voxels (Fig. 4B-swarm plot, two-sided t-test comparing the difference in error to zero: $t(25)=-2.11$, $p=0.045$).

Likewise, distinctiveness for adult faces in the union of selective voxels of right lateral VTC was significantly and positively related to face recognition performance ($\beta_{\text{distinctiveness}}=31.49$, $t(80)=3.82$, $p=0.0003$, LMM with random slope and random intercept across participants). That is, better face recognition performance was associated with higher values of distinctiveness for (Fig. 4C). When age was added to the LMM the effect of distinctiveness remained marginally significant ($\beta_{\text{distinctiveness}}=16.41$, $t(79)=1.97$, $p=0.053$; $\beta_{\text{age}}=3.94$, $t(79)=7.89$, $p<0.001$). Crucially, the link between face recognition and face distinctiveness was specific to the selective voxels of right lateral VTC, as no such effect was found for (i) face distinctiveness in the non-selective voxels of right lateral VTC ($\beta_{\text{distinctiveness}}=0.85$, $t(80)=0.09$, $p=0.93$) or (ii) pseudoword distinctiveness in the selective voxels in right lateral VTC ($\beta_{\text{distinctiveness}}=-8.7$, $t(80)=-1.18$, $p=0.24$). While face recognition ability was also linked to distinctiveness for words in the left hemisphere ($\beta_{\text{distinctiveness}}=24.56$, $t(80)=2.75$, $p=0.007$), this link was driven by age, as there was no significant relationship when age was added to the model ($\beta_{\text{distinctiveness}}=-2.69$, $t(79)=-0.36$,

$p=0.73$, $\beta_{\text{age}}=4.18$, $t(79)=7.83$, $p<0.001$). Similarly, while face recognition ability was marginally linked with distinctiveness for adult faces in the left lateral VTC ($\beta_{\text{distinctiveness}}=17.79$, $t(80)=1.97$, $p=0.052$), it was not significant when age was added to the model ($\beta_{\text{distinctiveness}}=-1.67$, $t(79)=-0.24$, $p=0.81$, $\beta_{\text{age}}=4.00$, $t(79)=7.35$, $p<0.001$).

Distinctiveness for adult faces over the selective voxels of right VTC also predicts face recognition performance in left out subjects (LOOCV, Fig. 4D-purple bar). Face recognition ability is predicted from face distinctiveness over the selective voxels of right lateral VTC in the left-out subject with a median prediction error of 10.93%. Again, adding age to the model did not significantly reduce the error ($t(28)=-0.42$, $p=0.68$, Fig. 4D-dark gray bar) and for most subjects, the prediction error was higher from a model based on the non-selective voxels than a model based on the union of selective voxels (Fig. 4D-light gray bar and swarm plot, t-test comparing the difference in error to zero: $t(28)=-3.19$, $p=0.004$). Together these data suggest distinctiveness of word and face representations in lateral VTC predict performance for these categories, and that developmental improvements in reading and face recognition abilities are linked to increases in word and face distinctiveness in the selective voxels of left and right lateral VTC, respectively.

Discussion

By combining longitudinal measurements of distributed VTC responses to multiple categories with behavioral data in children over several years, the present study reveals several novel findings regarding the functional development of category representations in children's brains. First, our data reveals development of multiple category representations in lateral VTC from age 5 to age 17. This development consists of both enhancement and diminution of the distinctiveness of distributed category representations, as well as development of representations for categories that are not associated with clustered regions in this cortical expanse. Second, the development of distributed responses was driven mainly by changes in representations over the union of voxels that exhibited category selectivity. Third, importantly, the developmental increases of category distinctiveness to words and faces in the union of selective voxels in the

dominant hemisphere (left and right, respectively) predicted visual recognition ability of words and faces in left-out data.

Our longitudinal and multivariate analyses of category distinctiveness in lateral VTC reveal increases of distinctiveness for faces, words, numbers, and houses as well as a decrease of distinctiveness for limbs. The increase in distinctiveness for faces and words in left lateral VTC is in line with prior findings showing that word- and face-selective regions grow with development and become increasingly more selective to their preferred category^{16,19,21,23,25,32} and with results from multivariate approaches showing increases in distinctiveness for words^{16,18}. Further, the decrease in distinctiveness in distributed lateral VTC responses for limbs is consistent with findings that limb-selective regions shrink from childhood to the teens and lose their limb-selectivity²³. Together this provides evidence that developmental changes in the degree of selectivity affect distributed responses. That is, developmental increases in selectivity lead to increases in category distinctiveness and developmental decreases in selectivity lead to decreases in distinctiveness.

Our data also reveal increases in distinctiveness for numbers and houses with age in the right lateral VTC. The former finding is in line with the expectation that schooling not only enhances word but also number representations in VTC. As prior cross-sectional¹⁸ and univariate²³ approaches did not detect development of number responses in lateral VTC, these data highlight the higher sensitivity of longitudinal, multivariate approaches to measure functional development in the brain. The latter finding of developmental increases in distinctiveness for houses in lateral VTC was unexpected as place-selective regions are located in the medial part of VTC^{10,12}. As both voxels with strong positive and a negative preference contribute to category information (e.g., words¹⁸, the increase in distinctiveness for houses may be driven both by increases in the number of house-selective voxels in lateral VTC²³, as well as increases in the number of voxels with a negative preference to houses, like face-selective voxels.

In addition to the changes in category distinctiveness, we also examined changes in the nature of the representational space of the 10 categories across lateral VTC. Like prior

cross-sectional studies^{15,17} we find that representations of faces are already separated from those of other categories at age 5 (Fig. 3) and yet our data show that they continue to be enhanced into adolescence. Our results also show that representations for words and numbers in the left hemisphere are not yet separated from those of other categories in the younger age group of 5-9-year-olds, but become enhanced by adolescence. As such, these data suggest (i) that the development of distributed face representations starts earlier than that of character representations, which appears to start later in childhood and (ii) that schooling and learning how to read changes category representational space in VTC. This idea is consistent with findings that extensive experience playing Pokémon during childhood also affects category representations in VTC³³.

In addition to the enhancement of character and face representations during childhood, our data also reveal that representations for limbs weakened from childhood to adolescence in the right hemisphere. That is, representations for limbs move towards the center of the MDS space as children get older (Fig. 3D). While the reason for this developmental change is currently unclear, it is possible that changes in visual experience with limbs during childhood contribute to this change. Future longitudinal studies can test this prediction empirically. These findings also raise questions on how the representational space for visual categories may be affected in individuals with altered or deprived visual experience such as children with cataracts^{34,35}, illiterate individuals, or late-literate individuals who only learned how to read when they were adults³⁶.

The present results also have important theoretical implications for the debate on how distributed¹³ vs. modular²⁸ category representations in VTC contribute to behavior. On the one hand, our data reveal that representations of animate (faces and bodies) categories are separable from representations of the other inanimate categories in both selective and non-selective voxels (Fig. 3), consistent with the predictions of the distributed hypothesis¹³. On the other hand, we find significant development of category distinctiveness across the union of the selective voxels in lateral VTC, but no evidence for development of category distinctiveness in the non-selective voxels, contrary to the

predictions of the distributed hypothesis¹³. Crucially, however, we find that distinctiveness over the union of selective voxels, but not the non-selective voxels, predicts recognition behavior. That is, distinctiveness for pseudowords over the union of selective voxels in left lateral VTC predicted pseudoword reading; likewise, distinctiveness for faces over the union of selective voxels in right lateral VTC predicted face recognition. These longitudinal measurements are consistent with our prior cross-sectional data that found that better reading ability in adults than children is coupled with higher distinctiveness in word-selective voxels¹⁸. Together, the present data support a sparsely distributed account of VTC functional organization (Weiner and Grill-Spector 2010). That is, while we find that distributed responses across both selective and non-selective voxels contain category information, development of the selective voxels is more enhanced and better predicts behavioral outcomes.

The described link between category distinctiveness in the union of selective voxels and recognition ability has important ramifications for studying the neural basis of atypical development and developmental disorders. For instance, future research can examine if children with dyslexia have lower distinctiveness for words in left lateral VTC than age-matched typically-developing children and whether distinctiveness for words contributes to explaining reading ability independent of other predictors such as phonological awareness³⁷, socio-economic status^{38,39}, white matter properties⁴⁰, gyration in auditory cortex⁴¹, or perceptual decision-making⁴². Similarly, future research can test if adults and children with developmental prosopagnosia^{43,44} show lower distinctiveness for faces, compared to typical age-matched controls, and conversely if super-face-recognizers^{45,46}, show higher distinctiveness for faces in right lateral VTC than typical age-matched controls.

In sum, our results not only provide novel insights on the development of category representations in VTC, but also elucidate how this development relates to behavioral changes in visual recognition of faces and written words during childhood and adolescence. Future studies can test the hypothesis that this link between category

distinctiveness and behavior generalizes to the representation of other categories and skills involving visual recognition, such as number representations and math.

Methods

Participants

Children aged 5-12 years with normal or corrected-to-normal vision were recruited for this study. This age range was selected for two reasons: First, face recognition and reading, the two behavioral measures assessed in this study, improve during this age range. Second, prior studies investigating the functional development of VTC have shown development in this age range^{16,23}.

Children were recruited from schools in and around Palo Alto, Ca. The diversity of the participants reflects the makeup of the region: 62.5% of children were Caucasian, 20% were Asian, 5% were Native Hawaiian, 5% were Hispanic, and 7.5% were multiracial or from other racial/ethnic groups. We collected fMRI data from 40 (26 female) children (onset age=5-12 years, $M=8.66$ years, $SD=2.34$ years). Data from 4 children had to be excluded because they dropped out of the study after participating only once, thus providing no longitudinal fMRI data. Data from 7 children were excluded because their data did not pass inclusion criteria (see below, Method details). In the remaining 29 children, 29 functional sessions were excluded due to motion, 1 session due to a technical error during acquisition, and 1 session due to aliasing artifacts during acquisition. The fMRI data has previously been reported in²³.

Therefore, in this study we report data of 29 neurotypical children (18 female, 11 male), who were between 5 to 12 years old ($mean=9.19$, $SD=2.13$) when they enrolled in the study. Children participated in the study for 1 to 5 years. When possible, children completed 1 to 2 functional scans and a structural scan each year. Each child participated in at least 2 and up to 10 fMRI sessions ($mean=4.41$, $SD=1.92$) with the time interval between the first and last fMRI scan ranging from 10 months to 5 years ($mean=45$ months, $SD=18$ months). We also report behavioral data collected on a subset of sessions of the same children (see below, behavioral data collection).

Statement on ethical regulations

This study was approved by the Institutional Review Board of Stanford University. Prior to the start of the study, parents gave written consent, and children gave written assent for their participation. Children received \$30 per hour for scanning, \$10 per hour for behavior, and a small toy for their participation.

Procedure

Before taking part in the actual fMRI session, children completed a training with a mock scanner to enhance the quality of pediatric neuroimaging data. During the mock scanner training children were acclimated to the scanner environment in a child-friendly way: Children practiced laying still while watching a short movie and receiving live feedback on how much they were moving. After the mock scanner training the child participated in the actual MRI session. Functional and anatomical scans were typically conducted on different days to avoid fatigue. Face recognition and reading tests were typically completed after one of the scanning sessions.

Magnetic resonance imaging

Structural imaging

We collected neuroimaging data at the Center for Cognitive Neurobiological Imaging at Stanford University using a phase-array 32 channel head coil and a 3 Tesla GE Discovery MR750 scanner (GE Medical Systems). Anatomical scans were collected using quantitative MRI (qMRI⁴⁷). Here, we used a spoiled gradient echo sequence with multiple flip angles ($\alpha=4^\circ, 10^\circ, 20^\circ, 30^\circ$), TR=14ms and TE=2.4ms. The scan resolution was 0.8x0.8x1.0mm³ (which was later resampled to 1mm isotropic). For T1-calibration spin-echo inversion recovery scans were acquired with an echo-planar imaging read-out, spectral spatial fat suppression and a slab inversion pulse. These scans were collected at TR=3s, inplane resolution=2mmx2mm, slice thickness=4mm and 2x acceleration, echo time=minimum full.

Functional imaging

The functional data was acquired with the same scanner and head coil as the structural images. We oriented slices parallel to the parieto-occipital sulcus. Data were collected using a simultaneous multi-slice, one-shot T2* sensitive gradient echo EPI sequence with a multiplexing factor of 3. This sequence had a FOV=192mm, TR=1s, TE=30ms, and flip angle=76°, a resolution of 2.4 mm isotropic and near whole brain coverage (48 slices).

10 category experiment

During functional imaging, children completed three runs of a 10-category experiment ^{23,48}. In each functional run, participants watched images from 10 categories which can be grouped into 5 domains (Supplemental Fig. 1B): faces (child faces, adult faces), characters (pseudowords, numbers), body parts (headless bodies, limbs), places (houses, corridors) and objects (cars, string instruments). Images contained category stimuli, which were placed on a phase-scrambled background generated from randomly selected images. All stimuli were grayscale. Images were presented at a rate of 2 Hz and did not repeat across the course of the experiment. Images were presented in 4 s blocks, which were intermixed with baseline blocks showing gray luminance screen. Blocks were counterbalanced across categories and baseline blocks. Participants were instructed to view the images while fixating on a central dot and perform an oddball task. The oddball task required participants to press a button whenever an image comprising only the phase-scrambled background appeared. Behavioral responses of the oddball task were recorded in 98 out of 128 sessions due to occasional button malfunction. Performance on the oddball task performed during scanning was overall high (median performance=91%, SD=18%).

Behavioral data collection

Assessing reading ability

In a subset of sessions participants also completed the word identification and word attack tests from the Woodcock Reading Mastery Test (WRMT). Reading assessment was performed outside the scanner. In the word identification task, participants are asked to read a list of words as accurately as possible. In the word attack task, participants are

supposed 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 read the complete list. The reading score for each test was obtained by dividing the number of words read correctly by the total number of words and multiplying the result by 100.

Assessing face recognition ability

In a subset of sessions participants also completed the Cambridge Face Memory Test (CFMT⁴⁹) using adult male faces and a version of the CFMT using male child faces⁵⁰. The CFMT is a self-paced face recognition test. In the learning phases, participants learn the identity of six target unfamiliar faces. In the test phase, in each trial they are shown a triplet of faces, and are asked to identify the learned faces amongst distractor faces. The test constitutes 72 trials that become increasingly difficult as faces appear in unknown views, lighting, and with superimposed noise. Accuracy was measured as the percent correct responses made during the test phase.

Data analysis

We used MATLAB version 2017b (The MathWorks, Inc.) and the mrVista software package (<https://github.com/vistalab/vistasoft/wiki/mrVista>) to analyze the data. Swarm plots in Fig. 4 were created using MATLAB version 2020b.

Inclusion criteria

There were two criteria relevant for inclusion of the fMRI data into the analysis. First, there needed to be at least 2 (out of 3) runs per session with within-run motion < 2 voxels and between-run motion < 3 voxels. Second, the child participated in at least two fMRI sessions that were at least 6 months apart. Because in several sessions only 2 out of 3 functional runs passed the motion quality criteria, final analyses include 2 runs per session to ensure equal amounts of data across participants and functional sessions. For sessions with 3 runs surviving motion quality thresholds, the 2 runs with lowest within-run motion were selected.

For analyses that relate fMRI data to behavioral data, behavioral datasets (face recognition and reading tests) were included in the analysis if the time between

acquisition of the behavioral data and the acquisition of fMRI data was < 1 year (Fig. S1A).

Analysis of structural MRI data and individual template creation

We processed quantitative whole brain images of each child and timepoint with the mrQ pipeline (<https://github.com/mezera/mrQ>⁴⁷) to generate synthetic T1 brain volumes. For each child, we then used the synthetic T1 brain volumes from their multiple timepoints to generate a within-subject brain volume template. The individual template of each child was generated using the FreeSurfer Longitudinal pipeline implemented in FreeSurfer version 6.0. (<https://freesurfer.net/fswiki/LongitudinalProcessing>⁵¹). We then manually edited the gray-white matter segmentations of each participant's within-subject brain template to fix segmentation errors (like holes and handles) to generate an accurate cortical surface reconstruction. The functional data from each child's multiple timepoints (see below) was then aligned to the within-subject template. The reasons for this procedure were to: (i) minimize potential biases which can occur from aligning longitudinal data to the anatomical volume from a single timepoint⁵¹ and to (ii) enable using the same anatomical regions of interest (ROIs) across different timepoints in the same brain volume for each participant. On average 2.48 (SD=0.69) synthetic T1s were used to generate the within-subject-template (min=2, max=5). In 17 participants the last fMRI session that was included was conducted after the within-subject template had been created. These functional sessions were acquired on average 11 ± 2 months after acquisition of the last synthetic T1 that was included in the within-subject-template (excluding 2 participants whose last synthetic T1 could not be used because of technical error during acquisition and subject motion).

Definition of lateral VTC Regions of Interest (ROIs)

We defined lateral VTC ROIs based on anatomical landmarks on the inflated cortical surface in each hemisphere of each participant as in prior publications^{18,23}. The posterior border of the lateral VTC ROI was defined along the posterior transverse collateral sulcus (ptCoS). The anterior border was aligned to the posterior end of the hippocampus, which typically aligns with the anterior tip of the mid fusiform sulcus (MFS). The medial border

of the ROI was placed along the mid-fusiform sulcus (MFS) and the lateral border of the ROI was along the inferior temporal gyrus (ITG).

Analysis of functional MRI data

Functional data from each session were aligned to the individual subject brain template (see above). Data was motion corrected both within and across functional runs. We applied no spatial smoothing and no slice-timing correction. Time courses were transformed into percentage signal change. To this end, each timepoint of each voxel's data was divided by the average response across the entire run. A general linear model (GLM) was fit to each voxel by convolving the stimulus presentation design with the hemodynamic response function (as implemented in SPM, <https://www.fil.ion.ucl.ac.uk/spm/>) to estimate the contribution of each of the 10 conditions.

Multivariate pattern analysis

For each category multivoxel patterns (MVPs) of responses are vectors of response amplitudes across all voxels in each ROI (left and right lateral VTC). Amplitudes were estimated from the GLM run at each voxel. Response amplitudes were transformed into z-scores to remove between voxels differences in amplitudes (e.g., because of distance from coil) and to down-weight noisy voxels (see^{18,52,53}). Next, we calculated all pair-wise correlations between MVP pairs from one functional run to the other resulting in a 10 x 10 representational similarity matrix (RSM¹⁴) for each session.

Assessment of category distinctiveness

For each session's RSM, we computed the distinctiveness for each of the 10 categories. The distinctiveness of a category is defined as the within-category minus between-category similarity of distributed responses leaving out the between-category similarity with the other category from the same domain. This is illustrated in the gray box in Fig. 1A: In the RSM the values on the diagonal represent the within-category similarity, while the off-diagonal values represent the between-category similarities. For instance, to compute the distinctiveness for words, the between-category similarities for all categories except numbers are subtracted from the within category similarity for words. The reason

for leaving out the other category from the same domain is that for some domains the two categories are very similar to each other (such as adult and child faces) while for other domains the two categories differ from each other to a greater extent (such as cars and string instruments) so we chose a procedure that would be least biased and will not differently affect stimuli from various domains.

Definition of voxel subsets: the union of selective voxels across all categories and of non-selective voxels

To define the union of selective voxels in each session, we first computed the selective voxels for each category and then took the union of selective voxels across all 10 categories. Selective voxels for each category were defined by contrasting responses to a given category vs. all other categories except the other category from the same domain (i.e., words vs. all other categories except numbers). Category-selectivity was defined as a t-value > 3 (voxel-level) for the contrast of interest. Then, we took the union of the selective voxels across all 10 categories. Non-selective voxels were the remainder of voxels in lateral VTC, that were not selective of any of the 10 categories.

Visualizing changes in the representational space

We used multidimensional scaling (MDS) to visualize the representational space in lateral VTC and determine how it changes with age (Fig. 3 A,B,D, E & Supplemental movie 1). As visualizing how the representational space changes longitudinally is challenging, we show MDS embeddings for young children (5-9-year-olds) and teenagers (13-17-year-olds). Supplemental movie 1 is a video that shows the developmental trajectory of the MDS embedding across age. These age groups are used for visualization only; statistics are run on the whole longitudinal sample (see below). For the MDS embedding shown in Fig. 3 A,B,D,E we first computed mean RSMs for 5-9-year-olds and 13-17-year-olds separately for the union of selective voxels and for the non-selective voxels in each hemisphere (Fig. S2B,C). The RSMs were then turned into dissimilarity matrices. We next applied classical multidimensional scaling (MDS) to visualize the representational space in 2D (Fig. 3 A,B,D,E). To facilitate visual comparison of different MDS embeddings across hemispheres and sets of voxels, we aligned all MDS embeddings of 13-17-year-

olds to the embedding of 13-17-year-olds across all voxels in left lateral VTC using the Procrustes transformation. We next used the Procrustes transformation to align the embedding of 5-9-year-olds to that of 13-17-year-olds. Procrustes analysis was used without scaling in both cases.

Assessing changes in the representational space

The line plots in Fig. 3C,F quantify the change that is visualized in the MDS embeddings (Fig. 3A,B,D,E). The goal of this analysis was to assess how the representational structure changes in the union of selective voxels compared to in the non-selective voxels in each child. To this end, we applied MDS to the RSM of each child's first session in the study and to the RSM of their last session. We next used Procrustes transformation without scaling to align the embedding of the first session to that of the last session. In each child we then measured the Euclidian distance for each category between the first to the last session, in the shared MDS embedding space and report the mean distance across all categories. This analysis was performed for both hemispheres and for both sets of voxels (union of selective voxels, non-selective voxels).

Statistics

Unless stated otherwise, tests reported in this manuscript are two-tailed.

We used linear mixed models (LMM) to test whether distinctiveness develops with age because LMMs can account both for the hierarchical data structure with multiple sessions being nested within each participant and for the uneven distribution of sessions across time (Supplemental Fig. 1A).

Statistical analyses related to Fig. 1

LMMs were fitted using the 'fitlme' function in MATLAB version 2017b (The MathWorks, Inc.). In these models, category distinctiveness was predicted by age using participant as a random factor. We first tested whether a random intercept or random slopes model fit the data best. In a random intercept model, intercepts are allowed to vary across participants, while in a random slope model both intercepts and slopes can vary across

participants. Since in the majority of cases random intercept models fit the data best, we used random intercepts for all analyses presented in Fig. 1 to enable comparability across models.

We next tested whether motion during scanning and tSNR contributed to the model fit. Adding motion as a predictor to the LMM did not significantly contribute to the model fit except for distinctiveness for string instruments. Importantly, adding motion in the model for distinctiveness for string instruments did not influence the result: There was no significant contribution of age to string instruments distinctiveness with or without adding motion. We next tested whether tSNR contributed to the model fit. Since adding tSNR contributed to the model fit independently from age in several cases, we included tSNR as an additional factor into the model. The bars in Fig. 1D show the slopes for the effect of age on category distinctiveness taking into account tSNR. LMMs related to Figure 1D can be expressed as: category distinctiveness \sim age in years + tSNR + (1|participant), in which category distinctiveness is the response variable, age and tSNR are the predictors and the term (1|participant) indicates that intercepts vary by participant. False-discovery rate (FDR) correction following the procedure by Benjamini and Hochberg⁵⁴ as implemented in MATLAB version 2017b (The MathWorks, Inc.) was applied to account for multiple comparisons for analyses shown in Fig. 1D.

Statistical analyses related to Fig. 2

The analyses in Fig. 2 were performed similar to those in Fig.1D and can be expressed as: category distinctiveness \sim age in years + tSNR + (1|participant). LMMs were run separately for each category, hemisphere, and voxel subsets (union of the selective voxels, non-selective voxels). Here, tSNR was obtained for each voxel and then averaged across voxels within each voxel subset (union of the selective voxels, non-selective voxels). In addition, we ran a LMM to test whether there are differences in tSNR in the subset of the union of the selective voxels and the non-selective voxels. This model can be expressed as: tSNR \sim voxelSubset + (1|participant), where voxelSubset is a binary predictor (selective/non-selective). For analyses shown in Fig. 2 FDR correction was performed as described for Fig. 1.

Statistical analyses related to Fig. 3

We used paired t-tests to evaluate the change in representation in the union of selective voxels and the non-selective voxels in each hemisphere (Fig. 3C,F). That is, in each hemisphere and voxel subset we first computed the mean Euclidian distance across all categories from each child's MDS embedding of their first session to the embedding of their last session. We then compared the mean distances across the union of selective voxels to that in the non-selective voxels.

Statistical analyses related to Fig. 4

For the analysis in Fig. 4A,C, we used LMMs to test if distinctiveness in the union of selective voxels is related to behavior.

Fig. 4A: While we acquired reading scores for both real words and pseudowords in our participants, we used the reading scores for pseudowords to test for a link between distinctiveness and behavior, because pseudowords were also used in the fMRI experiment. We used a random slopes model that can be specified as: Pseudoword reading score (%) ~ distinctiveness for pseudowords + (distinctiveness for pseudowords | participant). A random slopes model was used because (i) it fitted the data better compared to the random intercept model and (ii) the individual subject's slopes (random effects) visualize that there is a positive link between distinctiveness for pseudowords and reading score in most individuals.

Fig. 4C: In this analysis we tested if there is a link between distinctiveness for adult faces in the union of selective voxels in lateral VTC and face recognition scores as assessed with the CFMT using adult faces. While we had also acquired data of a version of the CFMT using child faces (CFMT-child), we used data of the CFMT-adults to link it to distinctiveness for two reasons. First, performance on the CFMT-child was overall higher resulting in ceiling effects in some subjects. Second, as the CFMT-adults is a widely used test it enables comparison across studies. As such, for the data presented in Fig. 4C we used a random slopes model that can be specified as: Face recognition score in CFMT-

adults (%) \sim distinctiveness for adult faces + (distinctiveness for adult faces | participant). While the random slopes model did not fit the data significantly better than the random intercept model, it enables visualizing the effect between face distinctiveness and face recognition in individual subjects. There are no significant differences to the results of the present analysis when a random intercept model is used instead of the random slopes model.

Predicting behavior from brain data

We used a leave-one participant-out-cross validation (LOOCV) approach to test if we can predict behavior from brain data. That is, we computed a LMM that predicts behavior using distinctiveness on all sessions except for one subject that was left out in each iteration. Then, we used the LMM estimates to compute the predicted behavioral score for each session of the left-out subject. The prediction error was defined as the difference between the predicted and the actual behavioral score. We then repeated this procedure for all subjects.

The bar plots in Fig. 4 B,D show the prediction error for three different models: (i) Selective: LMM predicting behavior from distinctiveness over the union of selective voxels (B: Pseudoword reading score \sim distinctiveness for pseudowords across the union of selective voxels + (distinctiveness for pseudowords | participant); D: Face recognition score across the union of selective voxels \sim distinctiveness for adult faces + (distinctiveness for adult faces | participant)). (ii) Non-selective voxels: LMM predicting behavior from distinctiveness over the non-selective voxels. (B: Pseudoword reading score \sim distinctiveness for pseudowords across the non-selective voxels + (distinctiveness for pseudowords | participant); D: Face recognition score \sim distinctiveness for adult faces across the non-selective voxels + (distinctiveness for adult faces | participant)). (iii) Selective & age: LMM predicting behavior from distinctiveness over the union of selective voxels with age as an additional factor (B: Pseudoword reading score \sim distinctiveness for pseudowords across the union of selective voxels + age + (distinctiveness for pseudowords | participant); D: Face recognition score \sim distinctiveness for adult faces across the union of selective voxels + age + (distinctiveness for adult faces | participant)).

In the analyses in the scatterplots in Fig. 4 B,D we calculated the difference in prediction error using a model that predicts behavior based on distinctiveness in the union of the selective vs. in the non-selective voxels. Next, we used one-sample t-tests to evaluate if the difference was significantly different from zero.

Acknowledgements

We thank Laura Villalobos, Erica Yeawon Hwang, Savana Huskins, Alema Fitisemanu, and Philip Eykamp, Cat Davis and Tai Anthony McMillan for manually editing gray-white matter brain segmentations. We thank Caitlyn Estrada and Nancy Lopez-Alvarez for help with data collection, and Rachel Hinds for help with data entry and management.

Funding: Fellowship of the German National Academic Foundation NO 1448/1-1 (MN); NIH grant 2RO1 EY 022318 (KGS); NIH training grant 5T32EY020485 (VSN); NSF Graduate Research Development Program DGE-114747 (JG); Ruth L. Kirschstein National Research Service Award F31EY027201 (JG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

M.N. collected data, developed, and coded the analysis pipeline, analyzed the data, and wrote the manuscript. V.S.N. and J.G. designed the experiment, collected data, and contributed to the manuscript. A.A.R., D.F. and H.K, collected the data, and contributed to the manuscript. K.G.S. designed the experiment, contributed to the analysis pipeline and data analyses, and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Materials & Correspondence

Correspondence and material requests should be addressed to the corresponding author M.N.

Data availability

Original code and data required to generate the main figures will be made available with publication at: <https://github.com/VPNL/distributedVTCDevelopment> (access with author permission). Due to the large size of the raw data, it will be made available from the corresponding author upon request.

References

1. Fausey, C. M., Jayaraman, S. & Smith, L. B. From faces to hands: Changing visual input in the first two years. *Cognition* **152**, 101–107 (2016).
2. Carey, S., Diamond, R. & Woods, B. Development of face recognition - a maturational component? *Dev. Psychol.* **16**, 257–269 (1980).
3. Germine, L. T., Duchaine, B. & Nakayama, K. Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition* **118**, 201–210 (2011).
4. Gaillard, R. *et al.* Direct Intracranial, fMRI, and Lesion Evidence for the Causal Role of Left Inferotemporal Cortex in Reading. *Neuron* **50**, 191–204 (2006).
5. Mégevand, P. *et al.* Seeing scenes: Topographic visual hallucinations evoked by direct electrical stimulation of the parahippocampal place area. *J. Neurosci.* **34**, 5399–5405 (2014).
6. Parvizi, J. *et al.* Electrical stimulation of human fusiform face-selective regions distorts face perception. *J. Neurosci. Off. J. Soc. Neurosci.* **32**, 14915–14920 (2012).
7. Kanwisher, N., McDermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* **17**, 4302–4311 (1997).
8. Peelen, M. V. & Downing, P. E. Selectivity for the human body in the fusiform gyrus. *J. Neurophysiol.* **93**, 603–608 (2005).
9. Schwarzlose, R. F. Separate Face and Body Selectivity on the Fusiform Gyrus. *J. Neurosci.* **25**, 11055–11059 (2005).
10. Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment. *Nature* **392**, 598–601 (1998).

11. Dehaene, S., Le Clec'H, G., Poline, J.-B., Le Bihan, D. & Cohen, L. The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport* **13**, 321–325 (2002).
12. Grill-Spector, K. & Weiner, K. S. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* **15**, 536–548 (2014).
13. Haxby, J. V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
14. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 1–28 (2008).
15. Cohen, M. A. *et al.* Representational similarity precedes category selectivity in the developing ventral visual pathway. *NeuroImage* **197**, 565–574 (2019).
16. Dehaene-Lambertz, G., Monzalvo, K. & Dehaene, S. The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLoS Biol.* **16**, 1–34 (2018).
17. Golarai, G., Liberman, A. & Grill-Spector, K. Experience Shapes the Development of Neural Substrates of Face Processing in Human Ventral Temporal Cortex. *Cereb. Cortex* **bhv314** (2017) doi:10.1093/cercor/bhv314.
18. Nordt, M. *et al.* Learning to Read Increases the Informativeness of Distributed Ventral Temporal Responses. *Cereb. Cortex* **1–16** (2019) doi:10.1093/cercor/bhy178.
19. Ben-Shachar, M., Dougherty, R. F., Deutsch, G. K. & Wandell, B. A. The Development of Cortical Sensitivity to Visual Word Forms. *J. Cogn. Neurosci.* **23**, 2387–2399 (2011).

20. Cantlon, J. F., Pinel, P., Dehaene, S. & Pelpfrey, K. A. Cortical representations of symbols, objects, and faces are pruned back during early childhood. *Cereb. Cortex* **21**, 191–199 (2011).
21. Golarai, G. *et al.* Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nat. Neurosci.* **10**, 512–522 (2007).
22. Golarai, G., Liberman, A., Yoon, J. M. & Grill-Spector, K. Differential development of the ventral visual cortex extends through adolescence. *Front. Hum. Neurosci.* **3**, 80 (2010).
23. Nordt, M. *et al.* Cortical recycling in high-level visual cortex during childhood development. *Nat. Hum. Behav.* (2021) doi:10.1038/s41562-021-01141-5.
24. Peelen, M. V., Glaser, B., Vuilleumier, P. & Eliez, S. Differential development of selectivity for faces and bodies in the fusiform gyrus. *Dev. Sci.* **12**, 16–25 (2009).
25. Scherf, K. S., Behrmann, M., Humphreys, K. & Luna, B. Visual category-selectivity for faces, places and objects emerges along different developmental trajectories. *Dev. Sci.* **10**, (2007).
26. Haxby, J. V., Connolly, A. C. & Guntupalli, J. S. Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annu. Rev. Neurosci.* **37**, 435–456 (2014).
27. Kanwisher, N. Domain specificity in face perception. *Nat. Neurosci.* **3**, 759–763 (2000).
28. Kanwisher, N. Functional specificity in the human brain: A window into the functional architecture of the mind. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 11163–11170 (2010).
29. Weiner, K. S. & Grill-Spector, K. Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage* **52**, 1559–1573 (2010).

30. Weiner, K. S. *et al.* The mid-fusiform sulcus: A landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *Neuroimage* **84**, 453–465 (2014).
31. Gomez, J., Natu, V., Jeska, B., Barnett, M. & Grill-Spector, K. Development differentially sculpts receptive fields across early and high-level human visual cortex. *Nat. Commun.* **9**, 1–12 (2018).
32. Brem, S. *et al.* Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proc. Natl. Acad. Sci.* **107**, 7939–7944 (2010).
33. Gomez, J., Barnett, M. & Grill-Spector, K. Extensive childhood experience with Pokémon suggests eccentricity drives organization of visual cortex. *Nat. Hum. Behav.* **3**, 611–624 (2019).
34. Le Grand, R., Mondloch, C., Maurer, D. & Brent, H. Early visual experience and face processing. *Nature* **410**, 890 (2001).
35. Röder, B., Ley, P., Shenoy, B. H., Kekunnaya, R. & Bottari, D. Sensitive periods for the functional specialization of the neural system for human face processing. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 16760–16765 (2013).
36. Dehaene, S. *et al.* How learning to read changes the cortical networks for vision and language. *science* **330**, 1359–1364 (2010).
37. Bradley, L. & Bryant, P. E. Categorizing sounds and learning to read - a causal connection. *Nature* **301**, 419–421 (1983).

38. Noble, K. G., Wolmetz, M. E., Ochs, L. G., Farah, M. J. & McCandliss, B. D. *Brain-behavior relationships in reading acquisition are modulated by socioeconomic factors*. *Developmental Science* vol. 9 642–654 (2006).

39. Zuk, J. *et al.* Multifactorial pathways facilitate resilience among kindergarteners at risk for dyslexia: A longitudinal behavioral and neuroimaging study. *Dev. Sci.* **24**, (2021).

40. Yeatman, J. D., Dougherty, R. F., Ben-Shachar, M. & Wandell, B. a. Development of white matter and reading skills. *Proc. Natl. Acad. Sci.* **109**, E3045–E3053 (2012).

41. Kuhl, U. *et al.* The emergence of dyslexia in the developing brain. *NeuroImage* **211**, (2020).

42. O'Brien, G. & Yeatman, J. D. Bridging sensory and language theories of dyslexia: Toward a multifactorial model. *Dev. Sci.* **24**, (2021).

43. Behrmann, M. & Avidan, G. Congenital prosopagnosia: face-blind from birth. *Trends Cogn. Sci.* **9**, 180–187 (2005).

44. Duchaine, B. C. & Nakayama, K. Developmental prosopagnosia: a window to content-specific face processing. *Curr. Opin. Neurobiol.* **16**, 166–173 (2006).

45. Ramon, M. Super-Recognizers – a novel diagnostic framework, 70 cases, and guidelines for future work. *Neuropsychologia* **158**, 107809 (2021).

46. Russell, R., Duchaine, B. & Nakayama, K. Super-recognizers: People with extraordinary face recognition ability. *Psychon. Bull. Rev.* **16**, 252–257 (2009).

47. Mezer, A. *et al.* Quantifying the local tissue volume and composition in individual brains with magnetic resonance imaging. *Nat. Med.* **19**, 1667–1672 (2013).

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

49. Duchaine, B. C. & Nakayama, K. The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia* **44**, 576–585 (2006).
50. Dalrymple, K. A., Gomez, J. & Duchaine, B. The Dartmouth Database of Children's Faces: Acquisition and Validation of a New Face Stimulus Set. *PLoS ONE* **8**, e79131 (2013).
51. Reuter, M., Schmansky, N. J., Rosas, H. D. & Fischl, B. Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage* **61**, 1402–1418 (2012).
52. Margalit, E. *et al.* Ultra-high-resolution fMRI of Human Ventral Temporal Cortex Reveals Differential Representation of Categories and Domains. *J. Neurosci.* **40**, 3008–3024 (2020).
53. Sayres, R. & Grill-Spector, K. Relating Retinotopic and Object-Selective Responses in Human Lateral Occipital Cortex. *J. Neurophysiol.* **100**, 249–267 (2008).
54. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–3300 (1995).