bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Both mOTS-words and pOTS-words prefer emoji stimuli over text stimuli during a reading task

Alexia Dalski[1,2], Holly Kular[3], Julia G. Jorgensen[3], Kalanit Grill-Spector[3,4], Mareike Grotheer[1,2]

[1] Department of Psychology, Philipps-Universität Marburg, Marburg 35039, Germany.
[2] Center for Mind, Brain and Behavior – CMBB, Philipps-Universität Marburg and Justus-Liebig-Universität Giessen, Marburg 35032, Germany.
[3] Department of Psychology, Stanford University, Stanford, CA 94305, USA.
[4] Wu Tsai Neurosciences Institute, Stanford University, CA 94305, USA.

Highlights

- OTS-words can be divided into two subregions
- The subregions can be identified by contrasting text with other stimuli
- pOTS-words prefers readable emojis over text independent of task
- mOTS-words prefers readable emojis during a reading task

Abstract

The visual word form area in the occipitotemporal sulcus (OTS), here referred to as OTS-words, responds more strongly to text than other visual stimuli and plays a critical role in reading. Here we hypothesized, that this region's preference for text may be driven by a preference for reading tasks, as in most prior fMRI studies only the text stimuli were readable. To test this, we performed three fMRI experiments (N=15) and systematically varied the participant's task and the visual stimulus, investigating mOTS-words and pOTS-words subregions. In experiment 1, we contrasted text stimuli with non-readable visual stimuli (faces, limbs, houses, and objects). In experiment 2, we used an fMRI adaptation paradigm, presenting the same or different compound words in text or emoji formats. In experiment 3, participants performed either a reading or a color task on compound words, presented in text or emoji format. Using experiment 1 data, we identified left mOTS-words and pOTS-words in all participants by contrasting text stimuli with non-readable stimuli. In experiment 2, pOTS-words, but not mOTS-words, showed fMRI adaptation for compound words in both text and emoji formats. In experiment 3, surprisingly, both mOTS-words and pOTS-words showed higher responses to compound words in emoji than text format. Moreover, mOTS-words, but not pOTS-words, also showed higher responses during the reading than the color task as well as an interaction between task and stimulus. Multivariate analyses showed that distributed responses in pOTS-words encode the visual stimulus, whereas distributed responses in mOTS-words encode both the stimulus and the task. Together, our findings suggest that the function of the OTS-words subregions goes beyond the specific visual processing of text and that these regions are flexibly recruited whenever semantic meaning needs to be assigned to visual input.

Keywords:
visual word form area, reading, occipito-temporal sulcus, category selectivity, vision

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction:

Reading texts is a fundamental building block of communication in our society and an inability to do so can have profound negative impacts on somebody's career, well-being, and socioeconomic status (Ardila et al., 2010; Cho et al., 2008; Weiss et al., 1991). However, unlike other common visual categories, such as faces and objects, the ability to assign meaning to written text needs to be acquired through formal education (Dehaene et al., 2015). This makes written text a unique visual category, and unraveling the neural underpinnings of reading acquisition can hence shed critical light on brain plasticity and organization (Dehaene et al., 2015; Dehaene & Cohen, 2011).

Previous studies have examined a region in the ventral-temporal cortex (VTC), which shows higher responses to text stimuli compared to other visual stimulus categories and is often referred to as the visual word form area (VWFA) (Cohen et al., 2000; Dehaene & Cohen, 2011; Jobard et al., 2003). Here we will follow an alternative naming convention, which takes into account the anatomical location and the functional preference of this brain region: As the VWFA is located in the occipitotemporal sulcus (OTS) and shows higher responses to text than other visual stimuli, it will be referred to as OTS-words. OTS-words is predominantly left-lateralized and is activated in literate participants during reading (Cohen et al., 2003; Dehaene et al., 2015). The causal role of the occipitotemporal sulcus (OTS) for reading was demonstrated in a lesion study by Gaillard et al. (2006), which showed that after the surgical removal of a word-sensitive patch in the OTS a patient developed a severe reading deficiency. Studies with dyslexic participants further support this notion, as subjects with reading disabilities show decreased activations along the occipitotemporal sulcus (Sandak et al., 2018; van der Mark et al., 2009).

Even though it is essential for reading, the emergence of OTS-words in the VTC is not fully understood. Considering that scripture is a relatively recent cultural invention, McCandliss et al. (2003) propose that the progressive specialization for words throughout reading acquisition follows the same principles as other forms of visual expertise (such as differentiating between different car brands or bird species). They argue that the visual cortex adapts to process salient stimulus groups, such as words, due to exposure and skill acquisition, not simply maturation (McCandliss et al., 2003). Findings contrasting response patterns of literate and illiterate subjects support this view, showing that the OTS is functionally reorganized when a person learns to read (Dehaene et al., 2015). However, there are different hypotheses regarding the details of this reorganization: The theory of functional refinement proposes that category selectivity develops on previously unspecialized cortex, which has advantageous properties for processing a certain category (Dehaene-Lambertz et al., 2018). For example, Szwed et al. (2009) suggest, that OTS-words develops on cortical theory, which shows an initial preference for line junctions and is therefore well equipped to process the shapes of letters. In contrast, the cortical recycling theory proposes that category selectivity emerges through competition for cortical resources, leading to neural recycling of the cortex when demands change, for example through reading acquisition (Dehaene et al., 2015). Some research suggests, that this rivalry may lead to face-sensitive cortex getting repurposed for the perception of words (Dehaene et al., 2010, 2015). In contrast, more recent longitudinal work indicated that neighboring limb selective cortex develops a sensitivity towards words when children learn to read (Nordt et al., 2021).

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A recent review paper, has further highlighted the importance of changes in visual diet throughout development, as during infancy limbs are more salient visual stimuli, while words become more important during literacy acquisition (Kubota, Grill-spector, et al., 2023).

Interestingly, several studies have proposed a posterior-to-anterior increase in functional processing level along the OTS, with simple features being processed in the more posterior parts and more complex information in the anterior ones (Caffarra et al., 2021; Carreiras et al., 2014; Vinckier et al., 2007). For example, there is evidence that the perception of letters happens in the more posterior part of the VTC, while perceiving entire words as language units takes place in the more anterior part (Thesen et al., 2012; Vinckier et al., 2007). In fact, recent studies have suggested that OTS-words can be further divided into two subregions, middle OTS-words and posterior OTS-words (Lerma-Usabiaga et al., 2018; Stigliani et al., 2015). Lerma-Usabiaga et al. (2018) further showed that these subregions differ in their functional properties; while pOTS-words is sensitive to the low-level visual features of text, mOTS-words is sensitive to the lexical content of the stimuli and prefers real words over pseudo-words. mOTS-words is hence proposed to be the region where vision and language are integrated. Accordingly, analyses of white matter connections revealed distinct connectivity patterns for pOTS and mOTS: While pOTS-words shows connections to the intraparietal sulcus via the vertical occipital fasciculus, mOTS-words is connected to the arcuate fasciculus and hence the language network (Kubota, Mareike, et al., 2023; Lerma-Usabiaga et al., 2018). A recent functional connectivity study further supports this notion, showing that pOTS-words responses are more strongly correlated with bilateral visual regions, while responses in mOTS-words are more strongly correlated with language regions, such as the inferior frontal gyrus (IFG) (Yablonski et al., 2023)

Importantly, the functional roles of the OTS-words subregions are not without debate. While the predominant view in the field suggests that these regions are selectively involved in processing text (Cohen et al., 2000a; Cohen & Dehaene, 2004; Szwed et al., 2011), several studies found surprising evidence against an exclusive preference for text stimuli in OTS-words: A study by Mei et al. (2010) demonstrated that OTS-words is involved in memory encoding of both text and faces, and does in fact not show higher activations when presented with text. Furthermore, OTS-words was shown to be activated when pseudowords, Amharic strings, and even line drawings are presented (Vogel et al., 2012). This is further supported by a study showing that while pOTS-words shows the highest responses to words, line drawings also elicit significant activations, leading to the conclusion that its shape extraction goes beyond text processing (Ben-Shachar et al., 2007). In a review paper, Vogel et al. (2014) hence conclude that OTS-words does not prefer written text over other visual stimuli and is not specifically used for reading, but rather has properties that are beneficial to multiple tasks, for example reading. Similarly, in a review by Price & Devlin (2003), authors concluded that OTS-words is involved in a number of tasks, that go beyond the visual processing of words, such as i) naming objects or colors (Moore & Price, 1999; C. J. Price & Friston, 1997), ii) hearing and answering yes/no questions about objects and animals (Thompson-Schill et al., 1999), and iii) Braille reading (C. Price et al., 1998; Reich et al., 2011). To explain these inconsistencies, Wright et al. (2008) conducted an fMRI study comparing the activations of OTS-words in response to pictures of objects and written text, using various statistical thresholds. Even though this study detected differential activations for pictures and text, these results did not withstand less liberal statistical thresholds, were not replicable between subjects, and were also not exclusive to OTS-words but found for other brain areas as well.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Here we hypothesized that this debate about the functional role of OTS-words may be resolved by evaluating the two sub-regions of OTS-words separately. After all, it is possible that one subregion is selective for text whereas the other subregion is not, which could explain the previous heterogeneous findings. Moreover, we postulate that it may be important to take into consideration not only the presented visual stimulus but also the task the participants are performing. After all, in many prior experiments, only the text stimuli were readable, so the preference for the visual features of text stimuli may have been confounded with a preference for reading tasks, which could have contributed to previous inconsistencies between studies. Indeed, a recent study illustrates the importance of the performed task in driving responses in OTS-words, as authors found activation in OTS-words when native speakers were reading Chinese characters but not when they performed a color task on the same stimuli (Qu et al., 2022). As such, we derived the following hypotheses from the current literature: H1: Both OTS-words subregions show higher responses to text stimuli compared to other visual stimulus categories, irrespective of the performed task (Cohen et al., 2000; Dehaene & Cohen, 2011) (Fig 1a). H2: Both OTS-words subregions show higher responses to reading tasks compared to other tasks, irrespective of the visual stimulus format (Fig 1b). H3: There is an anterior-to-posterior gradient in processing level between mOTS-words and pOTS-words, whereas the anterior mOTS-words is selective for reading tasks and the posterior pOTS-words is selective for the visual features of text (Caffarra et al., 2021; Lerma-Usabiaga et al., 2018; Taylor et al., 2019; Thesen et al., 2012; Vinckier et al., 2007a) (Fig 1c).

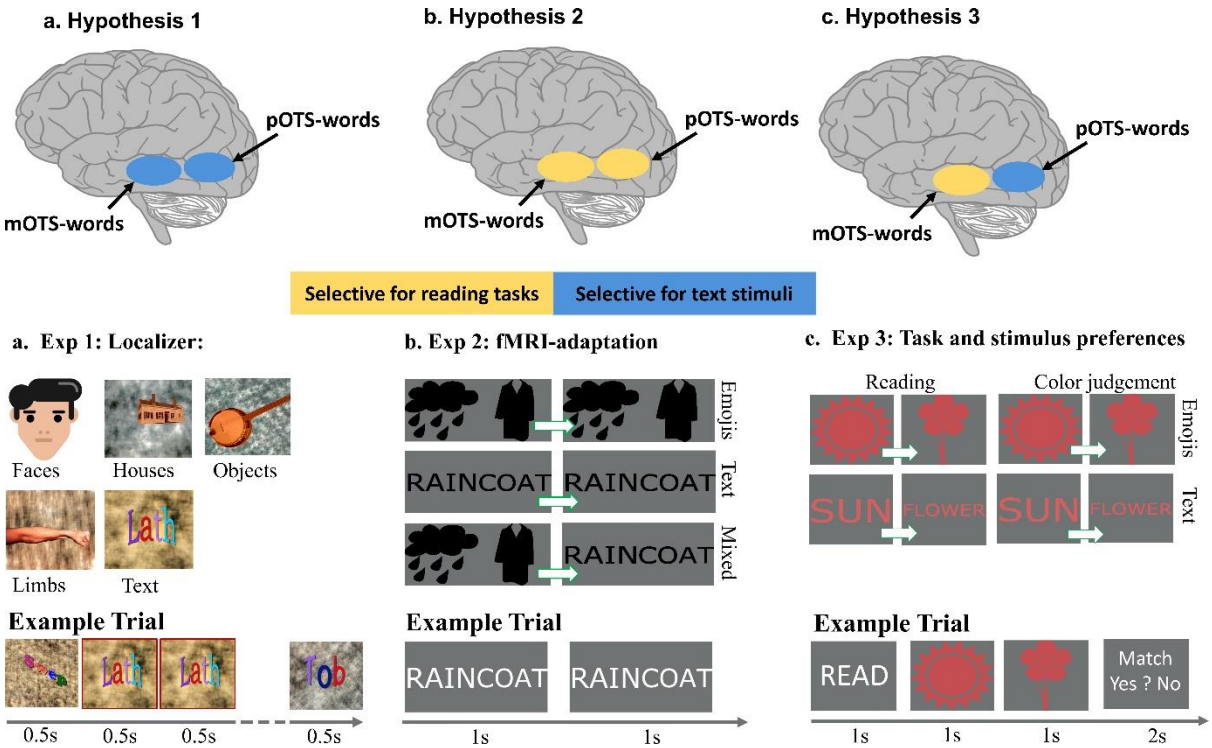

Figure 1. Summary of hypotheses and experimental design. a. Hypothesis 1. Both subregions of OTS-words show a preference for text stimuli. **b. Hypothesis 2.** Both subregions of OTS-words prefer reading tasks over other tasks. **c. Hypothesis 3.** pOTS-words is selective for the visual features of text and mOTS-words is selective for reading tasks. **d. Exp 1: Localizer.** Five visual categories (faces, houses, objects, limbs, text) were presented while subjects performed a 1-back task. Images were either colored or in grayscale. Faces were photographs of human faces; here we illustrate those with an icon. **e. Exp 2: fMRI-adaptation.** Two pairs of images were presented, whereas each pair formed an English compound word; Task: Participants had to indicate when the images were white rather than

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

black. All images in a trial were either in emoji, text, or mixed format. Pairs in a trial could be repeated or non-repeated. **f. Exp 3: Task and stimulus preferences.** Subjects were presented with two stimuli, presented one after the other, in emoji or text format. Tasks: Subjects were instructed to either read the word pair and indicate if they form a meaningful English compound word (reading task) or to compare the color hues of the two stimuli (color task).

To distinguish between these hypotheses, we conducted three fMRI experiments in which we systematically varied both the visual stimulus and the participant's task, and assessed mean and distributed responses in mOTS-words and pOTS-words. In experiment 1, we contrasted neural responses to text with other visual categories (limbs, objects, houses, faces, Fig 1d). We used experiment 1 data to identify mOTS-words and pOTS-words based on their preference for text, as in previous studies (Rosenke et al., 2021; Stigliani et al., 2015). In experiment 2, we used an fMRI adaptation paradigm to probe the stimulus sensitivity of mOTS-words and pOTS-words, comparing text and emoji stimuli (Fig 1e). Finally, in experiment 3, subjects performed a reading and a color judgment task on readable text and emoji stimuli (Fig 1f). Importantly, testing for differential stimulus and task preferences across the OTS-words subregions requires spatial precision, as the two subregions are closely neighboring in the brain (Lerma-Usabiaga et al., 2018). Accordingly, all analyses were conducted in the native brain space of each participant, without the application of spatial smoothing, to avoid creating erroneous overlap between the subregions (Weiner & Grill-Spector, 2013).

In the current study, we identified two subregions in the occipito-temporal sulcus, mOTS-words and pOTS-words based on their preference for text stimuli over other visual stimulus categories. Strikingly, even though these regions were identified through their sensitivity towards text, our univariate analyses showed higher responses for emoji stimuli than text stimuli in both subregions. In pOTS-words we further observed fMRI adaptation effects for both text and emoji stimuli, whereas mOTS-words showed a preference for a reading task over a color task. Multivariate analyses align with these results showing that pOTS-words encodes the visual stimulus, while mOTS-words encodes both the visual stimulus and the performed task. These results expand our understanding of the functional role of OTS-words, suggesting that the OTS-words subregions do not solely process text stimuli but might be involved in processing any kind of readable stimuli.

2. Materials and Methods

2.1 Participants

18 right-handed volunteers (11 female, 7 male, mean age 28.4 years, SD = 12.8 years) were recruited from Stanford University and surrounding areas and participated in two experimental sessions conducted on different days. Two participants had to be excluded due to excessive head motions, and one participant was excluded due to an incidental finding. The final data set contained 15 subjects (8 female, 7 male). Subjects had normal or corrected to normal vision and gave their informed written consent. The Stanford Internal Review Board on Human Subjects Research approved all procedures.

2.2 Data acquisition and preprocessing

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.2.1 Acquisition: Data was collected at the Center for Cognitive and Neurobiological Imaging at Stanford University, using a GE 3 tesla SIGNA Scanner with a 32-channel head coil. 48 slices were acquired, covering the occipitotemporal and most of the frontal cortex using a T2*-sensitive gradient echo sequence (resolution: 2.4 mm × 2.4 mm × 2.4 mm, TR: 1000 ms, TE: 30 ms, FoV: 192 mm, flip angle: 62°, multiplexing factor of 3). A whole-brain, anatomical volume was acquired as well, once for each participant, using a T1-weighted BRAVO pulse sequence (resolution: 1mm × 1 mm × 1 mm, TI=450 ms, flip angle: 12°, 1 NEX, FoV: 240 mm).

2.2.2 Preprocessing: The anatomical brain volume of each participant was segmented into grey and white matter using FreeSurfer (http://surfer.nmr.mgh.harvard.edu/). Manual corrections were made using ITKGray (http://web.stanford.edu/group/vista/cgi-bin/wiki/index.php/ItkGray), and each subject's cortical surface was reconstructed. Functional data was analyzed using the mrVista toolbox (http://github.com/vistalab) for Matlab. The fMRI data from each experiment was motion-corrected, both within and between runs, and then aligned to the anatomical volume. To increase precision, all analyses were conducted in the native brain space of each participant, and no spatial smoothing was applied. The time course of each voxel was high-pass filtered with a 1/20 Hz cutoff and transformed from arbitrary units to percentage signal change. For each experiment, we created a separate design matrix and convolved it with the hemodynamic response function (HRF) implemented in SPM (http://www.fil.ion.ucl.ac.uk/spm) to generate predictors for each experimental condition. Regularized response coefficients (betas) were estimated for each voxel and each predictor using a general linear model (GLM) indicating the magnitude of response for that condition.

2.3 Experimental Design

2.3.1 Experiment 1: Localizer

The first experiment was designed to detect voxels which show higher responses for text compared to other visual stimuli, as in previous work (Stigliani et al., 2015).

Stimuli: Stimuli were images belonging to one of five different categories: text, faces, limbs, objects, and houses. In order to also detect color-sensitive voxels, the stimuli were presented either in grayscale or in color.

Trial structure: All stimuli were shown for 500 ms, in blocks of 8 stimuli. Each participant completed 4 runs; each run was 5 minutes long and included 10 repetitions of each stimulus category.

Task: Participants were asked to fixate a black dot in the center of the screen and to indicate immediate repetitions of a stimulus by pressing a button (1-back task, Fig 1d). For further details on the experimental design see Stigliani et al. (2015).

2.3.2 Experiment 2: fMRI-adaptation

The goal of the second experiment was to leverage fMRI adaptation effects to probe stimulus preferences in the previously identified OTS-words and control regions, with high spatial accuracy.

Stimuli: Stimuli consisted of two pairs of images, presented at the center of the screen, whereas each pair formed an English compound word (e.g., raincoat). The compound words were presented in three different formats: as emojis, as text, or in a mixed format, in which one pair was an emoji and one pair was text. All stimuli were either white or black on a gray background.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Trial structure: In each trial, two image pairs were presented, which were either repeated (e.g., "raincoat" followed by "raincoat") or non-repeated (e.g., "raincoat" followed by "sunflower"). Each pair was presented for 1s (Fig 1e). Participants completed 6 runs; each run had a duration of ~5 minutes and contained 72 experiment trials. Trial conditions were balanced in regards to stimulus format and repetition within each run.
Task: The image pairs were presented either in black or white in ~50% of the trials. Subjects had to press a button whenever they saw a white image pair (black/white detection).

2.3.3 Experiment 3: Task and stimulus preferences
The goal of experiment 3 was to probe the influence of different tasks (reading task vs. color judgment task) and stimuli (emoji vs. text stimuli) on neural responses in the OTS-words subregions.
Stimuli: Participants were presented with two stimuli shown consecutively, each stimulus depicting an English word. The stimuli were either both in text or emoji format. Each stimulus pair either formed a meaningful English compound (sun + flower) or a meaningless term (flower + sun). All pairs were presented in color on a gray background. The pairs had either the exact same color or similar, yet distinguishable, hues (Fig 1f). Behavioral pilot experiments performed on Amazon Mechanical Turk (MTurk) and in the lab were used to i) ensure that the compound words are easily identifiable even in emoji format and, ii) titrate the hues used in the color task, until task difficulty was well-matched between the color and the reading task.
Trial structure: At the beginning of a trial, participants were given a cue indicating which task should be performed ("Read" or "Color"), which was shown for 1 s. After the cue, the two images were shown consecutively, each for 1s, followed by an answer screen presented for 2s (Fig 1f, bottom). Subjects completed 8 runs, each lasting ~5 minutes and containing 24 trials. Stimulus and task conditions were balanced across runs so that each compound word came up equally often in each experimental condition.
Task: The experiment included two tasks, a reading, and a color judgment task. When performing the reading task, participants had to read the stimuli and decide whether they form a meaningful English compound word (match) or not (mismatch). The color judgment task required participants to carefully compare the hues of the two presented stimuli, and to decide whether they are the same (match) or slightly different (mismatch). Participants responded by selecting "match" or "mismatch" on the answer screen via a button press, whereas the correct answer was presented randomly either on the left or the right side of the answer screen.

2.4 Regions of Interest Definition
Functional Regions of Interest (fROIs) were defined in the native brain space of each participant, using both functional and anatomical criteria. We used all runs from the first experiment (localizer) to define the fROIs. The fROIs defined during experiment 1 were then used to evaluate responses in the other two experiments. We did not apply any spatial smoothing and avoided group-based analyses, as those increase the risk of spurious overlap between closely neighboring fROIs (Weiner & Grill-Spector, 2013). We defined two sets of fROIs in the ventral temporal cortex:

2.4.1 OTS-words
We localized the OTS-words subregions by contrasting responses to text stimuli with responses to other visual categories (limbs, faces, objects, and houses). Consistent with prior results (Lerma-

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Usabiaga et al., 2018), we found two self-contained regions in the middle and posterior occipitotemporal sulcus (OTS) that showed a preference for text over other visual stimuli (T=3, voxel level, uncorrected). We refer to them as middle OTS-words (mOTS-words) and posterior OTS-words (pOTS-words). We were able to define pOTS-words bilaterally in all 15 participants. On average, the size of pOTS-words was 722.8 mm^3 (SD = 459.9 mm^3, N = 15) in the left and 266.9 mm^3 (SD = 163 mm^3, N = 15) in the right hemisphere. We were able to identify mOTS-words in all 15 subjects in the left and in 11 subjects in the right hemisphere. The average size of mOTS-words was 444.2 mm^3 (SD = 350.7 mm^3, N=15) in the left and 234.6 mm^3 (SD = 313.5 mm^3, N = 11) in the right hemisphere.

2.4.2 Color-sensitive region

In addition to the OTS-words subregions, we defined three color-sensitive patches located along the midfusiform sulcus, containing voxels which showed significantly higher responses for colored compared to grayscale images (T=3, voxel level, uncorrected) (Lafer-Sousa et al., 2016). We refer to them as the anterior, central, and posterior color regions. The anterior color region (Ac-Color) could be identified in all 15 subjects in the left hemisphere, with an average size of 103.5 mm^3 (SD = 95.5 mm^3, N = 15), and in 13 subjects in the right hemisphere, with an average size of 179.8 mm^3 (SD = 163.6 mm^3, N = 13). The posterior color region (Pc-Color) could also be defined in both hemispheres for all 15 subjects and included on average 384 mm^3 (SD = 228.6 mm^3, N = 15) in the left and 395.7 mm^3 (SD = 190.7 mm^3, N = 15) in the right hemisphere. A central color region, Cc-color, could be defined bilaterally for all 15 subjects, spanning 409.4 mm^3 (SD = 238.7 mm^3, N = 15) in the left and 399.7 mm^3 (SD = 193.8 mm^3, N = 15) in the right hemisphere. Since Cc-color could be identified bilaterally in all subjects, and most closely neighbors the OTS-words subregions, we chose this region as the control fROI. All individual mOTS-words, pOTS-words and Cc-color regions in the left hemisphere are shown in Supplementary Figure S1.

Additionally, we also defined constant size 7.5 mm disk fROIs for mOTS-words, pOTS-words, and Cc-color in both the left and right hemispheres, in order to match fROI sizes across participants for the multivoxel pattern analyses (MVPAs, see below). These disk fROIs were centered on the respective region, and the radius of the disk was chosen to match the average size of the OTS-words subregions and Cc-color.

2.5 Statistical analysis

2.5.1 Univariate Analysis

We first extracted the average time course in percentage signal change for each condition from mOTS-words, pOTS-words, and the control region. Then we applied a general linear model (GLM) to estimate betas, indicating the magnitude of response for each condition. Consequently, the bar graphs in Figures 2, 4, and 6 show betas in units of % signal change ± SEM. Importantly, we used independent data for fROI definition (experiment 1) and signal extraction (experiments 2 and 3). We conducted repeated measures analyses of variance (rmANOVAs) on the betas from each fROI. **Exp2:** rmANOVAs used stimulus (emoji, text, mixed) and repetition (repeated vs. non-repeated) as factors; **Exp3:** rmANOVAs used stimulus (emoji or text) and task (reading or color judgment) as factors. We also conducted rmANOVAs for both **Exp 2** and **Exp 3** where we added ROI (mOTS-words, pOTS-words) as an additional factor to compare responses across the two

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

OTS-words subregions. Post-hoc tests were applied, when the rmANOVAs revealed significant main effects or interaction effects between factors. Where applicable, we used Bonferroni correction for multiple comparisons.

2.5.2 Multivoxel Pattern Analysis (MVPA)

We conducted MVPAs (Haxby et al., 2001) on distributed responses across the 7.5 mm disk fROIs for mOTS-words, pOTS-words, and Cc-color. A GLM was calculated to estimate the responses to each experimental condition separately for each voxel. The responses were then normalized by subtracting each voxel's mean response and z-standardized.

For each experiment and each condition, we calculated the correlation among each pair of multivoxel patterns (MVPs), using a leave-one-run-out procedure. These correlations were summarized in representational similarity matrices (RSMs). By using the leave-one-run-out procedure as cross-validation, we were able to evaluate the reliability of the observed MVP across runs. The RSMs were calculated for each subject separately and then averaged across subjects. In addition to that, we trained winner-takes-all (WTA) classifiers separately on the responses of experiments 2 and 3, to observe which experimental conditions can be decoded from the MVPs of each fROIs. In experiment 2, classifiers were initially trained to decode stimulus (emoji, text, mixed) and repetition (repeated vs. non-repeated). However, as the mixed condition includes both emoji and text stimuli, the classifier could not learn patterns associated with each stimulus format when this condition was included. We therefore reran the MVPAs for experiment 2 excluding all mixed trials. The RSM containing all 3 stimulus conditions (text, emoji, mixed) can be found in the Supplementary Material (Fig S2). For experiment 3, WTA-classifiers were trained to decode stimulus (emoji vs. text) and task (reading task vs. color judgment). All classifiers used a leave-one-run-out procedure, where the training set consisted of all runs but one, and the test set included the left-out run. Results report the aggregated classifier performances across all iterations of leave-one-run-out combinations. To quantify the classifiers' performance, we computed paired t-tests against chance level for stimulus and task decoding. Where applicable, Bonferroni correction for multiple comparisons was applied.

We also created RSMs combining responses from experiments 2 and 3, to test MVPs for stimulus (emoji vs. text) and tasks (reading, color judgment, black/white judgment) across experiments. We included all conditions from experiment 3, and all emoji and text trials from experiment 2, averaging repeated and non-repeated trials. We normalized the data by subtracting the mean response separately for each experiment before merging them, as the average response levels differed between experiments. As experiments 2 and 3 also differed in their number of runs, a leave-one-run-out procedure was not feasible, which is why we divided the data set into even and odd runs for the RSA. RSMs were calculated for each subject separately and then averaged across subjects.

Finally, we conducted additional rmANOVAs on the WTA classifier decoding accuracy with decoding type (**Exp 2**: stimulus and repetition, **Exp 3**: stimulus and task) and ROI (mOTS-words, pOTS-words) as factors, to compare decoding performance across the OTS-words subregions. Post-hoc tests were applied, when the rmANOVAs revealed significant main effects or interaction effects between factors. Where applicable, we used Bonferroni correction for multiple comparisons.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.5.3 Analysis of behavioral responses during fMRI

We calculated each subject's average performance (measured in % correct ± *SE*) and reaction times separately for each experiment. Further, we tested for significant performance differences between experimental conditions, by conducting rmANOVAs on the behavioral responses. **Exp1:** ANOVAs used stimulus (words, limbs, objects, houses, faces) as a factor; **Exp2:** ANOVAs used stimulus (emoji, text, mixed) and repetition of stimulus (repeated vs. non-repeated) as factors; **Exp3.** ANOVAs used stimulus (emoji or text) and task (reading or color judgment) as factors. Post-hoc tests were applied, when the rmANOVAs revealed significant main effects or interaction effects between factors. Where applicable, Bonferroni correction for multiple comparisons was applied.

2.6 Code and Data availability

The fMRI and T1w-anatomical data were analyzed using the open-source mrVista software (available in GitHub: http://github.com/vistalab/vistasoft) and FreeSurfer (available at https://surfer.nmr.mgh.harvard.edu/), respectively. Source data and code for reproducing all figures and statistics are made available in GitHub as well (https://github.com/EduNeuroLab/read_emojis_ots). The raw data collected in this study will be made available by the corresponding author upon reasonable request.

3. Results

3.1 Task difficulty was well-matched across conditions in each of the three experiments

The goal of the first experiment was to localize mOTS-words and pOTS-words in each participant by contrasting responses to text stimuli with those elicited by other visual stimulus categories. Subjects (N=15) were presented with images of text, limbs, faces, objects, and houses while they performed a 1-back task. We used rmANOVAs to compare performance and reaction time across conditions. On average, participants' accuracy (± SE) was 60% (±4 %). There was no significant difference in accuracy between the different stimulus categories ($F(4,11) = 0.45$, $p = 0.77$) or between colored or grayscale images ($F(1,14) = 0.81$, $p = 0.38$) and no interaction between stimulus and color ($F(4,11) = 1.68$, $p = 0.17$). Participants' reaction time (± SE) was on average 260 ms (± 40 ms) and did not differ between stimulus categories ($F(4,11) = 0.47$, $p = 0.76$) or between colored or grayscale images ($F(1,14) = 2.13$, $p = 0.12$). There was also no interaction between stimulus and color ($F(4,11) = 0.53$, $p = 0.71$).

The goal of the second experiment (fMRI-adaptation) was to test whether there is a difference in neural adaptation for emojis compared to text stimuli in mOTS-words and pOTS-words. The images were shown either in white or black and the participant's task was to press a button when white images were shown. We tested for differences in subjects' performance (± SE) as well as in their reaction times (± SE) by conducting rmANOVA of behavioral responses with repetition and stimulus as factors. On average, participants responded correctly in 91% (± 1%) of the trials. Participants' performance did not significantly differ between emoji, text, and mixed trials $F(2,13) = 0.40$, $p = 0.68$) or between repeated and non-repeated trials ($F(1,14) = 0.20$, $p = 0.66$). We found no interaction effect between stimulus and repetition ($F(1,14) = 0.70$, $p = 0.55$). The subjects' reaction time (± SE) was on average 304 ms (± 10ms). We found no significant differences in response time between emoji, text, and mixed trials ($F(2,13) = 0.99$, $p = 0.38$) or

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

between repeated and non-repeated trials (F(1,14) = 0.10, p = 0.75) and no interaction between stimulus and repetition (F(2,13) = 0.69, p = 0.5).

The third experiment aimed to uncover the influence of different tasks (reading vs. non-reading) and stimuli (text vs. emoji) on responses in mOTS-words and pOTS-words. We performed rmANOVAs, using task and stimulus as factors to evaluate performance and reaction times. On average, subjects' performance (± SE) was 95 % (± 2%) correct. There were no effects of task (F(1,14) = 0.33, p = 0.67) or stimulus (F(1,14) = 0.22, p = 0.65) and no interaction effect between stimulus and task (F(1,14) = 0.21, p = 0.65) on participants' accuracy. Participant's average reaction time (± SE) was 669 ms (± 25ms). We found a significant main effect of stimulus (F(1,14) = 8.17, p = 0.01) and task (F(1,14) = 6.50, p = 0.02) and a significant interaction between stimulus and task (F(1,14) = 7.99, p = 0.01) on participant's reaction times. Post-hoc analyses revealed significantly longer reaction times for emoji compared to text stimuli in the reading task (p = 0.0004).

3.2 Left pOTS-words shows higher responses for emoji stimuli than text stimuli independent of the task

We identified pOTS-words by contrasting responses to text stimuli with responses elicited by other visual categories (faces, objects, limbs, and houses). In all participants, pOTS-words was successfully identified in both hemispheres (Fig 2a shows an example pOTS-words in a representative subject). To test if pOTS-words is only responsive towards text stimuli or also other, readable stimuli, such as emojis, in experiment 2 we leveraged an fMRI-adaptation design. In the left hemisphere, we found no significant main effect of stimulus (F(2,13) = 1.53, p = 0.23), but a main effect of repetition (F(1,14)=14.99, p= 0.002), as well as, an interaction between repetition and stimulus (F(2,13) = 5.30, p = 0.01). Post-hoc tests revealed a significant effect of repetition for both text (p=0.001) and emoji stimuli (p=0.03), but not for the mixed condition (p=0.39) (Fig 2b).

In experiment 3, we further probed task and stimulus preferences in pOTS-words by manipulating the participant's tasks and the presented stimuli orthogonally. In the left hemisphere, pOTS-words showed a main effect of stimulus (F(1,14) = 12.37, p = 0.003), but no significant main effect of task (F(1,14) = 3.85, p = 0.07) and no significant interaction effects (F(1,14) = 1.91, p = 0.18) (Fig 2c). Post-hoc test revealed higher responses for emoji stimuli than text stimuli in left pOTS-words (p = 0.003). Results of the right hemisphere pOTS-words are presented in detail in Supplementary Fig. S3.

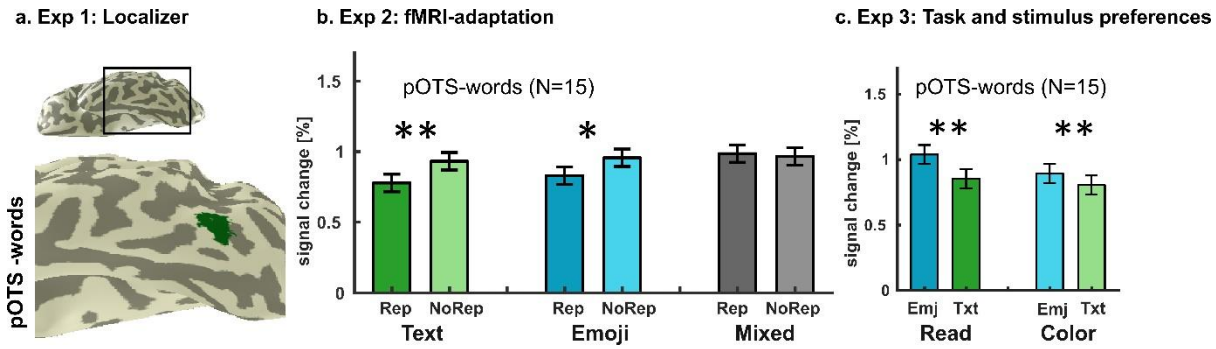

Figure 2. Left pOTS-words shows fMRI adaptation for both emoji and text stimuli in Exp 2 and an overall preference for emojis in Exp 3. a. Exp1: Localizer: Location of pOTS-words in the left hemisphere of a representative subject. pOTS-words was identified as a cluster of voxels in the posterior occipitotemporal sulcus that

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

responded higher to text than limbs, objects, houses, and faces (T = 3, voxel level, uncorrected). **b. Exp 2: fMRI-adaptation**: Mean responses ±SEM of pOTS-words across all subjects (N=15). A main effect of repetition and an interaction between repetition and stimulus was observed. Post-hoc tests revealed significant fMRI adaptation for text and emoji stimuli, but not the mixed condition. Stars indicate post-hoc results, showing significant differences between repeated and non-repeated trials; * p <0.05, ** p < 0.01. **c. Exp 3: Task and stimulus preferences.** Mean responses ± SEM of pOTS-words across all subjects (N=15). A main effect of stimulus was observed, whereas post-hoc tests showed higher responses for emoji than text stimuli for both tasks. Stars indicate post-hoc results; * p <0.05, ** p < 0.01, *** p<0.001. *Abbreviations*: Rep = repeated trials, NoRep = non-repeated trials, Emj = emoji trials, Txt = text trials.

3.3 Distributed responses across left pOTS-words encode the visual stimulus

In addition to the univariate results presented above, we also probed what information can be decoded from distributed responses within pOTS-words. For this, we performed MVPA within 7.5mm disk ROIs centered on pOTS-words. We used a constant disk ROI to equate the number of voxels used in the MVPA across participants.

In experiment 2, we generated an RSM comparing repeated and non-repeated emoji and text trials using a leave-one-run-out-approach. Thus, the diagonal of the RSM indicates similarity with the same condition across runs, and the off diagonal indicates similarity among different conditions across runs. The RSM reveals positive correlations for the same condition across runs as well as for the same stimuli across repeated and non-repeated trials (Fig 3a). To quantify if information about stimulus and/or repetition can be successfully decoded from the distributed responses in pOTS-words, we trained WTA (winner-takes-all) classifiers. We found that the stimulus but not repetition can be successfully decoded from distributed pOTS-word responses. Classifier performance reached an accuracy (± SEM) of 86.28% ± 4.72% for text and 88.33% ±3.93% for emoji stimuli, which in both cases significantly exceeded 50% chance level (text: p< 0.0001, emojis: p < 0.0001, significant with Bonferroni corrected threshold of p<0.01). Classifier performances for distinguishing between repeated and non-repeated trials reached an accuracy (± SEM) of 44.22% ± 4.76% for repeated and 57.11% ±4.30% for non-repeated trials, which did not exceed chance level (repeated trials: p = 0.88, non-repeated trials: p = 0.06) (Fig 3b). Accordingly, a direct comparison showed that the classifier decoding stimulus yielded a significantly higher performance than the classifier differentiating repeated and non-repeated trials (p < 0.0001). An RSM that also includes the mixed condition can be found in Supplementary Figure S2b.

In experiment 3, we created an RSM comparing stimuli (text, emoji) and tasks (reading, color, Fig 3c) again using a leave-1-run-out approach. As Fig 3c shows, the RSM revealed positive correlations for the same stimulus under different tasks. In addition, WTA classifiers were trained separately for stimulus and task, to determine which kind of information (stimulus, task, or both) can be successfully decoded from distributed pOTS-words responses. Classification performances (± SEM) for decoding text stimuli (85.42% ± 4.88%) and emoji stimuli (87.92% ± 4.30%) were significantly higher than 50% chance level (emoji trials: p<0.0001, text trials: p<0.0001, significant with Bonferroni corrected threshold of p<0.01). When differentiating between tasks, the classifier's performance (± SEM) was on average at 65%. Its performance (± SEM) was 68.54 % ± 3.50% for reading and 61.88% ± 3.92% for color judgments, with decoding accuracy for reading significantly exceeding 50% chance level (reading: p = 0.008, significant with Bonferroni corrected threshold of p<0.01; color judgment: p = 0.02, not significant with Bonferroni corrected threshold of p<0.01) (Fig 3d). A direct comparison showed that decoding accuracy for decoding the stimulus was significantly higher than for decoding the task (p < 0.0001). These results suggest that while

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

both the stimulus and the reading task are decodable from distributed responses in pOTS-words, pOTS-words contains more information regarding the visual stimulus than the task.

Finally, we merged data from experiments 2 and 3 to further probe distributed responses elicited by different stimuli and tasks across experiments. For experiment 2 we averaged across repeated and non-repeated trials. The resulting RSM shows that MVPs for the same stimulus (emoji and text) are positively correlated across experiments, suggesting that these stimuli induce stable and reproducible MVPs across the three tasks (reading, color detection, black/white detection). In contrast, there is no similarity in responses across stimuli within a task. Thus, our results suggest that pOTS-words is predominantly involved in encoding stimuli not tasks.

Results for the multivariate analysis for pOTS-words in the right hemisphere were similar to the left hemisphere results described above (Supplementary Fig S4).

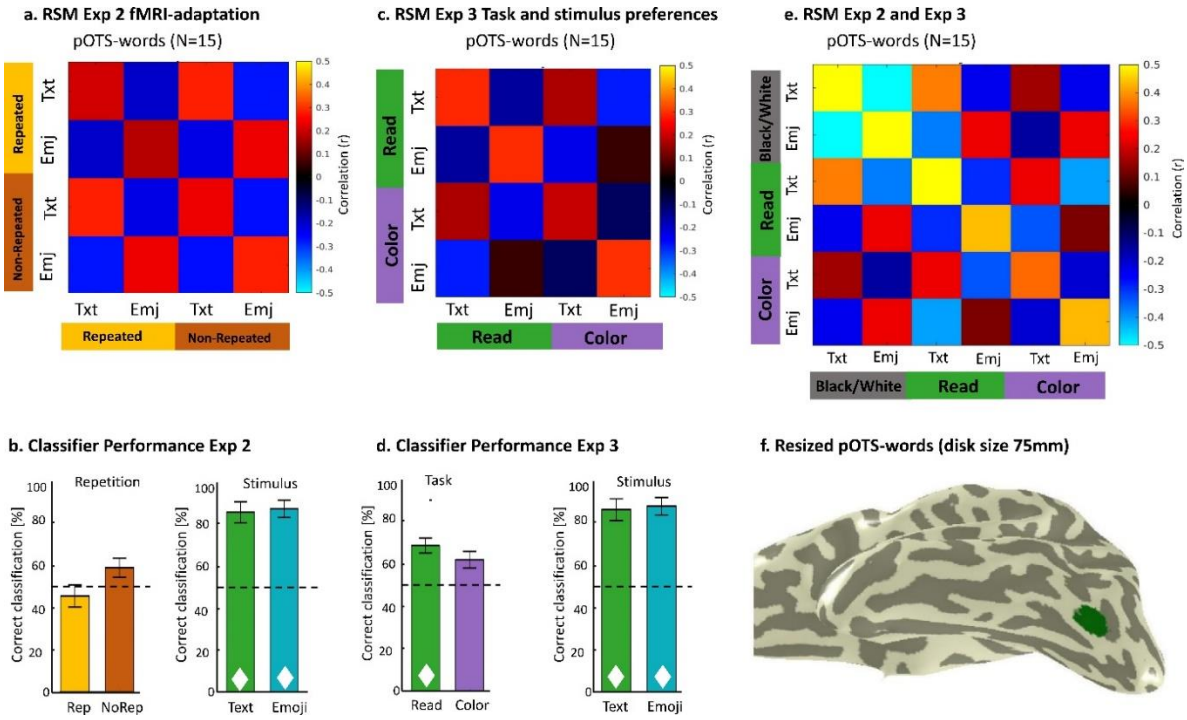

Figure 3. Distributed responses of left pOTS-words encode different stimuli, both within and across experiments. **a.** RSM for experiment 2 data from pOTS-words in the left hemisphere across all subjects. Conditions are arranged by stimulus (text vs. emoji) and grouped by repetition (repeated vs. non-repeated). **b.** Mean ± SEM WTA classification performance in experiment 2 for stimulus and repetition. Classifier performance significantly exceeded 50% chance level for decoding both stimulus categories but not repetition. **c.** RSM for experiment 3 data from pOTS-words in the left hemisphere across all subjects. Conditions are arranged by stimulus (text vs. emoji) and grouped by task (reading vs. color). **d.** Mean ± SEM of WTA classification performance in experiment 3 for stimulus and task. Classifier performance for decoding the reading task and for decoding both stimulus categories significantly exceeded chance level. **e.** Joined RSM for both experiments 2 and 3 data from pOTS-words in the left hemisphere across all subjects. Conditions are arranged by stimulus (text vs. emoji) and grouped by task (reading, color judgment, black/white detection). **f.** Disk fROI centered on pOTS-words is shown in a representative subject and was used for all MVPA analyses. *Dotted line*: chance level (50%); ✦ classifier performance significantly above chance level with a Bonferroni corrected threshold of $p < 0.01$; *Abbreviations*: RSM=representational similarity matrix, WTA=winner-takes-all, Rep: repeated trials, NoRep = non-repeated trials, Emj = emoji trials, Txt = text trials.

3.4 Left mOTS-words shows higher responses for emoji than text stimuli during the reading task

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

From experiment 1 data, we identified mOTS-words in each participant's native brain space by contrasting responses to text stimuli with those evoked by other visual stimuli. mOTS-words was successfully identified in all 15 subjects in the left hemisphere and in 11 subjects in the right hemisphere. As an example, mOTS-words from a representative subject is shown in Fig 4a. To test if mOTS-words is sensitive only to text or also to other readable visual stimuli, such as emojis, we examined the data from the fMRI adaptation experiment. We conducted rmANOVAs with factors of repetition (repeated, non-repeated) and stimulus format (emoji, text, mixed). In the left hemisphere, mOTS-words did not show significant effects of repetition ($F(1,14) = 1.12$, $p = 0.31$) or stimulus format ($F(2,13) = 0.90$, $p = 0.42$) (Fig 4b). Further, using the data from the third experiment, we tested task and stimulus preferences in mOTS-words. rmANOVAs from left mOTS-words with task (reading, color) and stimulus (text, emoji) as factors showed a main effect of stimulus ($F(1,14) = 7.54$ $p = 0.02$) and a main effect of task ($F(1,14) = 13.28$, $p = 0.003$) (Fig 4c) as well as an interaction between task and stimulus ($F(1,14) = 7.66$, $p = 0.02$). Post-hoc t-tests revealed higher responses for emoji than text stimuli during the reading task ($p = 0.0005$) as well as higher responses for the reading compared to the color task ($p=0.003$). Data from the right mOTS-words are presented in Supplementary Fig. S5).

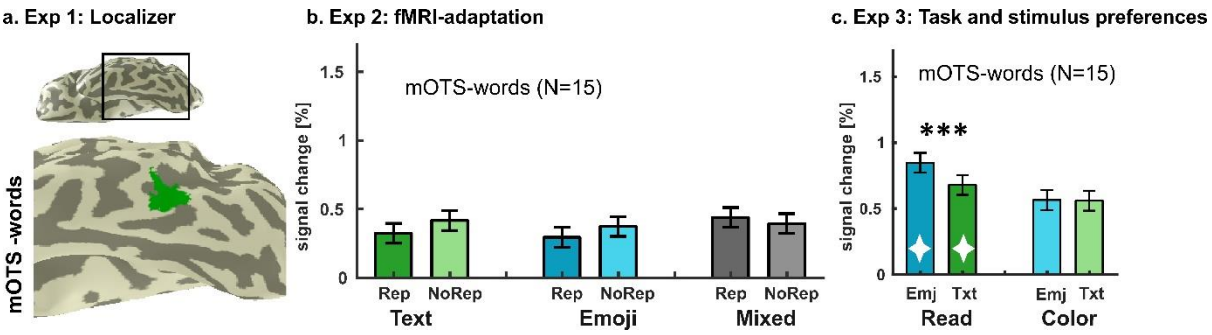

Figure 4. Left mOTS-words shows higher responses for emoji than text stimuli during a reading task. a. Exp 1: Localizer. Location of mOTS-words in the left hemisphere of a representative subject. mOTS-words was identified as a cluster of voxels that respond higher to text than limbs, objects, houses, and faces (T = 3, voxel level, uncorrected). **b. Exp 2: fMRI-adaptation**: Mean responses of mOTS-words across all subjects (N=15) ±SEM. No significant effects of repetition or stimulus were detected. **c. Exp 3: Task and stimulus preferences.** Mean responses ±SEM of mOTS-words across all subjects (N=15). We found a main effect for task and stimulus as well as an interaction effect between task and stimulus. Post-hoc tests revealed a preference for emoji stimuli during the reading task. Stars indicate results of post-hoc tests, showing significantly higher responses for emoji stimuli during a reading task, * $p < 0.05$, ** $p < 0.01$,*** $p<0.001$; diamonds indicate significantly higher responses for one task vs. the other task, $p < 0.05$. *Abbreviations*: Rep: repeated trials, NoRep = non-repeated trials, Emj = emoji trials, Txt = text trials.

3.5 Distributed responses across left mOTS-words encode stimulus and task

In addition to the univariate results presented above, we also probed what information can be decoded from distributed responses within 7.5mm disk ROIs centered on left mOTS-words.

In experiment 2, we calculated a RSM comparing MVPs of repeated and non-repeated emoji and text trials. We observed high MVP correlations for the same stimuli across repeated and non-repeated trials (Fig 5a; an RSM that also includes the mixed condition can be found in the Supplementary Fig S2a). Further, WTA classifiers performances (± SEM) reached 79.94±3.22% accuracy for text and 77.83±4.55% accuracy for emoji stimuli, which in both cases significantly exceeded 50% chance level (text : $p<0.0001$, emojis: $p<0.0001$, significant with Bonferroni

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

corrected threshold of $p<0.01$) (Fig 5b). Classifier performances (±SEM) for distinguishing between repeated and non-repeated trials was at 48.94±3.40%, for repeated and at 56.06±3.11% for non-repeated trials, which did not exceed 50% chance level (repeated: $p=0.62$, non-repeated: $p=0.04$, not significant with Bonferroni corrected threshold of $p<0.01$). Accordingly, the classifier decoding stimulus performed significantly better than the one decoding repetition ($p<0.0001$).

In experiment 3, we created a RSM comparing MVPs across stimuli (emoji, text) and tasks (reading, color judgment) (Fig 5c). The RSM shows high MVP correlations only along the diagonal, suggesting that MVPs in mOTS-words distinguish between both stimuli and tasks. To further quantify these effects, we trained WTA classifiers, separately for stimulus and task, to determine which kind of information (stimulus, task, or both) can be successfully decoded from MVPs in mOTS-words (Fig 5d). The classifier's accuracy (± SEM) for decoding the reading task was 78.54±4.17% and for decoding the color task was 74.38± 4.06%, which in both cases, was significantly higher than chance level (reading task: $p<0.0001$, color task: $p<0.0001$, significant with Bonferroni corrected threshold of $p<0.01$). The classifier trained to differentiate between emoji and text stimuli had an accuracy of 80.42±4.07% for decoding emojis and 72.50±3.29% for decoding text stimuli. Classifier performances significantly exceeded chance level for both stimulus conditions (emoji stimuli: $p<0.0001$, text stimuli: $p<0.0001$, significant with Bonferroni corrected threshold of $p<0.01$). We found no significant differences, when we compared the classifier's accuracy for stimulus vs. task decoding ($p = 1$).

Next, we calculated an RSM combining data from experiments 2 and 3 and compared MVPs for emoji and text trials across three tasks (black/white detection, reading, and color judgment) (Fig 5e). We used all trials from experiment 3 and the emoji and text trials from experiment 2, averaged across repeated and non-repeated trials. The resulting RSM showed high correlations mostly along the diagonal suggesting that changing either task or stimulus generates a distinct pattern of response in left mOTS-words.

Results for the multivariate analysis for mOTS-words in the right hemisphere can be found in Supplementary Fig S6.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

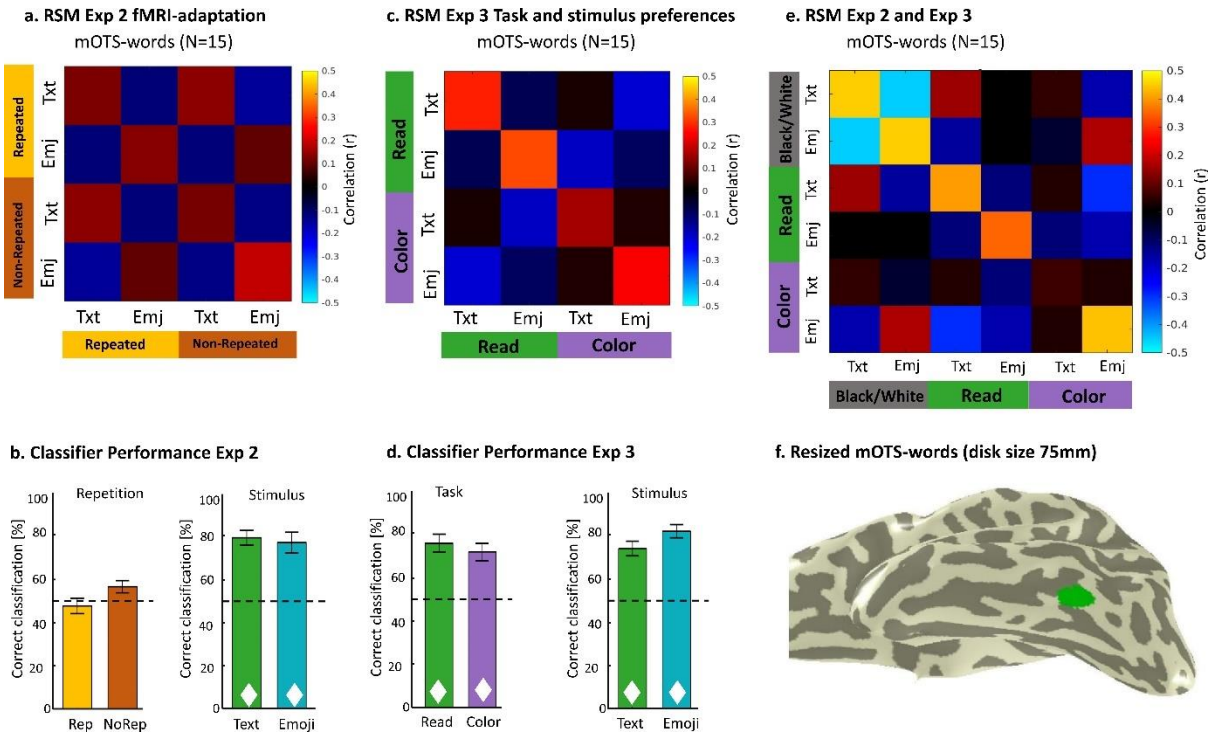

Figure 5. Distributed responses of left mOTS-words encode stimuli and tasks, both within and across experiments. **a.** Mean RSM for experiment 2 for mOTS-words in the left hemisphere across all subjects. Conditions are arranged by stimulus (text vs. emoji) and grouped by repetition (repeated vs. non-repeated trials). **b.** Mean ± SEM WTA classification performance in experiment 2 for stimulus and repetition. Classifier performance significantly exceeded 50% chance level for decoding both stimuli but not decoding repetition. **c.** Mean RSM for experiment 3 for mOTS-words in the left hemisphere across all subjects. Conditions are arranged by stimulus (text vs. emoji) and grouped by task (reading vs. color judgment). **d.** Mean ± SEM WTA classification performance in experiment 3 for stimulus and task. Classifier performance was significantly above 50% chance level for decoding both stimuli and tasks. **e.** Mean RSM across experiments 2 and 3 for mOTS-words in the left hemisphere across all subjects. Conditions are arranged by stimulus (text vs. emoji) and grouped by task (black/white detection vs. reading vs. color judgment). **f.** Disk fROI centered on mOTS-words is shown in a representative subject and was used for all MVPA analyses. *Dotted line*: chance classification level (50 %); ✦ classifier performance significantly above chance level with a Bonferroni corrected threshold of p<0.01; *Abbreviations*: RSM=representational similarity matrix, WTA=winner-takes-all, Rep: repeated trials, NoRep = non-repeated trials, Emj = emoji trials, Txt = text trials.

3.6 Cc-color shows higher responses for the color task than the reading task

In order to identify the color patches along the midfusiform sulcus, in experiment 1, we contrasted responses to colored images with gray-scale images. Three color regions were defined. Here we focus on Cc-color, as it was successfully identified in both hemispheres for all 15 subjects and is adjacent to both OTS-words subregions. Cc-color was used as a control region to evaluate the specificity of the response patterns observed in mOTS-words and pOTS-words (Fig 6a).

In experiment 2, we used fMRI adaptation to probe stimulus representations in Cc-color. In the left hemisphere, we found a main effect of stimulus ($F(2,13)=32.98$, $p<0.0001$) and a main effect of repetition ($F(1,14)=5.37$, $p=0.04$). Furthermore, we found a significant interaction between repetition and stimulus ($F(2,13)=4.11$, $p=0.03$) (Fig 6b). Post-hoc tests revealed a significant difference between repeated and non-repeated trials for the emoji stimuli ($p = 0.007$), but not for the text ($p=0.39$) or the mixed ($p=0.94$) condition.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In experiment 3, where tasks and stimuli were manipulated orthogonally, Cc-color showed a main effect of stimulus ($F(1,14)=81.62$, $p<0.0001$) with a preference for emoji stimuli over text stimuli. Moreover, we found a significant main effect of task ($F(1,14)=20.85$, $p<0.0001$), indicating a preference for the color judgment task over the reading task in Cc-color. This established a double-dissociation of task preferences with mOTS-words (Fig 6c). We found no interaction between task and stimulus ($F(1,14) = 2.16$, $p = 0.16$) in Cc-color. Results of the right hemispheric central color region are presented in Supplementary Fig S7.

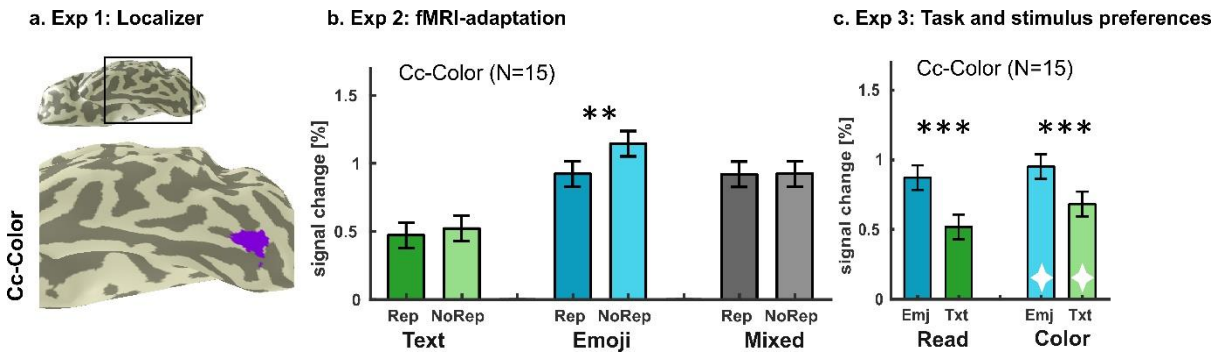

Figure 6. The left central color region shows preferences for emojis and the color judgment task. **a. Exp 1: Localizer**: Location of Cc-color in the left hemisphere of a representative subject. Cc-color was identified as voxels medial to the mfusiform gyrus that responded higher to colored compared to grey-scale images (T = 3, voxel level, uncorrected). **b. Exp 2: fMRI-adaptation:** Mean responses ±SEM of Cc-color across all subjects (N=15). Main effects of stimulus and repetition as well as an interaction between stimulus and repetition were observed. Post-hoc tests showed significant fMRI adaptation only for emoji stimuli. Stars indicate results of post-hoc tests, showing significant differences between repeated and non-repeated trials for the emoji condition; * $p<0.05$, ** $p < 0.01$. **c. Exp 3: Task and stimulus preferences.** Mean responses ±SEM of Cc-color across all subjects (N=15) in experiment 3. A main effect of stimulus was observed, with higher responses for emoji than text stimuli for both tasks. Stars indicate post-hoc results, * $p<0.05$, ** $p < 0.01$,*** $p<0.001$; diamonds indicate significantly higher responses for one task vs. the other task, $p<0.05$. *Abbreviations*: Exp: experiment, Rep: repeated trials, NoRep = non-repeated trials, Emj = emoji trials, Txt = text trials

3.7 Distributed responses across the left central color region encode the visual stimulus

In addition to the univariate results presented above, we also probed what information can be decoded from distributed responses within Cc-color. For this, we performed MVPA analyses within 7.5mm disk fROIs centered on Cc-color.

In experiment 2, we generated an RSM comparing MVPs across repeated and non-repeated text and emoji trials (an RSM that also includes the mixed condition can be found in the Supplementary Fig S2c). The RSM revealed a clear stimulus effect, showing positive correlations between repeated and non-repeated trials when the same stimulus was presented (Fig 7a). We then trained WTA classifiers to test what kind of information (stimulus, repetition, or both) can be successfully decoded from distributed responses in Cc-color (Fig 7b). Classification performance (mean ± SEM) for decoding text yielded 82.67± 5.61% accuracy and accuracy for decoding emojis was 80.72± 5.62%. In both cases, decoding accuracy was significantly above 50% chance level (text: $p<0.0001$, emojis: $p<0.0001$, significant with Bonferroni corrected threshold of $p<0.01$). Classification performance (±SEM) for detecting repeated trials was 52.50±3.06% and decoding

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

accuracy was 64.00±4.07% for non-repeated trials. Only when decoding non-repeated trials, classifier performance significantly exceeded 50% chance level with a Bonferroni corrected threshold of $p<0.01$ (repeated: $p = 0.21$, non-repeated: $p = 0.002$). When comparing the average performances for decoding stimulus vs. repetition, we found that the classifier decoding stimulus performed significantly better ($p = 0.01$).

In experiment 3, we created an RSM comparing MVPs across different tasks and stimuli (Fig 7c). Examination of this RSM revealed a strong stimulus effect, as MVPs for the same stimulus under different tasks were positively correlated. Accordingly, WTA classification performances (mean±SEM) for emoji trials (81.88±3.30%) and text trials (80.42±2.79%) were significantly higher than 50% chance level (emoji trials: $p<0.0001$, text trials: $p<0.0001$, significant with Bonferroni corrected threshold of $p<0.01$). The classifier's performance (±SEM) for distinguishing between the reading and color judgment tasks were 61.46± 4.18% for reading and 57.92±3.39%) for color judgment, neither of which significantly exceeded 50% chance level (reading: $p = 0.05$, color judgment: $p = 0.09$, not significant with Bonferroni corrected threshold of $p<0.01$) (Fig 7d). The average classification performance for decoding the stimuli was significantly higher than for decoding the tasks ($p < 0.0001$).

Finally, to investigate if similar stimuli induce reproducible MVPs in Cc-color across experiments, we calculated an RSM combining data from experiments 2 and 3. This RSM showed positive correlations for similar stimuli, even across experiments that included different tasks. In sum, these results suggest that MVPs in Cc-color contains information regarding the visual stimulus, but not regarding the task the participants are performing. A similar pattern is observable for the right hemisphere; these results are depicted in Supplementary Figure S8.

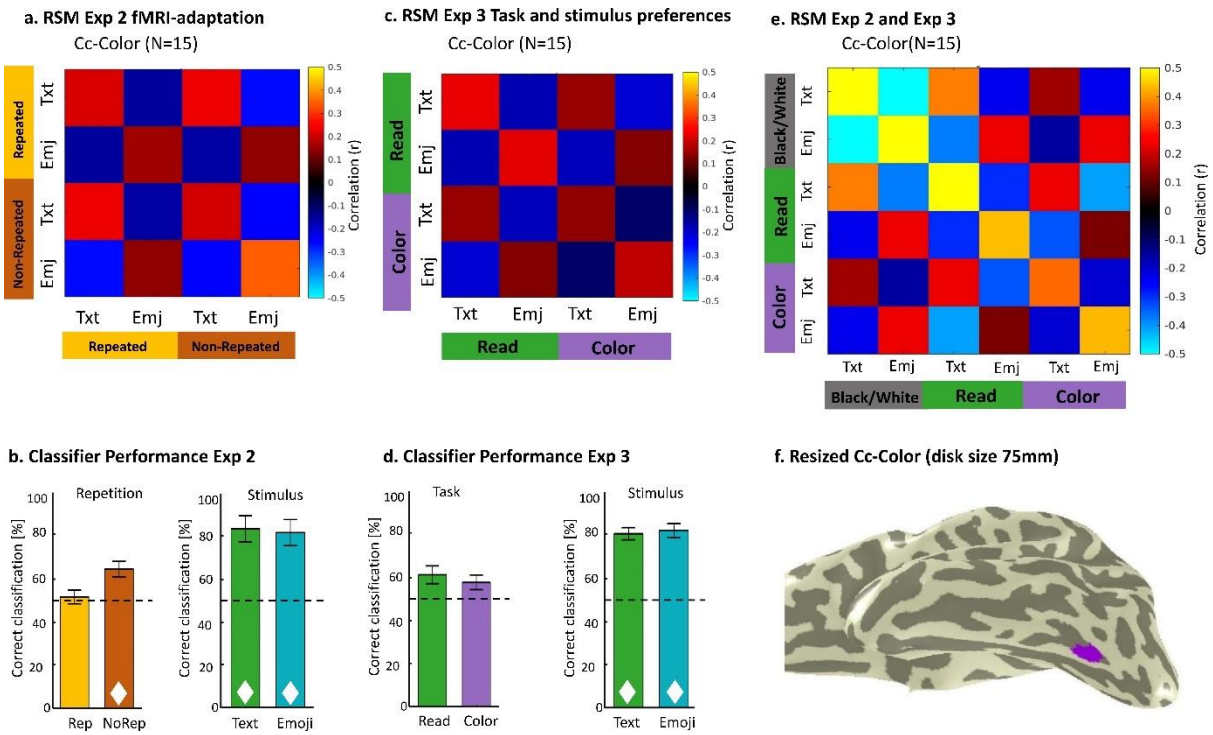

Figure 7. Distributed responses of left Cc-color encode visual stimuli, both within and across experiments. **a.** Mean RSM from experiment 2 for Cc-color in the left hemisphere across all subjects. Conditions are arranged by stimulus (text vs. emoji) and grouped by repetition (repeated vs. non-repeated). **b.** Mean±SEM WTA classification performance in experiment 2 for stimulus and repetition. Classifier performance significantly exceeded chance level

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

for decoding both stimulus categories and non-repeated trials. **c.** Mean RSM from experiment 3 for Cc-color in the left hemisphere across all subjects. Conditions are arranged by stimulus (text vs. emoji) and grouped by task (reading vs. color judgment). **d.** Mean±SEM WTA classification performance in experiment 3 for stimulus and task. Classifier performances for decoding stimulus but not task significantly exceeded 50% chance level. **e.** Mean RSM for joined analysis of experiments 2 and 3 for Cc-color in the left hemisphere across all subjects. Conditions are arranged by stimulus (text vs. emoji) and grouped by task (reading vs. color judgment vs. black/white detection). **f.** Disk fROI centered on Cc-color is shown in a representative subject and was used for all MVPA analyses. ✦ classifier performance significantly above chance level with a Bonferroni corrected threshold of $p < 0.01$; *Abbreviations*: RSM: representational similarity matrix, WTA = winner-takes-all, Rep = repeated trials, NoRep = non-repeated trials, Emj = emoji trials, Txt = text trials.

3.8 Task and stimulus decodability differ between the OTS-words subregions

After examining mOTS-words and pOTS-words separately, we wanted to explicitly test if there are significant differences in task and stimulus selectivity across these regions. To this end, for the univariate data, we conducted additional rmANOVAs adding ROI (mOTS-words, pOTS-words) as a factor. In experiment 2, we found no main effect of stimulus ($F(2,13) = 1.56$, $p = 0.23$), but a main effect of repetition ($F(1,14) = 5.10$, $p = 0.04$) and ROI ($F(1,14) = 34.00$, $p < 0.0001$). We also observed a significant interaction between stimulus and repetition ($F(2,13) = 4.94$, $p = 0.01$); post hoc tests revealed significant differences between repeated and non-repeated trials for text stimuli only (text: $p = 0.008$, emoji: 0.09, mixed: 0.23). We found no interaction between repetition and ROI ($F(1,14) = 1.88$, $p = 0.19$) and stimulus and ROI ($(2,13) = 0.63$, $p = 0.58$). The three-way interaction of ROI, stimulus and repetition was not significant as well ($F(2,13) = 0.17$, $p = 0.85$). In experiment 3, we found a main effect of task ($F(1,14) = 11.56$, $p = 0.004$), a main effect of stimulus ($F(1,14) = 27.48$, $p = 0.0001$) and a main effect of ROI ($F(1,14) = 6.86$, 0.02). No significant interaction effect between ROI and stimulus ($F=0.776$, $p=0.39$) was found. The interaction effect between task and stimulus was marginally significant ($F(1,14) = 4.53$, $p = 0.05$). The interaction between task and ROI was not significant, although a trend was observed ($F = 3.35$, $p = 0.09$). The three-way interaction between task, stimulus, and ROI was not significant ($F(1,14) = 2.13$, $p = 0.17$).

To further investigate if there are significant differences in the information contained in the distributed responses across the OTS-words subregions, we conducted rmANOVAs on the WTA classifier decoding accuracy with decoding type (Exp 2: stimulus and repetition, Exp 3: stimulus and task) and ROI (mOTS-words, pOTS-words) as factors. In experiment 2, we found a main effect of decoding type ($F(1,14) = 47.96$, $p < 0.0001$), but no main effect for ROI ($F(1,14) = 1.29$, $p = 0.28$), and no interaction effect ($F(1,14) = 3.18$, $p = 0.10$). For experiment 3, we found a significant main effect of decoding type ($F(1,14) = 9.10$, $p = 0.009$), no main effect of ROI ($F(1,14) = 0.039$, $p = 0.8$), but a significant interaction between ROI and decoding type ($F(1,14) = 11.55$, $p = 0.004$). Post-hoc tests revealed a significant difference between mOTS-words and pOTS-words in classifier performance for decoding both stimulus ($p =0.02$) and task ($p = 0.02$), whereas mOTS-words showed better task decoding and pOTS-words showed better stimulus decoding.

Discussion

The goal of this study was to disentangle task and stimulus preferences in the OTS-words subregions by introducing a novel stimulus type, readable emojis. As in prior studies, we identified two sub-regions of OTS-words, by contrasting responses to text stimuli with those for other visual stimulus categories. Our univariate analyses of these sub-regions showed fMRI-adaptation for

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

readable text and emoji stimuli in pOTS-words but not in mOTS-words. Further, pOTS-words showed a preference for emoji stimuli independent of task and mOTS-words showed a preference for emoji stimuli during a reading task. Multi-variate analyses aligned with the univariate results, showing that pOTS-words encodes the visual features of the stimulus, while mOTS-words appears to be sensitive to both the stimulus and the task.

Evidence for an anterior-to-posterior processing gradient along the occipito-temporal sulcus

With this study, we addressed a central debate about the functional role of OTS-words in reading, by testing if OTS-words subregions are more sensitive to text stimuli than to other visual stimuli. Moreover, we tested how different tasks influence responses in these sub-regions. We derived three hypotheses regarding the functional role of the OTS-words subregions from the current literature: H1: Both OTS-words subregions show higher responses to text stimuli compared to other visual stimulus categories, irrespective of the performed task (Cohen et al., 2000; Dehaene & Cohen, 2011) (Fig 1a). H2: Both OTS-words subregions show higher responses to reading tasks compared to other tasks, irrespective of the visual stimuli (Fig 1b). H3: There is an anterior-to-posterior gradient in processing level between mOTS-words and pOTS-words, whereas the anterior mOTS-words is selective for reading tasks and the posterior pOTS-words is selective for the visual features of text (Caffarra et al., 2021; Lerma-Usabiaga et al., 2018). We did not find evidence for H1, as both OTS-words subregions did not show higher activations for text stimuli when compared to readable emojis. Our findings also do not support H2, as we detected no preference for a reading task over a color judgment task in one of the subregions, pOTS-words, though we found this preference in mOTS-words. Our results align best with H3, especially as our multivariate results showed that pOTS-words contains more information about the visual stimulus than mOTS-words, which contains more information about the task the participants are performing. However, contrary to H3, our univariate results also revealed a stimulus preference in mOTS-words in addition to its preference for the reading task and the direct comparisons of univariate response profiles between the subregions were not significant.

The observed differences between mOTS-words and pOTS-words align with previous studies proposing an increasing processing level along the ventro-temporal cortex (Grill-Spector & Weiner, 2014; Stigliani et al., 2015). First, previous work showed that the perception of letter-form is processed in the more posterior part of the VTC, while the anterior part is more responsive to perceiving entire words as language units (Thesen et al., 2012; Vinckier et al., 2007b). Further, it was shown that pOTS-words can be identified by contrasting words with checkerboards, while mOTS-words can be identified by contrasting words with pseudo-words. Authors hence propose that pOTS-words is involved in the extraction of visual features, while mOTS-words plays a role in integrating information with other regions of the language network (Lerma-Usabiaga et al., 2018). This idea is further supported by a study in which participants were trained to read and understand an artificial language and scripture system: RSAs showed, that while the posterior part of the VTC was more sensitive towards basic visual similarities of words, the middle to anterior VTC encoded phonological and semantic similarity (Taylor et al., 2019). Moreover, Dehaene et al. (2005) theorized in the local combination detector (LCD) model, that words are encoded through a posterior-to-anterior hierarchy of neuronal populations, which are tuned to process increasingly more complex features of words. The observed differences between pOTS-words and mOTS-words also relate to the phenomenon that only one word can be recognized at a time (White et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2018). White et al. (2019) showed that while pOTS-words can respond to multiple words simultaneously at different locations in the visual field, mOTS-words processes words sequentially. The authors suggest that while pOTS-words responds to the visual features of written language, enabling parallel processing of multiple words, mOTS-words is tuned to orthographic and lexical properties and therefore processes only one word at a time (White et al., 2019). Differential functional properties of mOTS-words and pOTS-words is further supported by evidence from intracranial recordings, showing a temporal dissociation between early orthographic effects (~ 150–250 ms after stimulus onset) in the posterior regions of the VTC and later lexical responses in the more anterior VTC (Woolnough et al., 2021). In addition to that, studies show that even though dyslexic children show activations in OTS-words during reading tasks, the processing hierarchy within OTS-words is impaired, leading to reading difficulties (van der Mark et al., 2009). Together, these observed changes in processing level along VTC suggest that mOTS-words and pOTS-words should be evaluated separately to fully grasp the functional role of OTS-words in reading. For this, it is helpful to perform all analyses in the native brain space of the participants without using spatial smoothing, as normalization and smoothing can induce spurious overlap between closely neighboring regions (Weiner & Grill-Spector, 2013).

Are the OTS-words subregions specialized for processing text?

In our study, both OTS-words sub-regions showed a preference for emoji stimuli over text stimuli. This is surprising, as most prior studies have argued for a selectivity for text over other visual stimuli in OTS-words (Cohen et al., 2000b; Cohen & Dehaene, 2004; Szwed et al., 2011). However, other studies have questioned this preference for text, showing that activations in OTS-words can be elicited by other visual stimulus categories as well (Ben-Shachar et al., 2007; Mei et al., 2010; C. J. Price & Devlin, 2003). To determine what drives the activations in OTS-words, we identified this region in the same fashion as previous studies (Keller et al., 2022; Rosenke et al., 2021; Stigliani et al., 2015) but in addition to that introduced emojis as a new "readable" visual stimulus category. This allowed us to disentangle task and stimulus preferences, which is not possible if the only readable stimuli in the experiment are text stimuli. The observed preference for emojis over text stimuli in both OTS-words subregions provides evidence that these brain regions might not solely be selective for text. This difference between our findings and the previous literature may be explained by i) in our study responses to text are compared to responses to another readable stimulus, ii) here we introduced a task that required the participants to read all stimuli in the experiment, iii) we present emoji stimuli, with which the subjects have prior experience and which relate to archaic forms of writing. We discuss each of these differences separately in the paragraphs below.

The observed preference for emojis may be driven by their readability

The emoji stimuli presented in the current study differed from previous work in that they were readable, i.e., the two stimuli together formed meaningful English compound words. Typically in prior studies (Cohen et al., 2000b; Cohen & Dehaene, 2004; Szwed et al., 2011), the only "readable" stimuli were the text stimuli, which were contrasted with non-readable visual stimulus categories. The readability of the presented stimuli might be one of the driving mechanisms for increased responses in OTS-words. Similar to learning a new writing system, abstract stimuli can become "readable" when we learn to retrieve semantic meaning from visual features which previously had no significance to us. For example, in a recent study subjects were trained to read "HouseFont",

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

an artificial writing system, in which English phonemes were presented as different houses (Martin et al., 2019). Strikingly, the authors found activation in OTS-words after the training was completed and the participant were able to read the new font. This aligns with our findings, suggesting that the OTS-words subregions are activated when semantic meaning needs to be assigned to an abstract stimulus. One important difference between their work and ours is that Martin et al. used stimuli that initially were not related to the word that they represent, whereas our emoji stimuli were pictograms of the objects themselves. These differences raise the question which kind of stimuli can be turned into readable units. Future research could for example investigate if readability might also be invoked for stimuli such as photos of objects or natural scenes, if multiple stimuli together convey a word. Another difference between the studies is that, in the work by Martin et al. each house was mapped onto a phoneme, which then needed to be combined into words. In our study, each emoji represented a word by itself, and two emojis needed to be combined into a compound word. Interestingly, analogous to these differences, across different scriptures, the used symbols can either represent individual letters (e.g., English), individual phonemes (e.g., Finnish) or individual words (logograms, e.g., Japanese kanji). This suggests that OTS-words might be surprisingly flexible regarding the mapping between symbols and the language units they represent.

The observed preference for emojis may be driven by the reading task

Our findings highlight the importance of the performed task in driving responses in the OTS-words subregions. For example, during experiment 2 (fMRI adaptation), we did not find any effect of stimulus or repetition in mOTS-words. However, in experiment 3 (task and stimulus preferences), where a reading task was introduced, mOTS-words showed both task and stimulus effects. This observation might indicate the importance of providing an actual reading task, in addition to the readability of the stimulus when evaluating responses in OTS-words. This idea aligns with a recent study in which participants were instructed to either judge the color of a single word or to read it (Qu et al., 2022). The authors found higher activations in OTS-words for the reading task, compared to the color judgement task. Similarly, in our previous work we found higher activations in OTS-words during a reading task than a color judgement or adding task, even when all tasks were performed on morphs between letters and numbers (Grotheer et al., 2018, 2019). Further, others found that activity in OTS-words did not differ between reading words and trying to read non-words (e.g. strings of consonants) (Vigneau et al., 2005). These findings support the idea that the reading task introduced in experiment 3 might have been responsible for eliciting responses in OTS-words for readable emoji stimuli.

However, which task precisely is required to activate responses in OTS-words is not without debate. Considering that previous studies (Bookheimer et al., 1995; Moore & Price, 1999; Murtha et al., 1999) demonstrated that OTS-words is also activated during object naming, the question arises, if the preference for emojis during a reading task is driven by reading or by silently naming the presented objects. Studies with dyslexic children and adults showed that reading impairments are also correlated with difficulties in naming objects and colors as a consequence of problems in establishing phonological representation of both words and objects (Cantwell & Rubin, 1992; Catts et al., 2002; Katz & Lanzoni, 1997). Price et al. (2006) investigated how object naming and reading aloud differ on a neuronal level and found similar activations for reading written text in comparison to naming pictures of objects in OTS-words. However, both tasks in this case only

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

required the participants to engage with one stimulus at a time, while our reading task required subjects to associate two images with one another and test if they form a meaningful semantic unit. Evidence for differences between reading and object naming is provided by an fMRI study, which required participants to learn to read aloud artificial words and name artificial objects in two sessions (Quinn et al., 2017). During the first session participants were trained to name 18 foreign objects and to read 18 foreign words. During the second session their knowledge of the foreign stimuli was tested and they had to learn a second set of foreign objects and words. Behavioural results showed that while learning profiles remained the same for the first and second set of naming foreign objects, learning efficiency for learning to read the second set of foreign words was increased. fMRI results showed higher activation for the reading task in the posterior ventro-temporal regions compared to the naming task. The authors conclude, that there is a componential and systematic relationship between the stimuli which had to be read, that can be transferred to new stimuli. This kind of relationship is not present for the arbitrary relationship between the visual form of the object and saying its name. Another indicator for the specificity of OTS-words to reading tasks is a study investigating responses of OTS-words in blind participants when reading braille. Though these participants have no sight and read a sensory based scripture system OTS-words shows activation during reading real words, in comparison to pseudo-words or sensory controls (Reich et al., 2011). In sum, we propose that the observed preferences for emojis during reading are unlikely to be solely driven by an underlying object naming task. As participants had to combine two emojis into a semantic unit, our reading task required a higher level of lexical processing than object naming, and we predict that this reading instruction played an important role in eliciting the observed responses in OTS-words.

The observed preference for emojis may be driven by experience

Previous studies showed higher activations in OTS-words for scripts that the participants were trained to decipher compared to unfamiliar scripts (Dehaene et al., 2015; Martin et al., 2019; Szwed et al., 2011), highlighting the importance of training in driving OTS-words activations. A longitudinal fMRI study by Ben-Shachar et al. (2011) showed that the sensitivity of OTS-words for text emerges and increases with formal reading education. This emerging text-sensitivity of OTS-words is also not limited to grapheme based scripture systems, such as the Latin alphabet, as a study with Chinese-Korean bilingual participants showed (Bai et al., 2009). The Korean scripture is visually similar to the Chinese script, but has strong structural similarities to the Latin alphabet. The authors found, that both Korean and Chinese characters engage OTS-words, which is evidence for the cross-cultural importance of OTS-words for reading. This is further illustrated by a meta-analysis comparing a large body of studies investigating the responses in OTS-words across English, Japanese and Chinese (Bolger et al., 2005). Strikingly, OTS-words was found in a consistent location across all the different scripture systems, highlighting its importance for reading a large variety of scripts. All of these studies evaluate responses in OTS-words after extensive training for a given scripture, while the adult participants in our study were not formally trained to read emojis. Nonetheless they are likely to encounter emojis on a daily basis. With the increased usage of social media, written text is more often encountered in a digital format and frequently includes the usage of emojis (Prada et al., 2018). Boutet et al. (2021) have shown that emojis used in digital communication indeed influence emotional communication, social attribution, and overall information processing. One possible explanation for the popularity of emojis is that they might enhance clarity in written communication that often suffers from the lack of necessary non-

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

verbal cues, conveying emotions and attitude (Kaye et al., 2017). As our participant group stems from a demographic that likely encounters emojis daily (Prada et al., 2018), and as emojis are used to convey critical social information (Kaye et al., 2017), expertise for emojis might contribute to the observed preference for these stimuli in OTS-words. Future studies could evaluate participants who have no experience with social media to explicitly test the role of experience with emojis in driving responses in OTS-words. Moreover, while most modern day scripture systems are based on morpheme and phonographic signs, the earliest scripture systems, the Sumerian cuneiform and the Egyptian hieroglyphs, were iconographic (Trigger, 1998). This means that they were composed of pictograms, expressing objects, actions and ideas, similar to modern day emojis. The pictographic roots of our current writing systems might be an explanation why OTS-words responds strongly to emojis, as emojis can be viewed as readable modern pictograms. Hence, instead of interpreting emojis as a novel stimulus, they might actually represent a prior, more fundamental way of written communication (Alshenqeeti, 2016). We can speculate that as emojis are pictographic representations of real-world objects, it might be easier to associate them with their real-world counterpart, but it is still unclear why that would lead to higher activations in OTS-words. One possible explanation could be that reading emojis might require more cognitive resources. Considering that we observed longer response times when participants had to read emojis compared to text, one might argue that reading emojis leads to an increase in cognitive load and would therefore elicit higher responses in OTS-words. However, to fully understand why OTS-words shows a preference for emojis compared to texts, more research is required. Future research can also evaluate if the entire OTS-words subregions prefer emojis over text stimuli or whether there are distinct populations of voxels within these regions that selectively process one stimulus or the other. After all, the fact that we did not observe fMRI adaptation in the mixed condition in Exp 2 and could reliably decode stimulus from both subregions in all experiments may support the latter scenario.

Conclusion

In sum, our study characterized two distinct subregions in the occipitotemporal sulcus: pOTS-words is sensitive to the visual features of the stimuli with a preference for readable emojis, while mOTS-words exhibits both a stimulus and a task sensitivity with a preference for reading emojis. We conclude that both sub-regions are likely to be needed for fluent reading, and that they are flexibly recruited whenever semantic meaning has to be assigned to readable visual stimuli. This study enhances our understanding of the role of the OTS-words subregions and suggests that both stimulus features and task demands should be considered when investigating these regions.

Acknowledgements

This research was supported by the National Eye Institute (R01EY033835), by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation – project number 222641018 – SFB/TRR 135 TP C10), as well as by "The Adaptive Mind", funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art.

References

Alshenqeeti, H. (2016). Are Emojis Creating a New or Old Visual Language for New Generations? A Socio-semiotic Study. *Advances in Language and Literary Studies*, *7*(6). https://doi.org/10.7575/aiac.alls.v.7n.6p.56

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ardila, A., Bertolucci, P. H., Braga, L. W., Castro-Caldas, A., Judd, T., Kosmidis, M. H., Matute, E., Nitrini, R., Ostrosky-Solis, F., & Rosselli, M. (2010). Illiteracy: The neuropsychology of cognition without reading. *Archives of Clinical Neuropsychology*, *25*(8), 689–712. https://doi.org/10.1093/arclin/acq079

Bai, F., Watson, D. R., Yu, H., Shi, Y., Yuan, Y., & Zhang, Z. (2009). Abnormal resting-state functional connectivity of posterior cingulate cortex in amnestic type mild cognitive impairment. *Brain Research*, *1302*, 167–174.

Ben-shachar, M., Dougherty, R. F., Deutsch, G. K., & Wandell, B. A. (2011). *The Development of Cortical Sensitivity to Visual Word Forms*. 2387–2399.

Ben-Shachar, M., Dougherty, R. F., Deutsch, G. K., & Wandell, B. A. (2007). Differential sensitivity to words and shapes in ventral occipito-temporal cortex. *Cerebral Cortex*, *17*(7), 1604–1611. https://doi.org/10.1093/cercor/bhl071

Bolger, D. J., Perfetti, C. A., & Schneider, W. (2005). Cross-cultural effect on the brain revisited: Universal structures plus writing system variation. *Human Brain Mapping*, *25*(1), 92–104. https://doi.org/10.1002/hbm.20124

Bookheimer, S. Y., Zeffiro, T. A., Blaxton, T., Gaillard, W., & Theodore, W. (1995). Regional cerebral blood flow during object naming and word reading. *Human Brain Mapping*, *3*(2), 93–106. https://doi.org/10.1002/hbm.460030206

Boutet, I., LeBlanc, M., Chamberland, J. A., & Collin, C. A. (2021). Emojis influence emotional communication, social attributions, and information processing. *Computers in Human Behavior*, *119*(July 2020), 106722. https://doi.org/10.1016/j.chb.2021.106722

Caffarra, S., Karipidis, I. I., Yablonski, M., & Yeatman, J. D. (2021). Anatomy and physiology of word-selective visual cortex: from visual features to lexical processing. *Brain Structure and Function*, *226*(9), 3051–3065. https://doi.org/10.1007/s00429-021-02384-8

Cantwell, A., & Rubin, H. (1992). Object naming ability of adults with written language difficulties. *Annals of Dyslexia*, *42*(1), 179–195. https://doi.org/10.1007/BF02654945

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*, *18*(2), 90–98. https://doi.org/10.1016/j.tics.2013.11.005

Catts, H. W., Gillispie, M., Leonard, L. B., Kail, R. V., & Miller, C. A. (2002). The role of speed of processing, rapid naming, and phonological awareness in reading achievement. *Journal of Learning Disabilities*, *35*(6), 509–524. https://doi.org/10.1177/00222194020350060301

Cho, Y. I., Lee, S. Y. D., Arozullah, A. M., & Crittenden, K. S. (2008). Effects of health literacy on health status and health service utilization amongst the elderly. *Social Science and Medicine*, *66*(8), 1809–1816. https://doi.org/10.1016/j.socscimed.2008.01.003

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

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.-A., & Michel, F. (2000a). The visual word form area. *Brain*, *123*(2), 291–307. https://doi.org/10.1093/brain/123.2.291

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., & Michel, F. (2000b). The visual word form area. Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*. https://doi.org/10.1093/brain/123.2.291

Cohen, L., Martinaud, O., Lemer, C., Lehéricy, S., Samson, Y., Obadia, M., Slachevsky, A., & Dehaene, S. (2003). Visual Word Recognition in the Left and Right Hemispheres: Anatomical and Functional Correlates of Peripheral Alexias. *Cerebral Cortex*, *13*(12), 1313–1333. https://doi.org/10.1093/cercor/bhg079

Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

acquisition. In *PLoS Biology* (Vol. 16, Issue 3). https://doi.org/10.1371/journal.pbio.2004103

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

Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience*, *16*(4), 234–244. https://doi.org/10.1038/nrn3924

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: a proposal. *Trends in Cognitive Sciences*, *9*(7), 335–341.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science*, *330*(6009), 1359–1364. https://doi.org/10.1126/science.1194140

Gaillard, R., Naccache, L., Pinel, P., Clémenceau, S., Volle, E., Hasboun, D., Dupont, S., Baulac, M., Dehaene, S., Adam, C., & Cohen, L. (2006). Direct Intracranial, fMRI, and Lesion Evidence for the Causal Role of Left Inferotemporal Cortex in Reading. *Neuron*, *50*(2), 191–204. https://doi.org/10.1016/j.neuron.2006.03.031

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536–548. https://doi.org/10.1038/nrn3747

Grotheer, M., Jeska, B., & Grill-Spector, K. (2018). A preference for mathematical processing outweighs the selectivity for Arabic numbers in the inferior temporal gyrus. *NeuroImage*, *175*(November 2017), 188–200. https://doi.org/10.1016/j.neuroimage.2018.03.064

Grotheer, M., Zhen, Z., Lerma-Usabiaga, G., & Grill-Spector, K. (2019). Separate lanes for adding and reading in the white matter highways of the human brain. *Nature Communications*, *10*(1), 420216. https://doi.org/10.1038/s41467-019-11424-1

Katz, R. B., & Lanzoni, S. M. (1997). Activation of the phonological lexicon for reading and object naming in deep dyslexia. *Brain and Language*, *58*(1), 46–60. https://doi.org/10.1006/brln.1997.1766

Kaye, L. K., Malone, S. A., & Wall, H. J. (2017). Emojis: Insights, Affordances, and Possibilities for Psychological Science. *Trends in Cognitive Sciences*, *21*(2), 66–68. https://doi.org/10.1016/j.tics.2016.10.007

Keller, A. S., Jagadeesh, A. V., Bugatus, L., Williams, L. M., & Grill-Spector, K. (2022). Attention enhances category representations across the brain with strengthened residual correlations to ventral temporal cortex. *NeuroImage*, *249*, 1–35. https://doi.org/10.1016/j.neuroimage.2022.118900

Kubota, E., Grill-spector, K., & Nordt, M. (2023). Trends in Cognitive Sciences Rethinking cortical recycling in ventral temporal cortex. *Trends in Cognitive Sciences*, 1–10. https://doi.org/10.1016/j.tics.2023.09.006

Kubota, E., Mareike, G., Dawn, F., S, N. V., Jesse, G., & Kalanit, G. (2023). *White matter connections of high-level visual areas predict cytoarchitecture better than category-selectivity in childhood , but not adulthood. June 2022*, 2485–2506.

Lafer-Sousa, R., Conway, B. R., & Kanwisher, N. G. (2016). Color-biased regions of theventral visual pathway lie between face-andplace-selective regions inhumans,as in macaques. *Journal of Neuroscience*, *36*(5), 1682–1697. https://doi.org/10.1523/JNEUROSCI.3164-15.2016

Lerma-Usabiaga, G., Carreiras, M., & Paz-Alonso, P. M. (2018). Converging evidence for functional and structural segregation within the left ventral occipitotemporal cortex in reading. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(42), E9981–E9990. https://doi.org/10.1073/pnas.1803003115

Martin, L., Durisko, C., Moore, M. W., Coutanche, M. N., Chen, D., & Fiez, J. A. (2019). The VWFA is the home of orthographic learning when houses are used as letters. *ENeuro*, *6*(1),

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1–13. https://doi.org/10.1523/ENEURO.0425-17.2019

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, *7*(7), 293–299. https://doi.org/10.1016/S1364-6613(03)00134-7

Mei, L., Xue, G., Chen, C., Xue, F., Zhang, M., & Dong, Q. (2010). The "visual word form area" is involved in successful memory encoding of both words and faces. *NeuroImage*, *52*(1), 371–378. https://doi.org/10.1016/j.neuroimage.2010.03.067

Moore, C. J., & Price, C. J. (1999). Three distinct ventral occipitotemporal regions for reading and object naming. *NeuroImage*, *10*(2), 181–192. https://doi.org/10.1006/nimg.1999.0450

Murtha, S., Chertkow, H., Beauregard, M., & Evans, A. (1999). The neural substrate of picture naming. *Journal of Cognitive Neuroscience*, *11*(4), 399–423. https://doi.org/10.1162/089892999563508

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

Prada, M., Rodrigues, D. L., Garrido, M. V., Lopes, D., Cavalheiro, B., & Gaspar, R. (2018). Motives, frequency and attitudes toward emoji and emoticon use. *Telematics and Informatics*, *35*(7), 1925–1934. https://doi.org/10.1016/j.tele.2018.06.005

Price, C., Friston, K., & Bu, C. (1998). A multimodal language region in the ventral visual pathway. *Nature*, *394*(July), 14–17. http://www.nature.com/nature/journal/v394/n6690/abs/394274a0.html

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

Price, C. J., & Friston, K. J. (1997). Cognitive conjunction: A new approach to brain activation experiments. *NeuroImage*, *5*(4 I), 261–270. https://doi.org/10.1006/nimg.1997.0269

Price, C. J., McCrory, E., Noppeney, U., Mechelli, A., Moore, C. J., Biggio, N., & Devlin, J. T. (2006). How reading differs from object naming at the neuronal level. *NeuroImage*, *29*(2), 643–648. https://doi.org/10.1016/j.neuroimage.2005.07.044

Qu, J., Pang, Y., Liu, X., Cao, Y., Huang, C., & Mei, L. (2022). Task modulates the orthographic and phonological representations in the bilateral ventral Occipitotemporal cortex. *Brain Imaging and Behavior*, *16*(4), 1695–1707. https://doi.org/10.1007/s11682-022-00641-w

Quinn, C., Taylor, J. S. H., & Davis, M. H. (2017). Learning and retrieving holistic and componential visual-verbal associations in reading and object naming. *Neuropsychologia*, *98*, 68–84. https://doi.org/10.1016/j.neuropsychologia.2016.09.025

Reich, L., Szwed, M., Cohen, L., & Amedi, A. (2011). A ventral visual stream reading center independent of visual experience. *Current Biology*, *21*(5), 363–368. https://doi.org/10.1016/j.cub.2011.01.040

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

Sandak, R., Mencl, W. E., Frost, S. J., & Pugh, K. R. (2018). The neurobiological basis of skilled and impaired reading: Recent findings and new directions. *The Cognitive Neuroscience of Reading: A Special Issue of Scientific Studies of Reading*, *8438*, 273–292. https://doi.org/10.4324/9780203764442-6

Stigliani, A., Weiner, K. S., & Grill-Spector, K. (2015). Temporal processing capacity in high-level visual cortex is domain specific. *Journal of Neuroscience*, *35*(36), 12412–12424. https://doi.org/10.1523/JNEUROSCI.4822-14.2015

Szwed, M., Cohen, L., Qiao, E., & Dehaene, S. (2009). The role of invariant line junctions in object and visual word recognition. *Vision Research*, *49*(7), 718–725. https://doi.org/10.1016/j.visres.2009.01.003

Szwed, M., Dehaene, S., Kleinschmidt, A., Eger, E., Valabrègue, R., Amadon, A., & Cohen, L.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(2011). Specialization for written words over objects in the visual cortex. *NeuroImage*, *56*(1), 330–344. https://doi.org/10.1016/j.neuroimage.2011.01.073

Taylor, J. S. H., Davis, M. H., & Rastle, K. (2019). Mapping visual symbols onto spoken language along the ventral visual stream. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(36), 17723–17728. https://doi.org/10.1073/pnas.1818575116

Thesen, T., McDonald, C. R., Carlson, C., Doyle, W., Cash, S., Sherfey, J., Felsovalyi, O., Girard, H., Barr, W., Devinsky, O., Kuzniecky, R., & Halgren, E. (2012). Sequential then interactive processing of letters and words in the left fusiform gyrus. *Nature Communications*, *3*, 1–8. https://doi.org/10.1038/ncomms2220

Trigger, B. G. (1998). Writing systems: A case study in cultural evolution. *Norwegian Archaeological Review*, *31*(1), 39–62. https://doi.org/10.1080/00293652.1998.9965618

van der Mark, S., Bucher, K., Maurer, U., Schulz, E., Brem, S., Buckelmüller, J., Kronbichler, M., Loenneker, T., Klaver, P., Martin, E., & Brandeis, D. (2009). Children with dyslexia lack multiple specializations along the visual word-form (VWF) system. *NeuroImage*, *47*(4), 1940–1949. https://doi.org/10.1016/j.neuroimage.2009.05.021

Vigneau, M., Jobard, G., Mazoyer, B., & Tzourio-Mazoyer, N. (2005). Word and non-word reading: What role for the Visual Word Form Area? *NeuroImage*, *27*(3), 694–705. https://doi.org/10.1016/j.neuroimage.2005.04.038

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007a). Hierarchical Coding of Letter Strings in the Ventral Stream: Dissecting the Inner Organization of the Visual Word-Form System. *Neuron*, *55*(1), 143–156. https://doi.org/10.1016/j.neuron.2007.05.031

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007b). Hierarchical Coding of Letter Strings in the Ventral Stream: Dissecting the Inner Organization of the Visual Word-Form System. *Neuron*, *55*(1), 143–156. https://doi.org/10.1016/j.neuron.2007.05.031

Vogel, A. C., Petersen, S. E., & Schlaggar, B. L. (2012). The left occipitotemporal cortex does not show preferential activity for words. *Cerebral Cortex*, *22*(12), 2715–2732. https://doi.org/10.1093/cercor/bhr295

Vogel, A. C., Petersen, S. E., & Schlaggar, B. L. (2014). The VWFA: It's not just for words anymore. *Frontiers in Human Neuroscience*, *8*(MAR), 1–10. https://doi.org/10.3389/fnhum.2014.00088

Weiner, K. S., & Grill-Spector, K. (2013). Neural representations of faces and limbs neighbor in human high-level visual cortex: Evidence for a new organization principle. *Psychological Research*, *77*(1), 74–97. https://doi.org/10.1007/s00426-011-0392-x

Weiss, B. D., Hart, G., & Pust, R. E. (1991). The relationship between literacy and health. *Journal of Health Care for the Poor and Underserved*, *1*(4), 351–363. https://doi.org/10.1353/hpu.2010.0294

White, A. L., Palmer, J., & Boynton, G. M. (2018). Evidence of Serial Processing in Visual Word Recognition. *Psychological Science*, *29*(7), 1062–1071. https://doi.org/10.1177/0956797617751898

White, A. L., Palmer, J., Boynton, G. M., & Yeatman, J. D. (2019). Parallel spatial channels converge at a bottleneck in anterior word-selective cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(20), 10087–10096. https://doi.org/10.1073/pnas.1822137116

Woolnough, O., Donos, C., Rollo, P. S., Forseth, K. J., Lakretz, Y., Crone, N. E., Fischer-Baum, S., Dehaene, S., & Tandon, N. (2021). Spatiotemporal dynamics of orthographic and lexical processing in the ventral visual pathway. *Nature Human Behaviour*, *5*(3), 389–398. https://doi.org/10.1038/s41562-020-00982-w

Yablonski, M., Karipidis, I. I., Kubota, E., & Yeatman, J. D. (2023). *The transition from vision to language : distinct patterns of functional connectivity for sub-regions of the visual word form area*. 1–23.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.07.565794; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.