



Parallel spatial channels converge at a bottleneck in anterior word-selective cortex

Alex L. White^{a,b,c,1}, John Palmer^c, Geoffrey M. Boynton^c, and Jason D. Yeatman^{a,b}

^aInstitute for Learning & Brain Sciences, University of Washington, Seattle, WA 98195; ^bDepartment of Speech & Hearing Sciences, University of Washington, Seattle, WA 98195; and ^cDepartment of Psychology, University of Washington, Seattle, WA 98195

Edited by Marlene Behrmann, Carnegie Mellon University, Pittsburgh, PA, and approved March 13, 2019 (received for review December 31, 2018)

In most environments, the visual system is confronted with many relevant objects simultaneously. That is especially true during reading. However, behavioral data demonstrate that a serial bottleneck prevents recognition of more than one word at a time. We used fMRI to investigate how parallel spatial channels of visual processing converge into a serial bottleneck for word recognition. Participants viewed pairs of words presented simultaneously. We found that retinotopic cortex processed the two words in parallel spatial channels, one in each contralateral hemisphere. Responses were higher for attended than for ignored words but were not reduced when attention was divided. We then analyzed two word-selective regions along the occipitotemporal sulcus (OTS) of both hemispheres (subregions of the visual word form area, VWFA). Unlike retinotopic regions, each word-selective region responded to words on both sides of fixation. Nonetheless, a single region in the left hemisphere (posterior OTS) contained spatial channels for both hemifields that were independently modulated by selective attention. Thus, the left posterior VWFA supports parallel processing of multiple words. In contrast, activity in a more anterior word-selective region in the left hemisphere (mid OTS) was consistent with a single channel, showing (i) limited spatial selectivity, (ii) no effect of spatial attention on mean response amplitudes, and (iii) sensitivity to lexical properties of only one attended word. Therefore, the visual system can process two words in parallel up to a late stage in the ventral stream. The transition to a single channel is consistent with the observed bottleneck in behavior.

visual word recognition | visual word form area | spatial attention | divided attention | serial processing

Pages of text are among the most complex and cluttered visual scenes that humans encounter. You cannot immediately comprehend the hundreds of meaningful symbols on this page because of fundamental limits to the brain's processing capacity. How severe are those limits, and what causes them? Studies of eye movements during natural reading have fueled a long debate about whether readers process multiple words in parallel (1, 2). In a direct psychophysical test of the capacity for word recognition, we recently found that participants can report the semantic category of only one of two words that are briefly flashed and masked. Those data demonstrate that a serial bottleneck allows only one word to be fully processed at a time (3). Where is that bottleneck in the brain's reading circuitry?

Early stages of visual processing are spatially parallel. In retinotopic areas of the occipital lobe, receptive fields are small, such that neurons at different cortical locations simultaneously process objects at different visual field locations. This spatial selectivity of neurons in early visual cortex allows spatial attention to prioritize some objects: Activity is enhanced at cortical locations that represent task-relevant compared with irrelevant locations in the visual field (4, 5). During simple feature detection tasks, multiple attended locations can be enhanced in parallel with no cost (6).

It is not clear whether such parallel processing extends into the brain areas responsible for complex object recognition. Ventral occipitotemporal cortex (VOTC) contains a mosaic of regions that each respond selectively to stimuli of a particular category, such as faces, scenes, objects or words (7). Receptive fields in VOTC span much of the visual field, so it is unclear how any one region can

process multiple stimuli of its preferred category (8–10). In the case of word recognition, nearly every neuroimaging study has presented only a single word at a time. Moreover, while there are detailed models of attention in retinotopic cortex, we know relatively little about the function of spatial attention in human VOTC (11–13). Of specific relevance to the present study, there have been no investigations of selective spatial attention in word-selective cortex. The limits of parallel processing and attentional selection have important implications for explaining limits on human behavior, especially during complex tasks such as reading.

We measured fMRI responses while participants performed a semantic categorization task (Fig. 14). On each trial, participants viewed a masked pair of words, one on each side of fixation, and either focused attention on one side (focal cue left or right) or divided attention between both sides (distributed cue). Of particular interest is a specialized VOTC region for word recognition, termed the “visual word form area” (VWFA) (14, 15). Most authors refer to a single left-hemisphere VWFA that was originally proposed to be “invariant” to the visual field position of a word (14, 16–18). More recent fMRI studies, however, demonstrate that VWFA voxels have some spatial tuning but are not organized retinotopically (19–21). Words also activate right-hemisphere VOTC, and there may be a posterior-to-anterior hierarchy within word-selective cortex (17, 22). Recent work has further identified two distinct subregions in the left hemisphere occipitotemporal sulcus (OTS) (7, 23, 24). Compared with the more posterior region, the anterior region is more sensitive to abstract lexical properties and is more strongly connected to language areas (24).

Significance

Because your brain has limited processing capacity, you cannot comprehend the text on this page all at once. In fact, skilled readers cannot even recognize just two words at once. We measured how the visual areas of the brain respond to pairs of words while participants attended to one word or tried to divide attention between both. We discovered that a single word-selective region in left ventral occipitotemporal cortex processes both words in parallel. The parallel streams of information then converge at a bottleneck in an adjacent, more anterior word-selective region. This result reveals the functional significance of subdivisions within the brain's reading circuitry and offers a compelling explanation for a profound limit on human perception.

Author contributions: A.L.W., J.P., G.M.B., and J.D.Y. designed research; A.L.W. performed research; A.L.W. analyzed data; and A.L.W., J.P., G.M.B., and J.D.Y. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: Data and code related to this paper are available at https://github.com/yeatmanlab/White_2019_PNAS and <https://doi.org/10.5281/zenodo.2605455>.

See Commentary on page 9699.

¹To whom correspondence should be addressed. Email: alexlw@uw.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1822137116/-DCSupplemental.

Published online April 8, 2019.

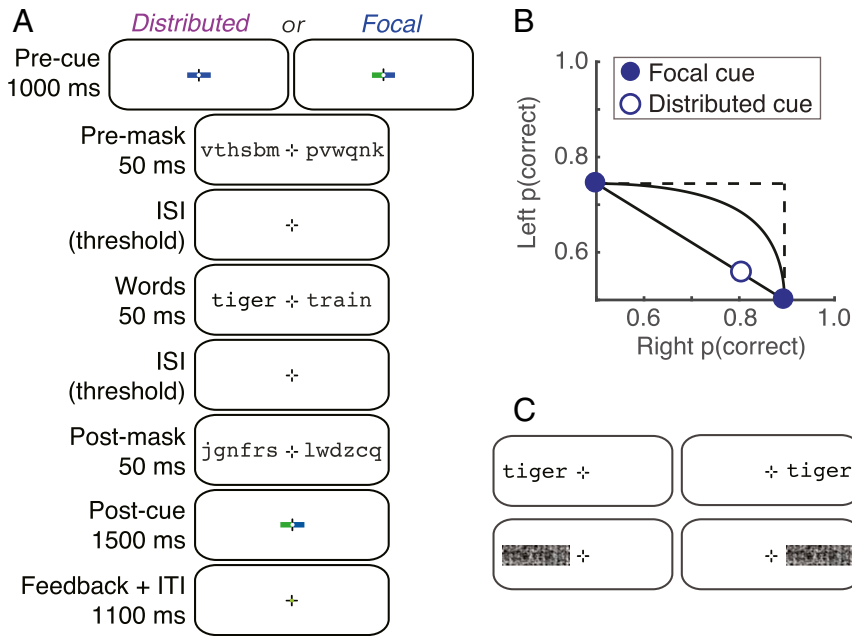


Fig. 1. Stimuli and behavioral performance. (A) Example trial sequence in the main experiment. Half the participants attended to blue cues and half to green cues. Note that there were two additional blank periods with only the fixation mark: a 50-ms gap between the precue and the premask; and a gap between the postmask and the postcue, with duration set to 200 ms minus the sum of the two ISIs between the masks and words. (B) Mean semantic categorization accuracy collected in the scanner, plotted on an AOC. Error bars (± 1 SEM, $n = 15$) are so small that the data points obscure them. (C) Examples of the four stimulus conditions in the localizer scans.

We consider two hypotheses for the bottleneck that prevents people from being able to divide attention and recognize two words simultaneously. The first hypothesis posits a capacity limit at an early stage when the left and right words are processed in separate spatial channels. We define a spatial “channel” as the set of neurons that respond to stimuli at a particular visual field location. Spatial attention serves to amplify the response of task-relevant channels, improving behavioral sensitivity. A severe capacity limit would prevent two channels from being amplified at once, causing a serial switching of attention from side to side across trials. As a result, the mean response when participants try to divide attention across two words would be reduced compared with when they focus on just one word. The opposing model is unlimited-capacity parallel processing, which predicts that responses are not reduced when attention is divided. We previously found such a pattern in retinotopic regions during a simple visual detection task (6). In the present study, we investigate whether the VWFAs contain sufficient spatial tuning to process two words in parallel channels, and how these channels are affected by spatial attention.

The second hypothesis is that the bottleneck lies after an early stage of spatially parallel processing, in a critical brain area that cannot maintain distinct representations of two words at once. This is a different kind of capacity limit: not on how many neurons or channels can be active at once, but on how much information one population of neurons can process. In our task, such a postbottleneck area would respond to only one attended word, regardless of visual field location. It should therefore have three diagnostic response properties: (i) relatively uniform spatial tuning, (ii) no effect of spatial attention on mean BOLD (blood oxygen level-dependent) amplitudes, and (iii) sensitivity to the lexical properties of only one attended word on each trial.

Results

Participants Can Recognize only One Word at a Time. In an fMRI experiment, participants viewed pairs of nouns, one on either side of fixation, which were preceded and followed by masks made of random consonants (Fig. 1A). At the end of each trial, participants were prompted to report the semantic category (living vs. nonliving) of one word. In the focal-cue condition, a precue directed their attention to the side (left or right) of the word they would need to report. In the distributed-cue condition, a precue directed them to divide attention between both words and at the end of the trial they could be asked about the word on

either side. In a training phase, we set the duration of the interstimulus intervals (ISIs) between the words and the masks to each participant’s 80% correct threshold in the focal-cue condition and then maintained that timing (mean = 84 ms) for all conditions. We excluded trials with eye movements away from the fixation mark.

Accuracy was significantly worse in the distributed-cue than in the focal-cue condition: The mean (\pm SE) difference in proportion correct was 0.14 ± 0.01 (95% CI = [0.12 0.16], $t(14) = 13.6$, $P < 10^{-8}$). Accuracy was also significantly worse on the left than on the right side of fixation, both in the focal-cue condition (mean difference = 0.15 ± 0.02 ; 95% CI = [0.11 0.19], $t(14) = 7.60$, $P < 10^{-5}$) and in the distributed-cue condition (0.25 ± 0.02 ; 95% CI = [0.20 0.29], $t(14) = 10.7$, $P < 10^{-7}$). This hemifield asymmetry for word recognition is well established (3, 25).

In Fig. 1B we plot these data on an “attention operating characteristic” (AOC) (26). Accuracy for the left word is plotted against accuracy for the right word. The focal-cue conditions are pinned to their respective axes. The distributed-cue condition is represented by the open symbol. Also shown are the predictions of three models for where that point should fall (26–29). First is unlimited-capacity parallel processing: Two words can be fully processed simultaneously just as well as one. This predicts that the distributed-cue point falls at the intersection of the dashed lines (no deficit), as has been found for simpler visual detection tasks (3, 6). Second is fixed-capacity parallel processing: The brain extracts a fixed amount of information from the whole display per unit time, using processing resources that must be shared between both words. Varying the proportion of resources given to the right word traces out the black curve in the AOC (3). Third is all-or-none serial processing: Words are recognized one at a time, and because of the time constraints imposed by the masking only one word can be processed per trial. Varying the proportion of trials in which the right word is processed traces out the diagonal black line in the AOC.

Mean dual-task accuracy fell significantly below the predictions of both the unlimited- and fixed-capacity parallel models and perfectly on top of the all-or-none serial model’s prediction (Fig. 1B). The average minimum distance from the serial model’s line (calculated such that points below the line have negative distances) was 0.0 ± 0.01 (95% CI = [−0.02 0.03]; $t(14) = 0.10$, $P = 0.93$). In sum, when participants tried to divide attention between the two words, they were able to accurately categorize one word (~80% correct) but were at chance for the other. This

behavior is consistent with the presence of a serial bottleneck at some point in the word recognition system (3).

Selectivity for the Contralateral Hemifield Decreases from Posterior to Anterior Regions. We analyzed BOLD responses in retinotopic visual areas (V1–hV4, VO, and LO) and two word-selective regions in VOTC: VWFA-1 located in the posterior OTS and VWFA-2 located more anterior (Fig. 2A). We use this nomenclature to denote these as subregions of the traditional “visual word form area.” The mean MNI (Montreal Neurological Institute) coordinates of each VWFA subregion are reported in Table 1. In the left hemisphere, all participants had both VWFA-1 and VWFA-2; 14 of 15 participants also had a right hemisphere region symmetric to left VWFA-1, but only 5 of 15 participants had a more anterior right VWFA-2, consistent with prior work (7, 30).

To assess how each region processes pairs of words that are presented simultaneously, we first need to know the region’s sensitivity to single words at the different visual field locations. We analyzed the mean responses to single words presented at either the left or right location in the localizer scan (stimuli in Fig. 1C; results in Fig. 2B). Consistent with the well-known organization of retinotopic cortex, the left hemisphere retinotopic areas responded positively only to words on the right of fixation, and vice versa.

The VFAs, in contrast, are only partially selective for the contralateral hemifield, and that hemifield selectivity decreases from VWFA-1 to VWFA-2 (Fig. 2B). Although all of the VFAs respond positively to words on both sides of fixation, most voxels

still prefer the contralateral side (SI Appendix, Fig. S1). We assessed lateralization with the index: $LI = 1 - R_I/R_C$, where R_I and R_C are the across-voxel mean responses to ipsilateral and contralateral words, respectively (19). Across participants, the mean LI values were 0.46 ± 0.04 in left VWFA-1, 0.36 ± 0.06 in left VWFA-2, 0.52 ± 0.08 in right VWFA-1, and 0.27 ± 0.03 in right VWFA-2. LI differed significantly between VWFA-1 and VWFA-2 [$F(1,45) = 6.71, P = 0.013$]. That difference did not interact with hemisphere [$F(1,45) = 0.92, P = 0.34$] and was present when the left hemisphere was analyzed separately [$F(1,28) = 5.33, P = 0.029$]. Note that the comparison of responses to words at the left and right locations was independent of the (words – scrambled words) contrast used to select the VWFA voxels (Materials and Methods).

In summary, retinotopic areas selectively process words in the contralateral visual field. In contrast, the VFAs of both hemispheres respond to single words at both locations, with a preference for the contralateral side. That preference is weaker in VWFA-2 than in VWFA-1, suggesting more integration across visual space. The magnitude of contralateral preference for single words, however, does not indicate whether either area could process two words at once in the main experiment. For example, right hemisphere VWFA-1 might process the left word while left VWFA-1 processes the right word (similar to right and left V1), or either area could process both words in parallel. We investigated those questions by analyzing data from the main experiment, when two words were present simultaneously and the participant attended to one or both.

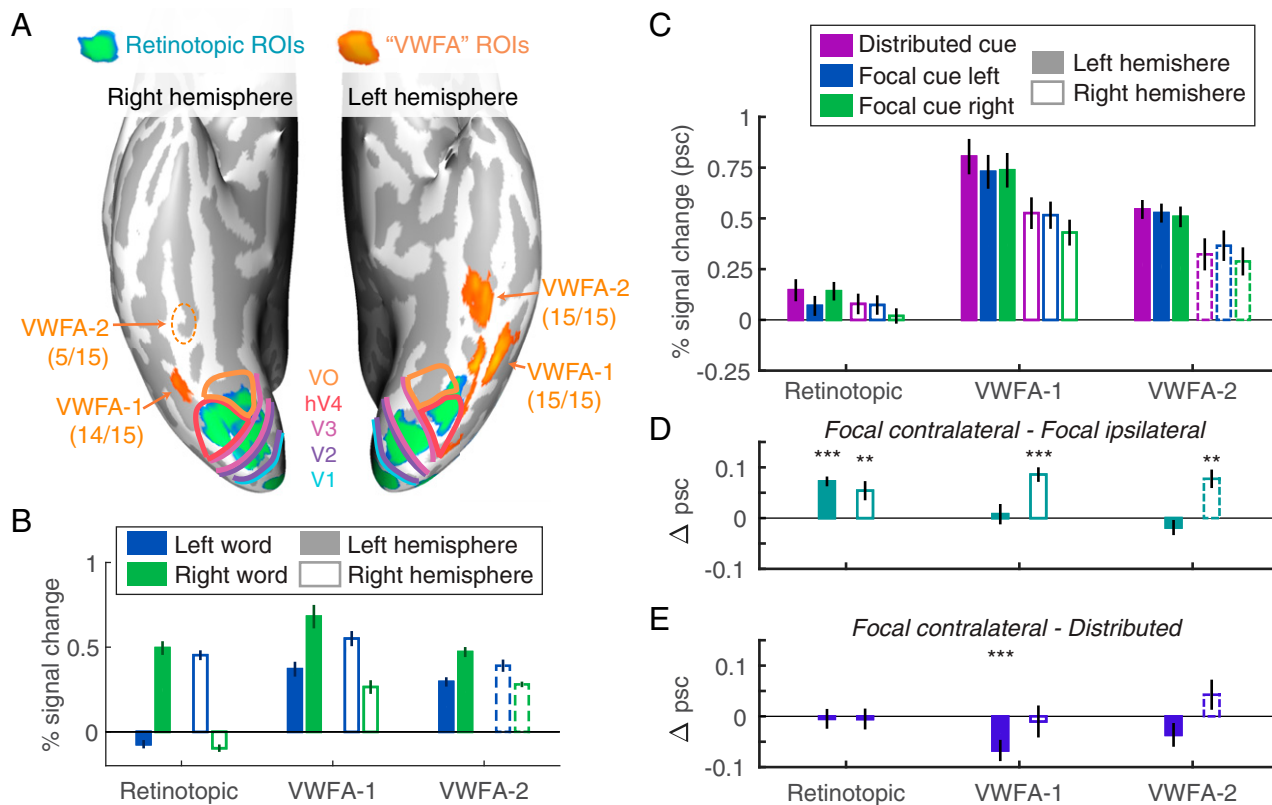


Fig. 2. ROIs and mean BOLD responses. (A) Ventral view of the inflated cortical surfaces of one representative participant. Colored lines are the boundaries between retinotopic areas. The numbers below each VWFA label indicate the number of participants in which that area could be defined. Consistent with previous studies, we find right VWFA-2 in a minority of participants and represent its data with dashed lines to indicate that it is not representative of the average participant. (B) Mean responses to words on the left and right of fixation during the localizer scans. (C) Mean BOLD responses in the main experiment. (D) Mean selective attention effects: differences between responses when the contralateral vs. ipsilateral word was focally cued. (E) Mean divided attention effects: differences between responses when the contralateral word was focally cued vs. both words were cued. Error bars are ± 1 SEM. Asterisks indicate two-tailed P values computed from bootstrapping: $***P < 0.001$; $**P < 0.01$.

Table 1. Across-participant mean (and SD) coordinates of the center of mass of each VWFA region, converted to the MNI-152 atlas

VWFA region	x	y	z
Left VWFA-1	-43.6 (3.4)	-69.1 (5.7)	-12.5 (4.4)
Left VWFA-2	-40.6 (3.0)	-49.1 (5.2)	-20.1 (3.6)
Right VWFA-1	41.2 (2.1)	-64.5 (7.1)	-14.2 (4.9)
Right VWFA-2	38.7 (9.8)	-43.2 (4.7)	-20.9 (2.6)

Left VFAs Respond Strongly During the Categorization Task. Fig. 2C plots the mean BOLD responses in each region of interest (ROI) and cue condition, averaging over all of the retinotopic areas (restricted to the portions that are sensitive to the locations of the words). See *SI Appendix, Fig. S2* for each retinotopic ROI separately. The VFAs responded more strongly to the briefly flashed words than retinotopic regions, and the left hemisphere responded more strongly than the right, especially in the VFAs. A linear mixed-effects model (LME) found reliable effects of region [$F(7,639) = 155.5, P < 10^{-132}$], hemisphere [$F(1,639) = 54.6, P < 10^{-12}$], and cue [$F(2,639) = 9.34, P = 0.0001$]. The effect of hemisphere interacted with region [$F(7,639) = 3.83, P = 0.0004$], but no other interactions were significant ($F_s < 0.25$).

Hemispheric Selective Attention Effects Are Reliable in Retinotopic Cortex but Not in the Left VFAs. The behavioral data demonstrate that participants cannot recognize both words simultaneously. In the focal-cue condition, therefore, the mechanisms of attention must select the relevant word to be processed fully. We first assess the selective attention effect in each region by comparing the mean BOLD responses when the contralateral vs. ipsilateral side was focally cued (Fig. 2D). No prior study has investigated such effects in the VFAs.

An LME model found a main effect of cue [contralateral vs. ipsilateral; $F(1,426) = 12.93, P = 0.0004$] that did not interact with region or hemisphere ($F_s < 0.5$). We also conducted planned comparisons of the focal cue contralateral vs. ipsilateral responses in each ROI (Fig. 2D). The selective attention effect was reliable in the retinotopic ROIs (mean effect: $0.06 \pm 0.01\%$ signal change) and in right hemisphere VWFA-1 ($0.09 \pm 0.01\%$) and VWFA-2 ($0.08 \pm 0.02\%$). However, it was absent in left hemisphere VWFA-1 ($0.01 \pm 0.02\%$) and VWFA-2 ($-0.02 \pm 0.02\%$). We propose two explanations for the lack of effects in the left VFAs: (i) Those areas process both words, so the mean response is a mixture of attended and ignored words, or (ii) they process only one attended word, which makes the mean BOLD response identical in the focal cue left and right conditions. The spatial encoding model described below allows us to discriminate between those possibilities.

Mean BOLD Responses Show No Evidence of Capacity Limits. Assuming that the BOLD response is proportional to the signal-to-noise ratio of the stimulus representation, a capacity limit should cause a reduction in the responses to each word when attention is divided compared with focused. Fig. 2E plots the mean divided attention effects, which are the differences between the focal cue contralateral and the distributed-cue conditions. There was no main effect of cue or interaction with region or hemisphere (all $F_s < 0.5$). Bootstrapping within each ROI found no significant divided attention effect except for an inverse effect (distributed > focal) in left VWFA-1 (mean: $0.07 \pm 0.02\%$ signal change). Therefore, this analysis revealed no evidence of a capacity limit in any area. The data are similar to those in a previous study that found unlimited capacity processing of simple visual features in retinotopic cortex (6).

Left VWFA-1 Contains Two Parallel Channels That Are Modulated by Selective Spatial Attention. Because the VFAs respond to words on both sides of fixation, we cannot isolate the response to each

word within them simply by computing the mean response in the contralateral region (as we can do for retinotopic regions). Instead, we capitalize on differences in spatial tuning across individual voxels and build a “forward encoding model” (31, 32). The model assumes that there are (at least) two spatial “channels” distributed across the region: one for the left word and one for the right word. Each voxel’s response is modeled as a weighted sum of the two channel responses. We estimate the two weights for each voxel as its mean responses to single words on the left and right in the localizer scans. We then “invert” the model via linear regression to estimate the two channel responses in each condition of the main experiment (Fig. 3A). Each channel response is expressed as the proportion of the BOLD response to single words in the localizer scan, in which the participant attended to the fixation mark. Comparing channel responses across cue conditions indexes the effect of spatial attention on voxels that are tuned to specific locations and reveals effects that were obscured by averaging over all voxels in an ROI.

However, not all regions necessarily contain two spatial channels; in fact, we predict that a region after the bottleneck should only have one channel. Therefore, we also fit a simpler

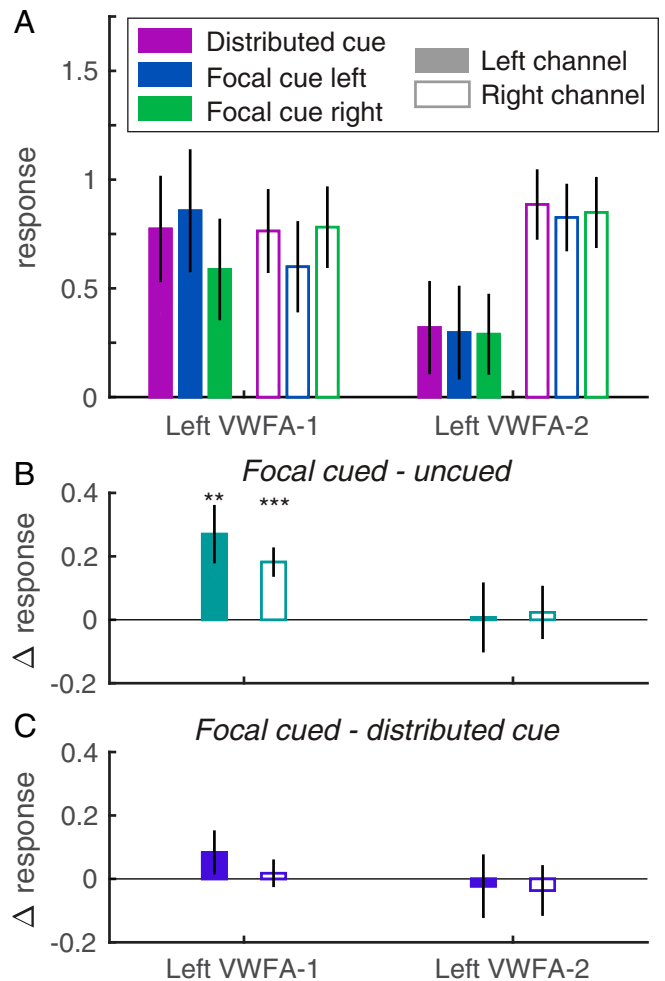


Fig. 3. Estimated left hemisphere channel responses from the spatial encoding model. (A) The left channel is plotted with solid bars and the right channel with open bars; the bar colors indicate precue conditions. (B) Selective attention effects: the differences between each channel’s responses when its visual field location was focally cued vs. uncued. (C) Divided attention effects: the differences between each channel’s response when its location was focally cued vs. when both sides were cued. Error bars are ± 1 SEM. Asterisks indicate two-tailed P values computed from bootstrapping: $***P < 0.001$; $**P < 0.01$. See *SI Appendix, Fig. S4* for right hemisphere data.

one-channel model to each region and compared its fit quality to the two-channel model. In the one-channel model, each voxel is given a single weight: the average of its localizer responses to left and right words. We then model the voxel responses in each condition of the main experiment by scaling those weights by a single channel response.

Adjusting for the number of free parameters, we found that the two-channel model fit significantly better than the one-channel model in left hemisphere VWFA-1: mean adjusted $R^2 = 0.63$ vs. 0.57 ; 95% CI of difference between models = $[0.01\ 0.19]$. In left VWFA-2, the two-channel model fit slightly worse than the one-channel model: 0.40 vs. 0.36 ; 95% CI of difference = $[-0.15\ 0.020]$. In right hemisphere VWFA-1 and VWFA-2, the one-channel model fit significantly better (*SI Appendix, Fig. S3*). We therefore only reject the one-channel model for left hemisphere VWFA-1. Given that not all participants have right hemisphere VFAs, and the one-channel model fit significantly better there, we plot the estimated two-channel responses only for the left hemisphere in Fig. 3. For the right hemisphere, see *SI Appendix, Fig. S4*.

We measured selective attention effects within each ROI as the difference between each channel's responses when its preferred location was focally cued vs. uncued (Fig. 3B). We fit LMEs to assess those cue effects and how responses differed across the left and right channels. In left VWFA-1, there was a main effect of cue [mean: 0.23 ± 0.07 ; $F(1,56) = 12.3$, $P = 0.001$], no main effect of channel [$F(1,56) = 0.01$, $P = 0.94$], and no interaction [$F(1,56) = 2.41$, $P = 0.13$]. The selective attention effect was significant in both channels. The average cued response was 1.38 times the average uncued response. Left VWFA-2 showed a very different pattern, with no significant effects of cue [mean: 0.02 ± 0.09 ; $F(1,56) = 0.03$, $P = 0.87$] or channel [mean right-left difference: 0.55 ± 0.34 ; $F(1,56) = 2.88$, $P = 0.10$] and no interaction [$F(1,56) = 0.09$, $P = 0.77$]. This pattern is consistent with the observation that the one-channel model is adequate for VWFA-2.

Channel responses in the right hemisphere VFAs (*SI Appendix, Fig. S4*) partially matched what was observed in the left hemisphere. Only the left channel within right VWFA-1 had a significant spatial attention effect. More detail is provided in *SI Appendix*, but note that the one-channel model was the best fit for right hemisphere areas, so there is limited value in interpreting those data. In summary, only in left VWFA-1 did we find evidence of two parallel channels, within a single brain region, that could both be independently modulated by selective spatial attention.

Spatial and Attentional Selectivity Are Correlated in VWFA-1. We performed one more test of whether each region supports parallel spatial processing and attentional selection before the bottleneck. If so, the magnitude of the spatial attention effect in each voxel should be related to its spatial selectivity. Consider a voxel that responds equally to single words on the left and right. When two words are presented at once, attending left would affect the voxel response in the same way as attending right. This voxel with no spatial selectivity should therefore have no selective

attention effect. In contrast, a voxel that strongly prefers single words on the right should respond much more in the focal cue right than focal left condition. We tested this prediction by evaluating the linear correlation between two independent measures from individual voxels in separate scans: (i) the difference between responses to single words on the left and right and (ii) the difference between responses in the focal cue left and right conditions. That correlation was significantly positive for all ROIs except left VWFA-2 (*SI Appendix, Fig. S5*). This means that VWFA-1 behaves like other visual areas: The differential spatial tuning of its voxels (and neurons) allows parallel processing of items at different spatial locations and attentional selection of task-relevant items. This also applies to the right hemisphere VFAs, which primarily process the left visual field. The one-channel model nonetheless fit them best because they do not simultaneously represent both the left word and the right word. Only left VWFA-1 appears capable of doing that.

Finally, left VWFA-2 is unique in that attention effects on individual voxels (which average to 0) are not related to their spatial preferences. This result further supports the hypothesis that left VWFA-2 represents a single word after the bottleneck, perhaps in a more abstract format that can be communicated to language regions.

In *SI Appendix* we report two analyses to rule out the hypothesis that the differences between VWFA-1 and VWFA-2 are due to lower signal-to-noise ratio of our measurements in VWFA-2.

Divided Attention Does Not Significantly Reduce VWFA Channel Responses. Finally, we assessed the divided attention effects in each region by comparing the responses of each channel when its location was focally cued vs. when both locations were cued. The divided attention effects for the left hemisphere are plotted in Fig. 3C. There was no effect of cue or interaction between cue and channel (all P s > 0.20). For data from the right hemisphere (where the one-channel model was the best fit) see *SI Appendix, Fig. S4C*.

Neuronal AOCs Assess Capacity Limits. We introduce an analysis of BOLD data: a neuronal AOC (Fig. 4) that can assess capacity limits similarly to the behavioral AOC (Fig. 1B). For a related analysis of EEG data, see ref. 33. In the behavioral AOC, the points pinned to the axes are the focal cue accuracy levels, relative to the origin of 0.5, which is what accuracy would be for an ignored stimulus. Correspondingly, on the neuronal AOC, we plot the differences between responses to attended and ignored words, with the right word on the x axis and the left word on the y axis. The solid points on the axes are the differences in response to each stimulus when it was focally cued vs. uncued. The single open point represents the distributed-cue condition: Its x value is the difference in right word response between the distributed cue and focal uncued (i.e., focal cue left) conditions. Similarly, the y value is the difference in left word response between the distributed cue and focal uncued conditions.

We compare that distributed-cue point to the predictions of two models. First is unlimited capacity parallel processing: As in the behavioral AOC, this model predicts that the distributed-cue

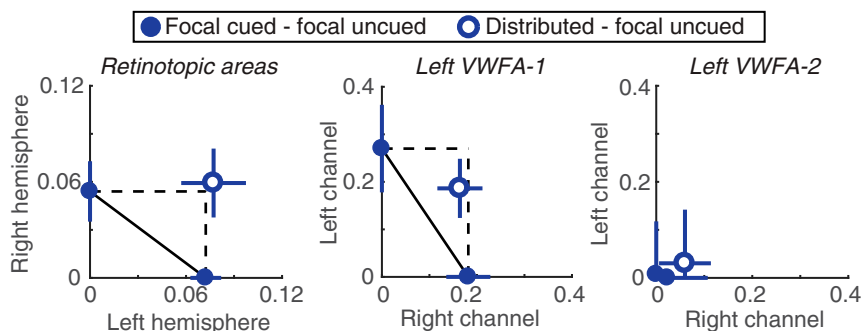


Fig. 4. Neuronal AOCs. (Left) BOLD responses averaged across all retinotopic ROIs and all participants. The left hemisphere, which processes the right word, is plotted on the horizontal axis, and the right hemisphere is on the vertical axis. (Middle and Right) Channel responses in left hemisphere VWFA-1 and VWFA-2, respectively. The right channel is on the horizontal axis and the left channel is on the vertical axis.

point falls on the intersection of the two dashed lines. That indicates no change in response magnitudes relative to when each word was focally cued. Second is serial switching of attention: The behavioral data suggest that on each distributed-cue trial participants recognize one word but not the other. It is as if they pick one side to attend to fully and switch sides sporadically from trial to trial. This model predicts that their average brain state should be a linear mixture of the focal-cue left state and the focal-cue right state. That prediction corresponds to the diagonal line connecting the two focal-cue points.

Fig. 4 contains the AOCs constructed from data averaged over all 15 participants. For retinotopic cortex, the mean data were clearly consistent with the unlimited-capacity model, as the distributed cue point fell just above the dashed intersection. We also assessed the distribution of AOC points across individual participants. The individual data in retinotopic cortex were limited by noise, as these areas were hardly responding above baseline (Fig. 2C). We could not construct the AOC for 4 of the 15 participants because they lacked a positive selective attention effect in at least one hemisphere, which put the whole AOC below one axis and rendered it uninterpretable. Among the remaining 11 participants, the mean distance from the nearest point on the serial switching line (calculated such that points below the line are negative) was 0.05 ± 0.03 (95% CI = $[-0.02, 0.09]$). We also computed the distance of each participant's distributed-cue point from the unlimited-capacity parallel point: mean = 0.0 ± 0.03 (95% CI = $[-0.07, 0.04]$). That distance was negative if, averaged across hemispheres, distributed-cue responses were less than focal-cued responses. In sum, we cannot definitively rule out the serial-switching model for retinotopic cortex, but average responses were more consistent with the unlimited-capacity parallel model (Fig. 4).

The AOC for channel responses in left VWFA-1, averaged over all 15 participants, is shown in Fig. 4, *Middle*. We were also able to construct these AOCs for 13 individuals. The mean distance from the nearest point on the serial switching line was 0.08 ± 0.04 . The 95% CI on that distance excluded 0: $[0.02, 0.16]$. The mean distance of the distributed cue point from the unlimited capacity point was -0.11 ± 0.09 and not significantly different from 0 (95% CI: $[-0.27, 0.05]$). In sum, although there was a modest reduction in channel responses when attention was divided, we can reject the serial switching model for left VWFA-1. That result suggests that the computations carried out in left VWFA-1 occur before the serial bottleneck that constrains recognition accuracy.

In left VWFA-2, the AOC collapses to the origin because responses were approximately equal under all cue conditions (Fig. 4, *Right*). This supports the hypothesis that left VWFA-2 responds to just one attended word, regardless of location. Indeed, two spatial channels are not necessary to explain the voxel responses in that region.

Left VWFA-2 Is Sensitive to the Lexical Frequency of only One Attended Word. Previous studies have shown that the VWFA responds more strongly to words that are uncommon in the lexicon (i.e., words with low lexical frequency; ref. 34). In contrast, we expect behavioral accuracy to be worse for low-frequency words. If left VWFA-2 operates after the bottleneck, it should be modulated by the frequency of only one attended word per trial, as should behavioral performance.

We sorted the words into two bins: below and above the median lexical frequency in the stimulus set. (Note that frequency was negatively correlated with word length, but length did not explain the findings below.) As expected, categorization accuracy (d') was better for high- than for low-frequency words. Fig. 5A shows the mean effect of each word's frequency (high bin - low bin) in each cue condition. According to an LME model, the frequency effect on d' was overall larger for the right than for the left word [$F(1,84) = 5.87, P = 0.018$], but that difference between sides interacted with cue condition [$F(2,84) = 6.62, P = 0.002$]. In the focal cue left condition, the left word's frequency had a large effect on d' (mean difference = 0.32 ± 0.09 , bootstrapped $P =$

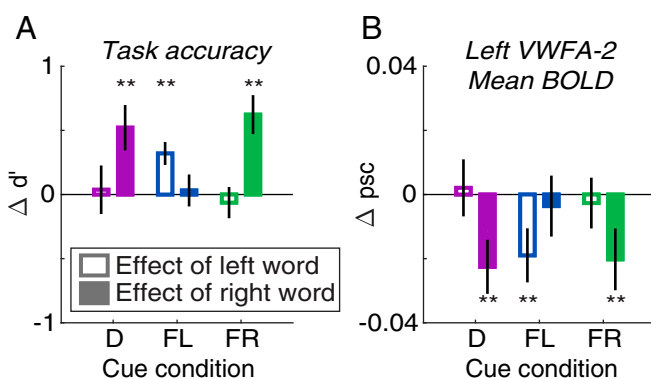


Fig. 5. Effects of lexical frequency. In each panel, the y axis is the mean difference between trials when the word's lexical frequency was above vs. below the median. Unfilled bars are the effect of the left word's frequency (averaging over the right word); filled bars are the effect of the right word's frequency (averaging over the left word). Bars are grouped along the x axis by cue condition: D, distributed; FL, focal left; FR, focal right. (A) Accuracy of semantic categorization judgments (d'). (B) Mean BOLD response in left hemisphere VWFA-2. Asterisks indicate two-tailed P values computed from bootstrapping: $**P < 0.01$.

0.002), and in the focal cue right condition, the right word's frequency had an even bigger effect ($0.62 \pm 0.15, P = 0.002$). In the focal-cue conditions we also computed d' for the uncued word by comparing the participant's categorization report to the semantic category of the uncued word. Those d' levels were near 0 and not affected by the frequencies of the uncued words (right word: 0.03 ± 0.13 ; left word: -0.06 ± 0.12).

In the distributed-cue condition, we examined the effect on d' of whichever word was postcued. There was a large effect of the right word's frequency ($0.52 \pm 0.18, P = 0.002$) but not of the left word's ($0.04 \pm 0.19, P = 0.93$). This indicates that participants strongly favored the right word when they tried to divide attention (consistent with the AOC in Fig. 1B). Together, these analyses show that behavioral responses depended on the frequency of single attended words.

We then performed an exploratory event-related analysis of our fMRI data to examine effects of the lexical frequencies of the two words on each trial. We quantified the effect of the left word as the mean difference between trials when the left word was high- vs. low-frequency, averaging over frequency levels of the right word, and vice versa. We then used LME models to compare those frequency effects across sides (left vs. right) and cue conditions.

In the mean BOLD responses of left VWFA-1, right VWFA-1, and right VWFA-2, we found no significant main effects of frequency or interactions with side or cue condition. However, in left VWFA-2 (Fig. 5B) there was a significant overall effect of frequency [$F(1,84) = 9.97, P = 0.002$] that depended on an interaction of side and cue [$F(2,84) = 3.13, P = 0.049$]. In the focal cue left condition, BOLD responses were lower when the left word was high than low frequency (mean difference in percent signal change = -0.019 ± 0.008 , bootstrapped $P = 0.009$), but there was no effect of the right word's frequency ($-0.004 \pm 0.010, P = 0.73$). The opposite was true in the focal cue right condition (effect of left word: $-0.003 \pm 0.008, P = 0.72$; effect of right word: $-0.020 \pm 0.010, P = 0.019$). In the distributed-cue condition, there was an effect of the right word's frequency ($-0.023 \pm 0.008, P = 0.002$) but not of the left's ($0.002 \pm 0.009, P = 0.84$). That is consistent with the behavioral result that participants favor the right word. Overall, left VWFA-2 is modulated by the frequency of only one attended word on each trial, consistent with processing after the bottleneck. This pattern perfectly mirrors the effects on behavioral accuracy (d' ; Fig. 5A).

We also used the forward encoding model to examine the effect of frequency on each spatial channel's response within left

VWFA-1 (*SI Appendix, Fig. S6*). The effects were weak and mostly isolated to the left channel. This could be because such word-level effects do not fully emerge until VWFA-2, and/or this analysis of individual channels is underpowered.

In summary, the effects of lexical frequency on behavioral sensitivity were matched only by BOLD responses in left VWFA-2. Although in the previous analyses we found no effects of selective attention on mean response magnitudes in VWFA-2, the effect of lexical frequency there was gated by spatial attention. That demonstrates that left VWFA-2 processes single attended words after other words have been filtered out by the bottleneck.

Discussion

A Bottleneck in the Word Recognition Circuit. The primary goal of this study was to determine how the neural architecture of the visual word recognition system forms a bottleneck that prevents skilled readers from recognizing two words at once. Activity in retinotopic cortex matched three criteria for parallel processing before the bottleneck: (i) The two words were processed in parallel spatial channels, one in each cerebral hemisphere, (ii) attended words produced larger responses than ignored words, and (iii) responses were equivalent when attention was divided between two words and focused on one word. These data support unlimited-capacity processing and are summarized in the neuronal AOC (Fig. 4A). We found a similar pattern in a prior study of retinotopic cortex with a simpler, nonlinguistic visual task in which accuracy was the same in the focal and divided attention conditions (6). The fact that there was a severe (completely serial) divided attention cost to accuracy in this semantic categorization task demonstrates that attentional effects in retinotopic cortex do not always predict behavior.

Critically, a word-selective region in the left posterior OTS (VWFA-1) also supported parallel processing before the bottleneck. This single region is not retinotopically organized and responds to words on both sides of fixation. Nonetheless, its individual voxels are spatially tuned to different locations in the visual field (19, 20). Here we demonstrate the functional significance of that tuning: We were able to recover the responses to both simultaneously presented words in parallel spatial channels within left VWFA-1. Those channels were independently modulated by spatial attention, and as shown in the AOC (Fig. 4B) the modest reduction caused by dividing attention was not sufficient to prevent parallel processing.

Finally, a relatively anterior word-selective region in the left hemisphere (VWFA-2 in the mid OTS) had properties consistent with serial processing of single words after the bottleneck. It had weaker and more homogenous spatial selectivity, its mean BOLD response was unaffected by spatial attention, and it could be explained by a model with only one spatial channel. Furthermore, the selectivity of the single channel within left VWFA-2 was revealed in a unique pattern of lexical frequency effects: it was modulated by the frequency of whichever word was selectively attended, and not by ignored words. This pattern mirrors the effects of lexical frequency on task performance, showing nearly perfect attentional selection of single words.

Despite their spatial proximity and similar category selectivity, VWFA-1 and VWFA-2 therefore play distinct roles in the visual word recognition system. Compared to adjacent retinotopic areas and to VWFA-2, left hemisphere VWFA-1 is unique in having intermingled spatial channels covering both visual hemifields.

On the basis of these findings, we propose the following model for how information flows through the word recognition circuit (Fig. 6). Visual signals from the retinas are first projected to contralateral retinotopic areas. Information about the left hemifield in right cortex then crosses over to left VWFA-1, presumably through the posterior corpus callosum. In the transition between VWFA-1 and VWFA-2 (or perhaps within VWFA-2 itself), there is a bottleneck, such that only one word can subsequently reach higher-level language and decision areas. Spatial attention can boost one relevant word before the bottleneck to increase the

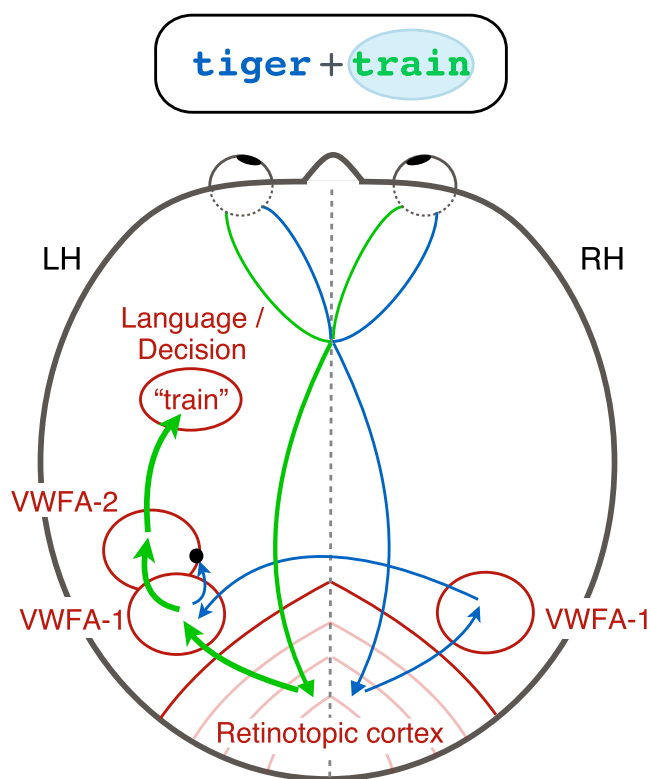


Fig. 6. Circuit diagram of visual processing of two words. This brain is viewed from above, so the left hemisphere is on the left. The bubble around the word on the right side of the display indicates that it is selectively attended, and therefore its representation is relatively enhanced (thicker arrows). A bottleneck (black dot) prevents the unattended word from getting into left VWFA-2.

likelihood that it is fully processed. Our data cannot determine the specific neuronal implementation of the bottleneck, but one possibility is winner-take-all normalization (35). Further research will refine the locus and nature of the bottleneck, which we conclude lies downstream of left VWFA-1.

Hierarchical Processing and White-Matter Connectivity. The differences we found between VWFA-1 and VWFA-2 build on previous models of the visual word recognition system. Several studies have concluded that the anterior portion of word-selective VOTC is more sensitive to higher-level, abstract, lexical properties (17, 22, 24). Another research group studied how VOTC integrates both halves of a single word that are split between hemifields (30, 36). They found that a posterior left VOTC region (the “occipital word form area,” roughly 12 mm posterior to our VWFA-1) represented both halves of a word but maintained them separately. In contrast, a more anterior area (near our VWFA-2) responded to entire words more holistically. That is consistent with our conclusion that left VWFA-1 contains two spatial channels while VWFA-2 contains only a single channel. Strother et al. (30, 36) also found that the right posterior region was biased for the left hemifield, which is consistent with our finding that right VWFA-1 contained a single channel and was especially responsive when attention was focused to the left. Our results go further to show how this circuit responds to pairs of whole words, and to relate the multivoxel patterns to selective attention and task performance.

Finally, our findings align with recently discovered differences in the tissue properties and white matter connections of VWFA subregions (24, 37, 38). Lerma-Usabiaga et al. (24) found that a posterior OTS region was strongly connected through the vertical occipital fasciculus to the intraparietal sulcus, which is implicated

in attentional modulations (39). The MNI position of that region overlaps our VWFA-1. That could explain why spatial attention modulates VWFA-1 but not VWFA-2. A mid-OTS region (near but slightly posterior to our VWFA-2) was found to be connected through the arcuate fasciculus (38) to temporal and frontal language regions. Lerma-Usabiaga et al. (24) postulated that this relatively anterior VWFA “is where the integration between the output from the visual system and the language network takes place.” According to the present results, that output has capacity for only one word.

Hemifield and Hemisphere Asymmetries. Another striking aspect of our data is that participants were much better at categorizing words to the right than left of fixation (3, 25). One potential explanation is the necessity of word-selective regions in the left hemisphere, which respond more strongly to words in the right than in left hemifield (Fig. 2B). VWFA-1 in the right hemisphere may help process letter strings in the left hemifield (19, 20, 30). However, we found that the left hemisphere has three advantages: (i) There were more roughly three times as many voxels in left than right VWFA-1, (ii) left VWFA-1 has two parallel channels, one for each hemifield, and (iii) only one-third of participants had a right VWFA-2, but all had a left VWFA-2. Left VWFA-2 may contain the single channel through which all words must pass on the way to left-hemisphere language regions. That means that the right hemisphere cannot fully process a word in the left hemifield while the left hemisphere processes another word in the right hemifield. Moreover, single words presented to the right hemifield evoke stronger and faster responses in the left hemisphere VWFA than words presented to the left hemifield (19). Therefore, in our paradigm, the right word may automatically win a competitive normalization in left VWFA-2, blocking the left word at the bottleneck. Such a pattern has been observed in macaque face-selective brain regions: The contralateral face in a simultaneous pair dominates the neural response (10). If left VWFA-2 behaves similarly, it could explain why accuracy for the left word is barely above chance in the distributed-cue condition. Focal attention can shift the bias in that competition in favor of the left word, but only partially.

Limitations and Further Questions. There are some limits to our interpretations. First, we were limited by the spatial resolution of our $3 \times 3 \times 3$ -mm functional voxels. We found little evidence for multiple spatially tuned channels in the left VWFA-2, but it is possible that it contains subpopulations of neurons with different spatial tuning that are more evenly intermingled than in VWFA-1. Second, it is possible that two words are in fact represented separately in left VWFA-2, but in channels that are not spatially tuned. It is difficult to imagine how such an architecture would avoid interference between the two words, given that the participant must judge them independently and location is all that differentiates them. Another possibility is that the core mechanism of the bottleneck lies in downstream language areas that influence BOLD activity in VWFA-2 via feedback connections. In any case, all of these speculative hypotheses are consistent with our primary finding that left VWFA-1 supports parallel processing before the bottleneck.

Regarding left VWFA-1, the two-channel spatial encoding model may seem to imply that some neurons in that region are totally selective for the left visual field location, and others for the right. That is uncertain; indeed, 92% of voxels in left VWFA-1 responded more strongly to the right location than the left (SI Appendix, Fig. S1). That is consistent with recent population receptive field measurements in VOTC (20). However, given that the two-channel model fit best, we suppose that many voxels contain some neurons with receptive fields shifted far enough to the left that, when the participant attends to the left, they are up-regulated so left word is represented most strongly. The correlation between spatial preference and selective attention effects supports this interpretation (SI Appendix, Fig. S5).

The conclusion that retinotopic areas had no capacity limit rests on an assumption that is common in the literature but deserves further scrutiny. Specifically, it assumes that the magnitude of the BOLD response is proportional to the signal-to-noise ratio of the stimulus representation used to make the judgment (40, 41). We found that responses in the distributed-cue condition were not lower than in the focally cued condition. We are aware of no prior results that could specifically explain that lack of divided attention effect, but some studies suggest that the total BOLD signal is a mixture of factors related to the stimulus, the percept, attention, anticipation, arousal, and perhaps other factors time-locked to the task (42–45). In principle, the total BOLD response in the distributed-cue condition could have been elevated by factors related to the increased task difficulty while the strength of the stimulus representation was actually lowered. However, no such factors could explain the selective attention effects, which are measured in trials with only focal cues. Our core conclusions hold even if we exclude the distributed-cue condition.

A final caveat is that the task we used differs markedly from natural reading. Specifically, participants fixated between two unrelated nouns and judged them independently. This study sets important boundary conditions for the limits of parallel processing of two words, but future work should attempt to generalize our model to conditions more similar to natural reading.

Conclusion

The experiment reported here advances our understanding of the brain’s reading circuitry by mapping out the limits of spatially parallel processing and attentional selection. Surprisingly, parallel processing of multiple words extends from bilateral retinotopic cortex into the posterior word-selective region (VWFA-1) in the left hemisphere. We propose that signals from the two hemifields then converge at a bottleneck such that only one word is represented in the more anterior VWFA-2. An important question for future work is whether similar circuitry applies to other image categories. Faces and scenes, for instance, are also processed by multiple category-selective regions arranged along the posterior-to-anterior axis in VOTC (7). Recognition for each category might rely on similar computational principles to funnel signals from across the retina into a bottleneck, or written words might be unique due to their connection to spoken language.

Materials and Methods

Participants. Fifteen volunteers from the university community (eight female) participated in exchange for a fixed payment (\$20/h for behavioral training and \$30/h for MRI scanning). The experimental procedures were approved by the Institutional Review Board at the University of Washington, and all participants gave written and informed consent in accord with the Declaration of Helsinki. All participants were right-handed, had normal or corrected-to-normal visual acuity, and learned English as their first language. All scored above the norm of 100 (mean \pm SEM: 116.6 ± 2.7) on the composite Test of Word Reading Efficiency (46). The sample size was chosen in advance of data collection on the basis of a power analysis (SI Appendix). Three participants had to be excluded and replaced. Two were excluded before any fMRI data collection because they broke fixation on at least 5% of trials during the behavioral training sessions. Another was excluded after one MRI session because he was unable to finish all of the scans and failed to respond on 9% of trials (compared with the mean of 0.7% across all included participants).

Stimuli and Task. Word stimuli were drawn from a set of 246 nouns (SI Appendix, Tables S1 and S2). The nouns were evenly split between two semantic categories: “living” and “nonliving.” The words were four, five, or six letters long, in roughly equal proportions for both categories. To estimate the lexical frequency of each word, we averaged the frequencies listed in two online databases (47, 48). The mean frequencies in the living and nonliving categories were 18.6 and 14.3 per million, respectively. On each trial, one word was selected for each side, with an independent 50% chance that each came from the “living” category. The same word could not be present on both sides simultaneously, and no word could be presented on two successive trials. Masks were strings of six random constants. Words and masks were presented in Courier font. The font size was set to 26 points during training and 50 points in the scanner, so that the size in degrees of

visual angle was constant. The word heights ranged from 0.54° to 0.96° (mean = 0.78°), and their lengths varied from 2.5° to 4.2° (mean = 3.25°). All characters were dark gray on a white background (Weber contrast = -0.85). A fixation mark was present at the screen center throughout each scan: a black cross, 0.43 × 0.43° of visual angle, with a white dot (0.1° diameter) at its center.

Each trial (Fig. 1A) began with a precue for 1,000 ms. The precue consisted of two horizontal line segments, each with one end at the center of the fixation mark and the other end 0.24° to the left or right. In the distributed-cue condition, both lines were the same color, blue or green. In the focal-cue condition, one line was blue and the other green. Each participant was assigned to either green or blue and always attended to the side indicated by that color. After a 50-ms ISI containing only the fixation mark, the premasks appeared for 50 ms. The masks were centered at 2.75° to the left and right of fixation and were followed by an ISI containing only the fixation mark (duration variable across participants; discussed below). Then the two words appeared for 50 ms, centered at the same locations as the masks. The words were followed by a second ISI with the same duration as the first, and then postmasks (different consonant strings) appeared for another 50 ms. After a third ISI, the postcue appeared. This consisted of a green and a blue line, which in the single-task condition matched the precue exactly. The line in the participant's assigned color indicated the side to be judged. The postcue remained visible for 1,500 ms. During that interval the participant could respond by pressing a button (task description below). A 450-ms feedback interval followed the postcue: The central dot on the fixation mark turned green if the participant's response was correct, red if it was incorrect, and black if no response was recorded. Finally, there was a 650-ms intertrial interval with only the fixation mark visible. Each trial lasted a total of 4 s.

The two ISIs between the words and masks had the same duration. That duration was adjusted for each participant during training to yield ~80% correct in the focal-cue conditions and then held constant in all conditions (mean = 84 ± 5 ms). The duration of the third ISI, between the postmasks and the postcue, was set such that the sum of all three ISIs was 200 ms.

The participant's task was to report whether the word on the side indicated by the postcue was a nonliving thing or a living thing. The participants used their left hand to respond to left words and their right hand to respond to right words. For each hand, there were two buttons, the left of which indicated "nonliving" while the right indicated "living."

In the focal-cue condition, the precue indicated with 100% validity the side to be judged on that trial. In the distributed-cue condition, the precue was uninformative, so the participant had to divide attention and try to recognize both words.

Main Experimental Scans. Each 6-min scan contained nine blocks of seven trials. All trials in each block were of the same precue condition: distributed, focal left, or focal right. During MRI scanning, there were 12-s blanks after each block, during which the participant maintained fixation on the cross. During behavioral training sessions, those blanks were shortened to 4 s. During the last 2 s of each blank, the precue for the upcoming trial was displayed with thicker lines.

Localizer Scans. We used localizer scans to define ROIs, presenting two types of stimuli one at a time at the same locations as the words in the main experiment (Fig. 1C). Each 3.4-min localizer scan consisted of 48 4-s blocks, plus 4 s of blank at the beginning and 8 s of blank at the end. Every third block was a blank, with only the fixation mark present. In each of the remaining blocks, a rapid sequence of eight stimuli was flashed at 2 Hz (400 ms on, 100 ms off). Each block contained one of two types of stimuli: words or phase-scrambled word images, all either to the left or right of fixation (center eccentricity 2.75°). Therefore, there were four types of stimulus blocks. Each scan contained eight of each in a random order.

The words were drawn from the same set as in the main experiment, in the same font and size, but with 100% contrast. We created phase-scrambled images by taking the Fourier transform of each word, replacing the phases with random values, and inverting the Fourier transform. Each image was matched in size, luminance, spatial frequency distribution, and rms contrast to the original word.

The participant's task was to fixate centrally and press a button any time the black cross briefly became brighter. Those luminance increments occurred at pseudorandom times: The intervals between them were drawn from an exponential distribution with mean 4.5 s, plus 3 s, and clipped at a maximum of 13 s. Hits were responses recorded within 1 s after luminance increment; false alarms were responses more than 1 s after the most recent increment. An adaptive staircase (49) adjusted the magnitude of the luminance increments to keep the task mildly difficult (maximum reduced hit rate = 0.8).

MRI Data Analysis. We performed all analyses in individual brains, averaging only the final parameter estimates extracted from each individual's ROIs.

After standard initial preprocessing (*SI Appendix*), we analyzed functional scans (combining across sessions) with the glmDenoise package in MATLAB (50). The glmDenoise algorithm fits a general linear model (GLM) to the task blocks and includes noise regressors estimated from voxels that were uncorrelated with the experimental protocol. To analyze the effects of lexical frequency of the two words on each trial, we performed an additional event-related GLM. It included predictors for each trial type defined by the combination of high- or low-frequency words (e.g., left low – right low, left low – right high, etc.) in each cue condition.

ROIs in retinotopic areas (V1–V4, VO, and LO) were defined from the localizer data by contrasting responses to scrambled words on the left minus scrambled words on the right. We defined each ROI as the intersection of voxels within that retinotopic area and the voxels that responded more to the contralateral stimuli at a conservative threshold of $P < 10^{-6}$.

The VWFA ROIs were defined by the contrast of words – scrambled words, regardless of side, with the false discovery rate $q < 0.01$. Voxels in all retinotopic regions were excluded from the VWFAs. Consistent with the emerging view that the visual word recognition system contains two separate regions in VOTC (at least in the left hemisphere), we separately defined VWFA-1 and VWFA-2 for each participant and hemisphere. In both hemispheres, VWFA-1 was anterior to area V4, often lateral to area VO, near the posterior end of the OTS. VWFA-2 was a second patch anterior to VWFA-1. In the left hemisphere, VWFA-2 was always anterior and/or lateral to the anterior tip of the midfusiform sulcus. Left VWFA-2 was usually also in the OTS, although in four cases it appeared slightly more medial, encroaching on the lateral fusiform gyrus. In a few cases, VWFA-1 and VWFA-2 were contiguous with each other at the chosen statistical threshold for the words – scrambled contrast. However, raising the threshold always revealed separate peaks, and VWFA-1 and VWFA-2 were defined to be centered around those peaks.

In the right hemisphere, there were fewer word-selective voxels that met our statistical threshold (Table 2). We found VWFA-1 in 14 of 15 participants, and VWFA-2 in 5 of 15 participants. Three of those VWFA-2s were medial of the OTS, on the fusiform gyrus. Previous studies have also reported less word selectivity in the right hemisphere than in the left, and constrained to a single region (7, 30). A representative participant's brain is illustrated in Fig. 2A, and Table 2 lists the numbers of voxels within each ROI. The mean MNI coordinates of the center of mass of each VWFA are listed in Table 1.

We analyzed the main experiment with GLM regressors for each type of block: distributed cue, focal cue left, and focal cue right. Blocks with one or more fixation breaks were flagged with a separate regressor. An average of $5 \pm 1.4\%$ of blocks were excluded in this way.

Spatial Encoding Model. For each VWFA, the two-channel model assumes that there is one "spatial channel" for each word. Left and right channel response strengths are denoted c_L and c_R , respectively. Each voxel i 's response d_i is a weighted sum of the two channel responses:

$$d_i = w_{iL}c_L + w_{iR}c_R + \text{noise}.$$

Weights w_{iL} and w_{iR} describe how strongly the two channels drive each voxel. We estimated those weights as the mean localizer scan responses to single words on the left (which evoke channel responses $c_L = 1$, $c_R = 0$) and single words on the right ($c_L = 0$, $c_R = 1$). That yielded a v -by-2 matrix W of voxel weights, where v is the number of voxels, with one column for each channel. Each condition of the main experiment produced a v -by-1 vector D of voxel responses. The spatial encoding model can then be expressed as

$$D = WC + \text{noise}.$$

Table 2. Mean (SE) numbers of voxels in each ROI

ROI	Left hemisphere	Right hemisphere
V1	68 (5)	68 (7)
V2	82 (8)	90 (10)
V3	129 (15)	126 (14)
V4	77 (9)	90 (9)
VO	45 (7)	50 (10)
LO	61 (15)	82 (16)
VWFA-1	56 (7)	17 (4)
VWFA-2	41 (5)	19 (5)

We were able to define each ROI in all 15 participants except right VWFA-1 (14 of 15 participants) and right VWFA-2 (5 of 15 participants).

Linear regression gives the best-fitting estimate of C , a 2-by-1 vector of channel responses:

$$C_{est} = (W^T W)^{-1} W^T D.$$

We compared that two-channel model to a one-channel model, in which each voxel i was assigned a single weight $w_{i,avg}$ that was the average of its responses to left and right words in the localizer. Then there is a single channel that responds with strength c_{avg} , such that

$$D = W_{avg} c_{avg}.$$

We estimated c_{avg} using linear regression as well. For each model, in each participant and each ROI, we computed the proportion of variance explained, R^2 . We then adjusted each model's R^2 for the number of free parameters, p (i.e., the number of channels):

$$R^2_{adj} = 1 - (1 - R^2) * (v - 1) / (v - p - 1).$$

See [SI Appendix](#) for more detail on the display equipment, eye tracking, MRI data acquisition and preprocessing, retinotopy, procedure, and statistical analyses.

Data and analysis code are publicly available online (51).

ACKNOWLEDGMENTS. We thank an anonymous reviewer for suggesting the lexical frequency analysis and Kevin Weiner for a helpful discussion on VWFA nomenclature. This work was funded by National Eye Institute Grants K99 EY029366 (to A.L.W.), F32 EY026785 (to A.L.W.), and R01 EY12925 (to G.M.B. and J.P.); National Institute of Child Health & Human Development Grant R21 HD092771 (to J.D.Y.); and NSF/U.S.-Israel Science Foundation Behavioral and Cognitive Sciences Grant 1551330 (to J.D.Y.).

- Murray WS, Fischer MH, Tatler BW (2013) Serial and parallel processes in eye movement control: Current controversies and future directions. *Q J Exp Psychol (Hove)* 66: 417–428.
- Reichle ED, Liversedge SP, Pollatsek A, Rayner K (2009) Encoding multiple words simultaneously in reading is implausible. *Trends Cogn Sci* 13:115–119.
- White AL, Palmer J, Boynton GM (2018) Evidence of serial processing in visual word recognition. *Psychol Sci* 29:1062–1071.
- Gandhi SP, Heeger DJ, Boynton GM (1999) Spatial attention affects brain activity in human primary visual cortex. *Proc Natl Acad Sci USA* 96:3314–3319.
- Beck DM, Kastner S (2014) Neural systems for spatial attention in the human brain. *The Oxford Handbook of Attention*, eds Nobre AC, Kastner S (Oxford Univ Press, Oxford), pp 1–43.
- White AL, Runeson E, Palmer J, Ernst ZR, Boynton GM (2017) Evidence for unlimited capacity processing of simple features in visual cortex. *J Vis* 17:19.
- Grill-Spector K, Weiner KS (2014) The functional architecture of the ventral temporal cortex and its role in categorization. *Nat Rev Neurosci* 15:536–548.
- Agam Y, et al. (2010) Robust selectivity to two-object images in human visual cortex. *Curr Biol* 20:872–879.
- Gentile F, Jansma BM (2010) Neural competition through visual similarity in face selection. *Brain Res* 1351:172–184.
- Bao P, Tsao DY (2018) Representation of multiple objects in macaque category-selective areas. *Nat Commun* 9:1774.
- Kay KN, Weiner KS, Grill-Spector K (2015) Attention reduces spatial uncertainty in human ventral temporal cortex. *Curr Biol* 25:595–600.
- Reddy L, Kanwisher NG, VanRullen R (2009) Attention and biased competition in multi-voxel object representations. *Proc Natl Acad Sci USA* 106:21447–21452.
- Zumer JM, Scheeringa R, Schoffelen JM, Norris DG, Jensen O (2014) Occipital alpha activity during stimulus processing gates the information flow to object-selective cortex. *PLoS Biol* 12:e1001965.
- Cohen L, et al. (2000) The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* 123:291–307.
- Dehaene S, Le Clec'h G, Poline J-B, Le Bihan D, Cohen L (2002) The visual word form area: A prelexical representation of visual words in the fusiform gyrus. *Neuroreport* 13:321–325.
- Cohen L, et al. (2002) Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain* 125:1054–1069.
- Dehaene S, et al. (2004) Letter binding and invariant recognition of masked words: Behavioral and neuroimaging evidence. *Psychol Sci* 15:307–313.
- Price CJ, Devlin JT (2011) The interactive account of ventral occipitotemporal contributions to reading. *Trends Cogn Sci* 15:246–253.
- Rauschecker AM, Bowen RF, Parvizi J, Wandell BA (2012) Position sensitivity in the visual word form area. *Proc Natl Acad Sci USA* 109:E1568–E1577.
- Le R, Witthoft N, Ben-Shachar M, Wandell B (2017) The field of view available to the ventral occipito-temporal reading circuitry. *J Vis* 17:6.
- Gomez J, Natu V, Jeska B, Barnett M, Grill-Spector K (2018) Development differentially sculpts receptive fields across early and high-level human visual cortex. *Nat Commun* 9:788.
- Vinckier F, et al. (2007) Hierarchical coding of letter strings in the ventral stream: Dissecting the inner organization of the visual word-form system. *Neuron* 55: 143–156.
- Stigliani A, Weiner KS, Grill-Spector K (2015) Temporal processing capacity in high-level visual cortex is domain specific. *J Neurosci* 35:12412–12424.
- Lerma-Usabiaga G, Carreiras M, Paz-Alonso PM (2018) Converging evidence for functional and structural segregation within the left ventral occipitotemporal cortex in reading. *Proc Natl Acad Sci USA* 115:E9981–E9990.
- Mishkin M, Forgays DG (1952) Word recognition as a function of retinal locus. *J Exp Psychol* 43:43–48.
- Sperling G, Melchner MJ (1978) The attention operating characteristic: Examples from visual search. *Science* 202:315–318.
- Shaw ML (1980) Identifying attentional and decision-making components in information processing. *Attention and Performance VIII*, ed Nickerson R (Routledge, New York), pp 277–296.
- Bonnel A-M, Prinzmetal W (1998) Dividing attention between the color and the shape of objects. *Percept Psychophys* 60:113–124.
- Scharff A, Palmer J, Moore CM (2011) Extending the simultaneous-sequential paradigm to measure perceptual capacity for features and words. *J Exp Psychol Hum Percept Perform* 37:813–833.
- Strother L, Coros AM, Vilis T (2016) Visual cortical representation of whole words and hemifield-split word parts. *J Cogn Neurosci* 28:252–260.
- Brouwer GJ, Heeger DJ (2009) Decoding and reconstructing color from responses in human visual cortex. *J Neurosci* 29:13992–14003.
- Sprague TC, et al. (2018) Inverted encoding models assay population-level stimulus representations, not single-unit neural tuning. *eNeuro* 5:ENEURO.0098-18.2018.
- Mangun GR, Hillyard SA (1990) Allocation of visual attention to spatial locations: Tradeoff functions for event-related brain potentials and detection performance. *Percept Psychophys* 47:532–550.
- Kronbichler M, et al. (2004) The visual word form area and the frequency with which words are encountered: Evidence from a parametric fMRI study. *Neuroimage* 21: 946–953.
- Carandini M, Heeger DJ (2011) Normalization as a canonical neural computation. *Nat Rev Neurosci* 13:51–62.
- Strother L, Zhou Z, Coros AK, Vilis T (2017) An fMRI study of visual hemifield integration and cerebral lateralization. *Neuropsychologia* 100:35–43.
- Yeatman JD, et al. (2014) The vertical occipital fasciculus: A century of controversy resolved by in vivo measurements. *Proc Natl Acad Sci USA* 111:E5214–E5223.
- Weiner KS, Yeatman JD, Wandell BA (2017) The posterior arcuate fasciculus and the vertical occipital fasciculus. *Cortex* 97:274–276.
- Kay KN, Yeatman JD (2017) Bottom-up and top-down computations in word- and face-selective cortex. *eLife* 6:1–29.
- Boynton GM, Demb JB, Glover GH, Heeger DJ (1999) Neuronal basis of contrast discrimination. *Vision Res* 39:257–269.
- Ress D, Backus BT, Heeger DJ (2000) Activity in primary visual cortex predicts performance in a visual detection task. *Nat Neurosci* 3:940–945.
- Cardoso MMB, Sirotnin YB, Lima B, Glushenkova E, Das A (2012) The neuroimaging signal is a linear sum of neurally distinct stimulus- and task-related components. *Nat Neurosci* 15:1298–1306.
- Sirotnin YB, Cardoso M, Lima B, Das A (2012) Spatial homogeneity and task-synchrony of the trial-related hemodynamic signal. *Neuroimage* 59:2783–2797.
- Jack AI, Shulman GL, Snyder AZ, McAvoy M, Corbetta M (2006) Separate modulations of human V1 associated with spatial attention and task structure. *Neuron* 51:135–147.
- Ress D, Heeger DJ (2003) Neuronal correlates of perception in early visual cortex. *Nat Neurosci* 6:414–420.
- Torgesen J, Rashotte C, Wagner R (1999) *TOWRE-2: Test of Word Reading Efficiency* (Pro-Ed, Austin, TX), 2nd Ed.
- Marian V, Bartolotti J, Chabal S, Shook A (2012) CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One* 7:e43230.
- Medler DA, Binder JR (2005) MCWord: An on-Line orthographic database of the English language. Available at www.neuro.mcw.edu/mcword/. Accessed January 7, 2018.
- Kaernbach C (1990) A single-interval adjustment-matrix (SIAM) procedure for unbiased adaptive testing. *J Acoust Soc Am* 88:2645–2655.
- Kay KN, Rokem A, Winawer J, Dougherty RF, Wandell BA (2013) GLMdenoise: A fast, automated technique for denoising task-based fMRI data. *Front Neurosci* 7:247.
- White AL, Palmer J, Boynton GM, Yeatman JD (2019) White_2019_PNAS: Manuscript data and code. Zenodo. Available at <https://doi.org/10.5281/zenodo.2605455>. Deposited March 25, 2019.