PNAS
Proceedings of the National Academy of Sciences of the United States of America www.pnas.org

Visual word processing and experiential origins of functional selectivity in human extrastriate cortex

Chris I. Baker, Jia Liu, Lawrence L. Wald, Kenneth K. Kwong, Thomas Benner, and Nancy Kanwisher

PNAS published online May 14, 2007;
doi:10.1073/pnas.0703300104

This information is current as of May 2007.

Supplementary Material	Supplementary material can be found at: www.pnas.org/cgi/content/full/0703300104/DC1
	This article has been cited by other articles: www.pnas.org#otherarticles
E-mail Alerts	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.
Rights & Permissions	To reproduce this article in part (figures, tables) or in entirety, see: www.pnas.org/misc/rightperm.shtml
Reprints	To order reprints, see: www.pnas.org/misc/reprints.shtml

Notes:

Visual word processing and experiential origins of functional selectivity in human extrastriate cortex

Chris I. Baker*†, Jia Liu‡, Lawrence L. Wald§, Kenneth K. Kwong§, Thomas Benner§, and Nancy Kanwisher†§¶

*Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, 10 Center Drive, Building 10, Room 4C104, Bethesda, MD 20892; ‡State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China; §Massachusetts General Hospital/Massachusetts Institute of Technology/Harvard Medical School Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, 13th Street, Charlestown, MA 02129; and ¶McGovern Institute for Brain Research and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

Contributed by Nancy Kanwisher, April 11, 2007 (sent for review March 13, 2007)

How do category-selective regions arise in human extrastriate cortex? Visually presented words provide an ideal test of the role of experience: Although individuals have extensive experience with visual words, our species has only been reading for a few thousand years, a period not thought to be long enough for natural selection to produce a genetically specified mechanism dedicated to visual word recognition *per se*. Using relatively high-resolution functional magnetic resonance imaging (1.4 × 1.4 × 2-mm voxels), we identified a small region of extrastriate cortex in most participants that responds selectively to both visually presented words and consonant strings, compared with line drawings, digit strings, and Chinese characters. Critically, we show that this pattern of selectivity is dependent on experience with specific orthographies: The same region responds more strongly to Hebrew words in Hebrew readers than in nonreaders of Hebrew. These results indicate that extensive experience with a given visual category can produce strong selectivity for that category in discrete cortical regions.

learning | vision | fMRI | experience | ventral visual pathway

Human extrastriate cortex contains a number of regions that respond selectively to specific categories of visual stimuli (1): the fusiform face area (FFA), which responds selectively to faces (2, 3); the parahippocampal place area, which responds selectively to scenes (4, 5); and the extrastriate body area and fusiform body area, which respond selectively to human bodies and body parts (6–8). Each of these regions can be found in roughly the same anatomical location in most subjects. How do they arise in cortex? Here we test the hypothesis that category-selective regions in the ventral visual pathway can be created through visual experience without a strong genetic predisposition for that specific selectivity.

Visually presented words are an ideal stimulus class for testing this hypothesis because they are (*i*) extensively learned, (*ii*) visually distinctive, and (*iii*) unlikely to have been the focus of evolutionary selection pressures (9). Each of us has had extensive visual experience with written words, yet our species has only been reading for a few thousand years (and literacy has only been prevalent for a few hundred years), which is not thought to be long enough for natural selection to produce a genetically specified brain region dedicated to visual word recognition (10–12). Therefore, if a cortical region is found that responds selectively to visually presented words, compared with other visual stimuli, this finding will serve as an existence of proof that extensive experience can be sufficient to produce selectively responsive regions of visual cortex.

Previous neuroimaging (10, 12–17), intracranial recording (18), magnetoencephalography (19, 20), electroencephalography (EEG) (21, 22), and neuropsychology studies (23–25) in both normal and dyslexic readers (26–29) have identified a region in the left inferior occipitotemporal cortex, often termed the "visual word form area" (VWFA) (16, 17, 30), that may be involved in visual word recognition (for review, see ref. 11). However, two important questions remain, which we addressed here.

First, how specific is the response of this region to visual words? That is, does the VWFA respond more to words than to any other category of visual stimuli? Cohen and Dehaene (16, 17, 30), who coined the term VWFA, define it as the region in the nonretinotopic cortex that responds more to words than checkerboards, leaving the question of specificity open. Many prior studies either fail to address the selectivity of word-responsive regions in detail or have potential confounds, such as differences in attention between word and nonword stimuli. Although selectivity for words compared with stimuli other than checkerboards has been reported [e.g., line drawings (14, 18), geometric symbols (20), and black-and-white pictures of faces and houses (24, 31)], this selectivity could still arise from simple visual feature differences. For example, Hasson *et al.* (14) identified a region responding more to alphanumeric character strings and written words than line drawings of faces, buildings, or tools, but did not test other more visually similar classes of stimuli. Given the large number of low-level features that differ between alphanumeric characters and line drawings, these findings leave the specificity of this region unclear. Polk and Farah (10, 12) identified voxels in some subjects that responded more to visually presented words than to digit strings, but these voxels often showed no difference in response between words and geometric shapes, suggesting that they were not selective for words. Further, a number of studies have reported no evidence for a region of ventral visual cortex selectively responsive to visual words or letters (32–36). Finally, it has been argued that the VWFA is activated in a number of different tasks that do not require visual word form processing (37) and may instead subserve general perceptual processing for any meaningful visual stimulus (38, 39). Thus, no consensus exists in the literature on the selectivity of the putative VWFA.

Second, if strong selectivity for visually presented words in fact exists, does it depend on experience with words? An alternative hypothesis is that an apparently word-selective region might instead reflect a preexisting selectivity for visual features present in word stimuli regardless of experience. To test whether the specific selectivity (if it exists) comes from experience, it is

Author contributions: C.I.B., J.L., and N.K. designed research; C.I.B. and J.L. performed research; L.L.W., K.K.K., and T.B. contributed new reagents/analytic tools; C.I.B., J.L., and N.K. analyzed data; and C.I.B. and N.K. wrote the paper.

The authors declare no conflict of interest.

Abbreviations: cLSSR, candidate letter string-selective region; FFA, fusiform face area; fMRI, functional MRI; VWFA, visual word form area; WvD ROI, words versus drawings region of interest.

†To whom correspondence may be addressed. E-mail: ngk@mit.edu or bakerchris@mail.nih.gov.

This article contains supporting information online at www.pnas.org/cgi/content/full/0703300104/DC1.

© 2007 by The National Academy of Sciences of the USA

NEUROSCIENCE

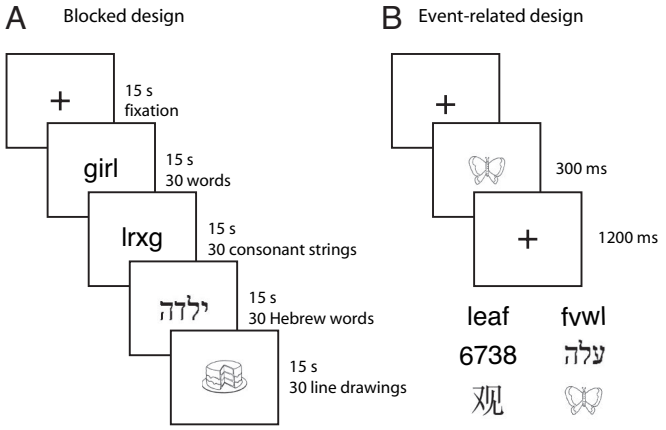


Fig. 1. Experimental design and stimuli. In blocked-design runs (*A*), participants saw 15-sec blocks of English words, consonant strings, Hebrew words, and line drawings interleaved with blocks of fixation only and either fixated passively (half the runs) or responded whenever there was an immediate repeat of the same stimulus. In event-related runs (*B*), participants saw English words, consonant strings, digit strings, Hebrew words, Chinese characters, and line drawings on interleaved trials and reported whether the stimulus was moving downward or to the right.

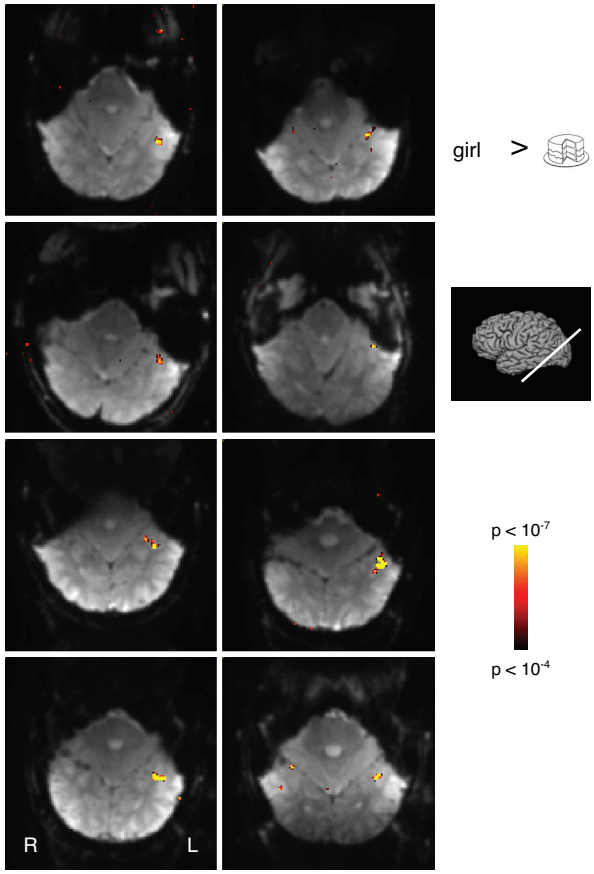


Fig. 2. Regions selectively responsive to English words. Representative functional slices from 8 non-Hebrew readers showing voxels selectively responsive to English words compared with line drawings ($P < 10^{-4}$). Selectively responsive voxels were predominantly found in the left (L) hemisphere close to the occipitotemporal sulcus.

necessary to compare neural responses in people with and without relevant experience.

To answer these two questions of the selectivity and origin of the putative VWFA, we used relatively high-resolution (1.4 × 1.4 × 2 mm) functional MRI (fMRI) to scan both Hebrew and non-Hebrew readers while they viewed English words, consonant strings, Hebrew words, digit strings, Chinese characters, and line drawings. We found a small region in the left hemisphere of most participants selectively responsive to letter strings (responding more to words and consonant strings than to other stimuli). Furthermore, this region responded more strongly to Hebrew words in Hebrew readers than in non-Hebrew readers. Taken together, our data support the hypothesis that extensive experience can produce strong selectivity in regions of the extrastriate cortex.

Results

Word Selectivity in the Ventral Visual Cortex. First we looked for selectivity for English words in English readers (who could not read Hebrew). Our initial screen entailed testing for a higher response to English words than line drawings of objects in a blocked-design experiment (Fig. 1). This criterion represents a necessary but not sufficient condition for a region to be regarded as word-selective. In 20 of 23 participants, we identified clusters of voxels outside of the posterior occipital cortex in the region of the occipitotemporal sulcus and fusiform gyrus that responded more to English words than to line drawings (Fig. 2). We refer to the candidate word-selective cluster in each participant as the words versus drawings region of interest (WvD ROI). The location of the WvD ROI in our participants is consistent with the location of the putative VWFA reported previously (14, 16, 17, 30). Voxels with a greater response to words than line drawings were also sometimes observed in the occipital cortex, probably reflecting differences in the retinotopic spatial envelope of line drawings and words (14), and on the lateral surface of the brain close to the superior temporal sulcus, likely corresponding to regions previously identified as involved in phonological processing (33, 40). However, coverage of the occipital and superior temporal regions was limited by the restricted slice prescription available during high-resolution scanning, and data from these regions were not obtained in all participants. Here we focused on clusters of apparently word-selective voxels in the ventral visual cortex.

In 18 of 20 participants, a WvD ROI was found only in the left hemisphere. In two participants, clusters were found in both hemispheres, and we collapsed the data across hemispheres. The WvD ROI averaged 45 voxels in size. For comparison, in 12 participants, we also localized face-selective voxels on the fusiform gyrus (defined by the contrast faces versus objects). The FFA was identified in every participant, and, consistent with previous reports (14), the left FFA was largely adjacent and medial to the WvD ROI. The FFA averaged 485 voxels bilaterally (right, 315 voxels; left, 170 voxels) and was significantly larger than the WvD ROI in the same participants [$t(11) = 5.0$, $P < 0.0001$].

As noted previously, a higher response to English words than line drawings represents a necessary but not sufficient criterion for a word-selective region. To provide a stronger test of the selectivity of the WvD ROI in more detail, we measured the response in this region in separate event-related runs in 11 participants viewing English words, consonant strings, Hebrew words, digit strings, Chinese characters, and line drawings [Fig. 3*A* and supporting information (SI) Fig. 5]. Note that the resulting response magnitudes are not subject to a selection bias (41) because the data used to define the region were independent from the data used to quantify the response. To characterize the selectivity in this region, we performed a series of planned comparisons contrasting the blood oxygen level-dependent response in each of the new stimulus classes to that

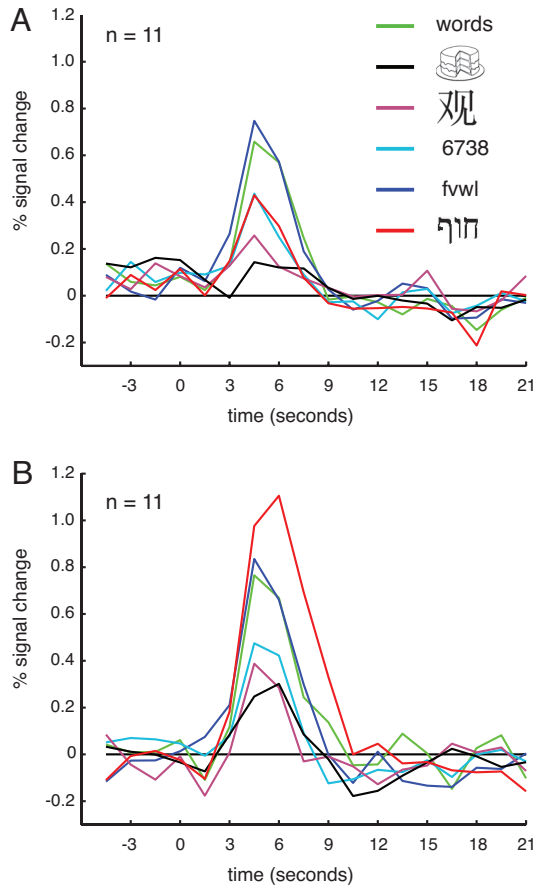


Fig. 3. Estimated hemodynamic response functions for the English words versus line drawings ROI in the event-related runs. In non-Hebrew readers (*A*), the response to English words was significantly greater than to all other stimuli except consonant strings. A similar pattern of response was found in the Hebrew readers (*B*), except that the response to Hebrew words was much greater than that seen in non-Hebrew readers.

elicited by the two benchmark conditions used in the localizer experiment, English words and line drawings. We found a significantly higher response to English words than to each of the other stimulus classes [Hebrew words, $t(10) = 3.933$, $P < 0.003$; digit strings, $t(10) = 2.559$, $P < 0.03$; Chinese characters, $t(10) = 7.655$, $P < 0.0001$] except for consonant strings [$t(10) = 0.517$, $P > 0.6$]. Words differ from all other stimulus classes in terms of pronounceability, but the similarity in response to words and the harder-to-pronounce consonant strings, as well as the difference in response between consonant and digit strings (which are at least as pronounceable as consonant strings), argue against pronounceability as an explanation of the selectivity observed. The similarity in response to English words and consonant strings suggests that this region is better characterized as being selective for letters or letter strings than for words *per se* (see SI Table 1).

Significantly increased responses relative to line drawings were also found for digit strings [$t(10) = 2.711$, $P < 0.022$] and Hebrew words [$t(10) = 5.537$, $P < 0.0001$]. However, there was no significant difference in the response to Chinese characters and line drawings [$t(10) = 1.634$, $P > 0.132$]. Thus, the WvD ROI also exhibits some selectivity for digits strings and Hebrew words even in non-Hebrew readers possibly due to their visual similarity to English words, which is much greater than for Chinese characters (see *SI Text*).

To determine whether any regions show selectivity for English words versus consonant strings, we directly compared the response to English words and consonant strings in the blocked-design experiment. Although we were able to identify scattered voxels in many participants exhibiting a preference for words over consonant strings (54 voxels per participant on average), these were not found in any consistent anatomical location across participants. The most significant voxels were found close to the superior temporal gyrus and likely reflect phonological activation for words over consonant strings (33, 40). In the ventral visual cortex, voxels showing a stronger response to English words than consonant strings could only be identified in 13 of 20 participants in whom we localized a WvD ROI and averaged 17 voxels. However, this selectivity for English words over consonant strings was only replicated in event-related data in one of six participants in whom we had collected both blocked-design and event-related data. Thus, we found no clear evidence for regions of the ventral occipitotemporal cortex showing selectivity for English words compared with consonant strings.

In summary, in the vast majority of participants, we were able to identify a region of the extrastriate cortex, predominantly in the left hemisphere and close to the occipitotemporal sulcus and fusiform gyrus, responding selectively to visual words and consonant strings, even when compared with similar visual stimuli (digit strings and Hebrew words). For ease of reference, we refer to this region as the candidate letter string-selective region (cLSSR).

Role of Experience. The strong response to letter strings in the cLSSR suggests that the selectivity of this region is based on our extensive experience with visual words. However, it remains possible that this selectivity might exist even if we had never learned to read, and that it instead results from some preexisting selectivity for the visual features present in letter string stimuli. The only way to firmly establish the role of experience in generating the observed selectivity is to test people with and without the relevant experience. Therefore, we next compared the response in the cLSSR to Hebrew words in individuals who either did or did not read Hebrew.

Using the same contrast of English words versus line drawings described previously, we identified a cLSSR in 11 of 13 Hebrew readers (who were also proficient English readers) (Fig. 4). In seven subjects these selective clusters were found only in the left hemisphere, in two subjects only in the right hemisphere, and in two subjects in both hemispheres. The average size of the cLSSR was 41 voxels, which was not significantly different from the size found in non-Hebrew readers [$t(29) = 0.305$, $P > 0.75$].

Having isolated a cLSSR in Hebrew readers, we compared the selectivity observed in separate event-related runs to that in the non-Hebrew readers tested previously. For all stimulus classes except Hebrew words, the pattern of response across the stimulus categories (Fig. 3*B*) was similar to that observed in the non-Hebrew readers (Fig. 3*A*): The response to English words was significantly greater than to digit strings [$t(10) = 2.531$, $P < 0.03$] and Chinese characters [$t(10) = 7.993$, $P < 0.0001$], and there was no significant difference in the response to English words and consonant strings [$t(10) = 0.401$, $P > 0.69$]. Critically, however, the response to Hebrew words was much greater in the Hebrew readers than in the non-Hebrew readers. A two-way ANOVA with group and stimulus type as factors showed a main effect of stimulus [$F(3.3, 66.2) = 31.071$, $P < 0.0001$], but no main effect of group [$F(1, 20) = 4.137$, $P > 0.055$] and a stimulus-by-group interaction [$F(3.3, 66.2) = 7.787$, $P < 0.0001$]. This stimulus-by-group interaction was no longer present when Hebrew words were excluded from the analysis [$F(2.7, 53.8) = 0.106$, $P > 0.94$]. Paired t tests comparing the responses for each stimulus class found that only for Hebrew words was there a significant difference in response between Hebrew and non-Hebrew readers [Hebrew words, $t(20) = 4.425$, $P < 0.0001$; all

NEUROSCIENCE

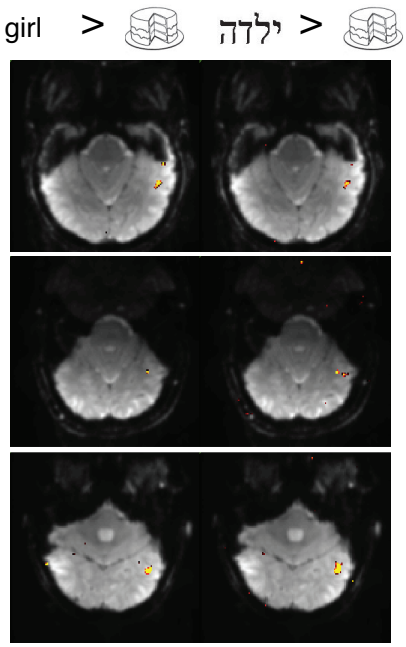


Fig. 4. Regions selectively responsive to English and Hebrew words in Hebrew readers. Representative functional slices from three Hebrew readers showing voxels selectively responsive to English words (*Left*) and Hebrew words (*Right*) compared with line drawings ($P < 10^{-4}$). Color scale as for Fig. 2.

others, $t(20) < 1.4$, $P > 0.16$]. This stronger response to Hebrew words in the Hebrew readers compared with the non-Hebrew readers suggests that experience shapes the selectivity of this region.

An alternative account of the data, however, is that the increased response to Hebrew in Hebrew readers reflects increased attention (i.e., top–down modulation) and not the effect of long-term experience on cortical selectivity (i.e., bottom–up processing). Consistent with this attentional account, the response to Hebrew words in Hebrew readers was significantly greater than the response to English words [$t(10) = 2.685$, $P < 0.023$], although the Hebrew readers were also proficient readers of English. We therefore examined the data from the one-back task, where attentional biases should, if anything, work in the opposite direction: greater attention to Hebrew words for non-readers of Hebrew than readers of Hebrew, arising from greater difficulty with Hebrew words for non-Hebrew readers (as confirmed by analysis of the behavioral data; see SI Table 2). Despite this opposite attentional bias, the blocked one-back data showed the same effect seen in the event-related data: a significantly stronger response for Hebrew words in the Hebrew readers than in non-Hebrew readers (see SI Fig. 6).

Do Hebrew and English words selectively activate the same regions in Hebrew readers? To compare the cLSSR identified with English words in Hebrew readers with regions showing Hebrew word selectivity, we directly contrasted Hebrew words and line drawings in the blocked-design data. In every Hebrew reader with a region selective for English words versus line drawings, we found a similar region selective for Hebrew words versus line drawings (Fig. 4). In every participant, these two regions overlapped considerably, with 79% of the voxels in the English word-selective region also found in the Hebrew word-selective region. The average size of the Hebrew word region was 100 voxels, almost twice the size of the cLSSR in the same participants, although this difference only just reached significance [$t(10) = 2.2$, $P < 0.05$]. Thus, similar occipitotemporal regions are selective for Hebrew and English words in readers of both languages (see also SI Fig. 7).

In summary, the response in the cLSSR to Hebrew words is much greater in Hebrew readers than in non-Hebrew readers. Thus, visual experience shapes the selectivity of the visual cortex and can lead to functionally selective cortical regions visible with fMRI.

Discussion

We report here a small region in the ventral occipitotemporal cortex identified in >85% of participants tested that responds selectively to visually presented words and letter strings compared with a variety of control stimuli. For ease of reference, we refer to this region as the cLSSR. Further, and most importantly, we show here that the selectivity of the cLSSR is shaped by experience. The response of this region to visually presented Hebrew words was substantially stronger in Hebrew readers than in non-Hebrew readers. Taken together, these findings indicate that the striking selectivity of at least one region in the ventral visual pathway for a particular stimulus class originates from extensive experience with that stimulus class. Although these findings do not demonstrate the experiential origins of other selective regions in the ventral visual pathway (FFA, parahippocampal place area, extrastriate body area, and fusiform body area), they serve as proof that extensive experience can produce selective regions of cortex.

Of course, the role for experience demonstrated here does not rule out some role for genes, which are necessary for many aspects of brain development and adult function; here we argue only that experience, rather than genetics, specifies the precise selectivity of the cLSSR. Indeed, the fairly consistent anatomical location of the cLSSR across subjects suggests that genetic contributions might predispose this region to develop the letter string-selectivity we observed (42). For example, the location of the cLSSR might be specified genetically by its connectivity (i.e., receiving inputs from visual regions and sending outputs to language-related regions), with its precise selectivity arising from learning mechanisms based on this connectivity. Although we and others (9) consider it unlikely that the precise selectivity of the cLSSR could be specified genetically in the very short period since people began reading, the hypothesis cannot be ruled out entirely until we gain a better understanding of how genes control cortical organization.

Our study demonstrates strong and robust selectivity for visual words in the extrastriate cortex, even when compared with visually similar categories of stimuli such as digit strings and Hebrew words. As detailed in the introduction, prior studies have provided only weak and inconsistent evidence for such selectivity in part because they tested few stimulus conditions (43). We speculate that many of the inconsistencies in the prior literature result from the fact that the cLSSR is very small compared with the size of typical imaging voxels, so when imaging at standard resolutions the response profile of this region will inevitably be averaged with neighboring cortex, and in some cases the region may be missed altogether (see also ref. 7). The present data resolve these ambiguities by showing that when scanning is conducted at relatively high resolution, a region that is strongly selective for letter strings (but not for words vs. consonant strings) can be found in most subjects tested.

Our finding that the cLSSR did not respond differentially to words and consonant strings does not necessarily imply that words and letter strings are processed in the same way in this region. We have shown only that there is no large-scale clustering of responsiveness to words versus consonant strings (but see refs. 17 and 44 and SI Fig. 8). Thus, it is still possible that the cLSSR contains groups of neurons separately encoding only words or only consonant strings, but that these neurons are not segregated on a scale visible at the current resolution. Indeed, fMRI

adaptation studies have shown differential sensitivity to words and their anagrams within the putative VWFA (45). However, recent behavioral work (46) suggests that letter strings and words are processed by common mechanisms in the left hemisphere: Consonant string distractors (but not Chinese or line drawing distractors) interfere with word processing to the same extent as word distractors.

An experiential origin of letter string-selectivity suggests that we should see changes in selectivity within the left hemisphere of children learning to read. Although a cross-sectional fMRI study of people ages 6 to 22 years found no relationship between reading ability and activity in the left inferotemporal cortex (35), a recent electrophysiological study (47) reported the absence of selectivity for letter strings versus symbols in the left hemisphere of children who could not yet read words, but robust selectivity in adults. Further, developmental changes in word processing in the ventral stream may extend into adolescence (48).

The experiential origins of selectivity we describe here are also consistent with studies of novel object learning. In a recent study (49), we found that 10 h of training on a novel object category produces a nearly 2-fold response to trained compared with untrained categories in some voxels in the ventral visual pathway, which was not present before training. Similarly, neurophysiological studies in nonhuman primates have found that training on discrimination and recognition of visual objects can lead to changes in the selectivity of neural responses in the inferior temporal cortex, a region critical for object recognition (50, 51). Finally, experience with visual objects can lead to clustering of neurons with similar selectivity within the inferior temporal cortex (52).

Although the present study resolves longstanding questions about the existence and experiential origins of letter string-selectivity, important questions remain. What exact computations are conducted in this region, and what kinds of representations does it extract from visually presented words? Important clues come from studies using fMRI adaptation, which suggest that representations in this region are case-independent (45, 53). Such case-independent representations also point to the role of experience in shaping the selectivity of letter string-selective regions of cortex. Second, although our findings demonstrate the role of experience in shaping the selectivity of the cLSSR, they leave unanswered the question of why this region lands in a very similar location across subjects and why it apparently cannot "move over" to adjacent cortex when damaged in adulthood (54, 55). Finally, what is the time course of the development of the cLSSR (56), what does this region do before children learn to read, and (how) does it differ in dyslexic versus normal children (26–29)?

Experimental Procedures

All experiments and procedures were approved by the institutional review boards of Massachusetts Institute of Technology and Massachusetts General Hospital. All participants gave informed consent before starting the experiments.

Participants. There were 41 right-handed participants (ages 18–45): 25 (11 male, 14 female) native English speakers with no experience of reading Hebrew (non-Hebrew readers) and 16 (11 male, 5 female) Hebrew readers who started to read Hebrew in childhood either before or shortly after learning to read English. Seven Hebrew readers were native Hebrew speakers, and the remaining nine were native English speakers with extensive experience of Hebrew. No significant differences were observed between native and nonnative Hebrew speakers, either in the size of the cLSSR or in the pattern of selectivity. Therefore, the data for Hebrew readers were collapsed across native language.

All participants were scanned in an initial blocked-design experiment. Twelve native English speakers and all Hebrew readers were also scanned in event-related runs.

One non-Hebrew reader and two Hebrew readers were excluded for excessive head motion, and one non-Hebrew reader and one Hebrew reader terminated the scan session early.

Stimuli. Participants saw concrete English words (uppercase), English consonant strings (uppercase), Hebrew words, line drawings, digit strings, and Chinese characters. Separate sets of stimuli were used for the blocked-design and event-related runs. English words, consonant strings, digit strings, and Hebrew words all subtended ≈1.5° of visual angle in height and ≈1° in width for each character. Chinese characters subtended ≈4° square, and line drawings subtended up to 6° along their longest axis. For further details on the characteristics of the orthographic stimuli, see *SI Text*.

Scanning. Participants were scanned at the Martinos Center for Biomedical Imaging at Massachusetts General Hospital in a 3T Siemens (Erlangen, Germany) Trio magnet with a custom-built eight-channel phased-array surface coil. Functional images were acquired with an EPI sequence with GRAPPA (echo time = 46 msec; 128 × 128 matrix; field of view 180 × 180 mm; 12–15 slices approximately parallel to the base of the temporal lobe with a slice thickness of 2 mm and an interslice gap of 0.4 mm). For blocked experiments, repetition time = 3 sec; for event-related designs, repetition time = 1.5 sec.

Participants were scanned in both blocked-design and event-related runs. Blocked-design runs were used to identify word-selective regions. Event-related runs were used to characterize the responsiveness of these regions across different types of stimuli.

Blocked-Design Runs. Participants viewed four runs during which 15-sec blocks (30 stimuli per block) of English words, consonant strings, Hebrew words, or line drawings were presented interleaved with blocks of no visual presentation. Stimuli were presented for 200 msec with an interstimulus interval of 300 msec. Each run contained 21 blocks and lasted for 5 min and 15 sec. For half of these runs, participants simply had to maintain fixation. In the other half, participants performed a one-back task, responding every time the same image was presented twice in a row (see SI Table 2 for behavioral data).

Event-Related Runs. Participants viewed five runs during which English words, consonant strings, digit strings, Chinese characters, line drawings of objects, and Hebrew words were presented on interleaved trials. Each trial lasted for 1.5 sec, with a 300-msec stimulus presentation followed by 1,200 msec of fixation. During each trial, the stimulus moved either to the right or down, and the participant's task was to report the direction of motion using a button box (see SI Table 3 for behavioral data). Each run contained 150 trials (25 per stimulus class) interspersed with periods of fixation and lasted for 5 min and 39 sec.

FFA Localizer. If time permitted, participants also viewed two blocked-design runs during which 15-sec blocks of photographs of faces or objects were presented interleaved with blocks of no visual presentation. Each run comprised 21 blocks with 20 images in each stimulus block and lasted 5 min and 15 sec; participants performed a one-back task.

Analysis of Imaging Data. Data were analyzed with Freesurfer (www.surfer.nmr.mgh.harvard.edu/) (57, 58), froi software (www.froi.sourceforge.net), and custom Matlab code. Data were motion-corrected and, for the blocked-design runs only, smoothed with a Gaussian kernel of 3-mm FWHM. Significance maps of the brain were computed by performing *t* tests for

pairwise comparisons of conditions and thresholded at $P = 0.0001$ (uncorrected).

Preprocessing did not involve any spatial normalization of subjects in a common reference space (e.g., Talairach). Given the anatomical variability between subjects, such normalization would obscure finer spatial patterns in activations, certainly at the high resolution that we used.

To avoid selection bias, all analyses were performed by using one data set to define ROIs and an independent data set to examine the pattern of response across conditions.

Within the WvD ROI, statistical analysis was performed on the peak responses during event-related and blocked-design runs. For event-related data, we averaged the responses at 4.5 and 6 sec poststimulus onset. For blocked-design data, we averaged responses at 6, 9, 12, and 15 sec after onset of the block.

We thank N. Knouf for technical support and M. Behrmann and E. Vul for helpful comments on the manuscript. This research was supported by National Institutes of Health Grants EY13455 and NS049052 (to N.K.), National Center for Research Resources Grant P41-RR14075, Biomedical Informatics Research Network Morphometric Project BIRN002, and the Mental Illness and Neuroscience Discovery Institute.

1. Grill-Spector K, Malach R (2004) *Annu Rev Neurosci* 27:649–677.
2. Kanwisher NG, McDermott J, Chun MM (1997) *J Neurosci* 17:4302–4311.
3. Puce A, Allison T, Gore JC, McCarthy G (1995) *J Neurophysiol* 74:1192–1199.
4. Epstein R, Kanwisher N (1998) *Nature* 392:598–601.
5. Epstein R, Harris A, Stanley D, Kanwisher N (1999) *Neuron* 23:115–125.
6. Peelen MV, Downing PE (2005) *J Neurophysiol* 93:603–608.
7. Schwarzlose RF, Baker CI, Kanwisher N (2005) *J Neurosci* 25:11055–11059.
8. Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) *Science* 293:2470–2473.
9. Tooby J, Cosmides L (2000) in *The New Cognitive Neurosciences*, ed Gazzaniga MS (MIT Press, Cambridge, MA), pp 1167–1178.
10. Polk TA, Farah MJ (1998) *Proc Natl Acad Sci USA* 95:847–852.
11. McCandliss BD, Cohen L, Dehaene S (2003) *Trends Cogn Sci* 7:293–299.
12. Polk TA, Stallcup M, Aguirre GK, Alsop DC, D'Esposito M, Detre JA, Farah MJ (2002) *J Cogn Neurosci* 14:145–159.
13. Petersen SE, Fox PT, Snyder AZ, Raichle ME (1990) *Science* 249:1041–1044.
14. Hasson U, Levy I, Behrmann M, Hendler T, Malach R (2002) *Neuron* 34:479–490.
15. Puce A, Allison T, Asgari M, Gore JC, McCarthy G (1996) *J Neurosci* 16:5205–5215.
16. Cohen L, Dehaene S, Naccache L, Lehericy S, Dehaene-Lambertz G, Henaff MA, Michel F (2000) *Brain* 123:291–307.
17. Cohen L, Lehericy S, Chochon F, Lemer C, Rivaud S, Dehaene S (2002) *Brain* 125:1054–1069.
18. Nobre AC, Allison T, McCarthy G (1994) *Nature* 372:260–263.
19. Tarkiainen A, Cornelissen PL, Salmelin R (2002) *Brain* 125:1125–1136.
20. Tarkiainen A, Helenius P, Hansen PC, Cornelissen PL, Salmelin R (1999) *Brain* 122:2119–2132.
21. Maurer U, Brandeis D, McCandliss BD (2005) *Behav Brain Funct* 1:13.
22. Bentin S, Mouchetant-Rostaing Y, Giard MH, Echallier JF, Pernier J (1999) *J Cogn Neurosci* 11:235–260.
23. Behrmann M, Nelson J, Sekuler EB (1998) *Neuropsychologia* 36:1115–1132.
24. Gaillard R, Naccache L, Pinel P, Clemenceau S, Volle E, Hasboun D, Dupont S, Baulac M, Dehaene S, Adam C, *et al.* (2006) *Neuron* 50:191–204.
25. Leff AP, Crewes H, Plant GT, Scott SK, Kennard C, Wise RJ (2001) *Brain* 124:510–521.
26. Salmelin R, Service E, Kiesila P, Uutela K, Salonen O (1996) *Ann Neurol* 40:157–162.
27. Salmelin R, Helenius P (2004) *Sci Stud Read* 8:257–272.
28. Shaywitz BA, Shaywitz SE, Pugh KR, Mencl WE, Fulbright RK, Skudlarski P, Constable RT, Marchione KE, Fletcher JM, Lyon GR, *et al.* (2002) *Biol Psychiatry* 52:101–110.
29. Paulesu E, Demonet JF, Fazio F, McCrory E, Chanoine V, Brunswick N, Cappa SF, Cossu G, Habib M, Frith CD, *et al.* (2001) *Science* 291:2165–2167.
30. Cohen L, Dehaene S (2004) *Neuroimage* 22:466–476.
31. Fiebach CJ, Rissman J, D'Esposito M (2006) *Neuron* 51:251–261.
32. Tagamets MA, Novick JM, Chalmers ML, Friedman RB (2000) *J Cogn Neurosci* 12:281–297.
33. Moore CJ, Price CJ (1999) *Neuroimage* 10:181–192.
34. Bookheimer S, Zeffiro T, Blaxtron T, Gaillard WD, Theodore W (1995) *Hum Brain Mapp* 3:93–106.
35. Turkeltaub PE, Gareau L, Flowers DL, Zeffiro TA, Eden GF (2003) *Nat Neurosci* 6:767–773.
36. Joseph JE, Gathers AD, Piper GA (2003) *Cogn Brain Res* 17:56–67.
37. Price CJ, Devlin JT (2003) *Neuroimage* 19:473–481.
38. Devlin JT, Jamison HL, Gonnerman LM, Matthews PM (2006) *J Cogn Neurosci* 18:911–922.
39. Starrfelt R, Gerlach C (2007) *Neuroimage* 35:334–342.
40. Buchsbaum BR, Hickok G, Humphries C (2001) *Cognit Sci* 25:663–678.
41. Baker CI, Hutchison TL, Kanwisher N (2007) *Nat Neurosci* 10:3–4.
42. Martin A (2006) *Neuron* 50:173–175.
43. Downing PE, Chan AW, Peelen MV, Dodds CM, Kanwisher N (2006) *Cereb Cortex* 16:1453–1461.
44. Binder JR, Medler DA, Westbury CF, Liebenthal E, Buchanan L (2006) *Neuroimage* 33:739–748.
45. Dehaene S, Jobert A, Naccache L, Ciuciu P, Poline JB, Le Bihan D, Cohen L (2004) *Psychol Sci* 15:307–313.
46. Finkbeiner M, Almeida J, Caramazza A (2006) *Cogn Neuropsych* 23:1083–1103.
47. Maurer U, Brem S, Bucher K, Brandeis D (2005) *J Cogn Neurosci* 17:1532–1552.
48. Brem S, Bucher K, Halder P, Summers P, Dietrich T, Martin E, Brandeis D (2005) *Neuroimage* 33:749–758.
49. Op de Beeck HP, Baker CI, DiCarlo JJ, Kanwisher NG (2006) *J Neurosci* 26:13025–13036.
50. Baker CI, Behrmann M, Olson CR (2002) *Nat Neurosci* 5:1210–1216.
51. Kobatake E, Wang G, Tanaka K (1998) *J Neurophysiol* 80:324–330.
52. Erickson CA, Jagadeesh B, Desimone R (2000) *Nat Neurosci* 3:1143–1148.
53. Dehaene S, Naccache L, Cohen L, Bihan DL, Mangin JF, Poline JB, Riviere D (2001) *Nat Neurosci* 4:752–758.
54. Henry C, Gaillard R, Volle E, Chiras J, Ferrieux S, Dehaene S, Cohen L (2005) *Neuropsychologia* 43:1983–1989.
55. Cohen L, Henry C, Dehaene S, Martinaud O, Lehericy S, Lemer C, Ferrieux S (2004) *Neuropsychologia* 42:1768–1780.
56. Golarai G, Ghahremani DG, Whitfield-Gabrieli S, Reiss A, Eberhardt JL, Gabrieli JD, Grill-Spector K (2007) *Nat Neurosci* 10:512–522.
57. Dale AM, Fischl B, Sereno MI (1999) *Neuroimage* 9:179–194.
58. Fischl B, Sereno MI, Dale AM (1999) *Neuroimage* 9:195–207.

SI Text

Selectivity for Words Versus Consonant Strings

Why do we find no difference in response in the cLSSR to words and consonant strings, whereas previous studies of the putative VWFA have reported a higher response for words than consonant strings (1-5)? One possible explanation for this discrepancy is that attention confounds in these earlier studies that might have elevated the response to words over consonant strings (6). For example, Cohen et al. (4) and Polk and Farah (3) asked participants to passively view word and consonant strings. However, because words are more engaging than consonant strings, the higher responses to words than consonant strings could simply reflect greater attention to words. Similarly, the higher response to words than consonant strings reported by Buchel, Price, and colleagues (1, 2, 5) could reflect increased attentional engagement due to task difficulty. Participants were asked to perform a feature detection task for ascending letters within words and consonant strings, a task which may be attentionally more demanding for words than consonant strings due to the need to suppress lexical/phonological processing while identifying local features of the stimuli [as suggested by slower reaction times for both words and pseudowords (2, 5)].

To address whether differences in attention could explain the previously observed differences between words and consonant strings, we compared the response to words and consonant strings in the blocked-design runs of our data both when subjects were passively viewing the stimuli and when they were performing the one-back task. We analyzed data across all participants (non-Hebrew readers and Hebrew readers) and to avoid selection bias, used a split-half design. Half the data (one passive viewing and one one-back run) were used to identify the cLSSR and the remaining passive viewing and 1-back runs were used to determine the pattern of response in the two tasks (SI Fig. 8). One Hebrew reader was excluded from this analysis because the cLSSR could not be localized with only half the data set. We found no difference between the responses for words and consonant strings while subjects were performing the 1-back task [$t(29) = 0.17$, $P > 0.86$], consistent with our findings for the event-related data. However, there was a small but significantly higher response to words than consonant strings during passive viewing [$t(29) = 2.54$, $P < 0.017$]. Confirming this task-dependence of the higher response to words than consonant strings, a two-way ANOVA with stimulus (words vs. consonant strings) and task (passive viewing vs. one-back) as factors revealed a main effect of task [$F(1, 29) = 11.96$, $P < 0.002$], no main effect of stimulus [$F(1, 29) = 3.28$, $P > 0.08$], but a significant task by stimulus interaction [$F(1, 29) = 4.92$, $P < 0.035$]. Although it could be argued that the greater difficulty of the one-back task for consonant strings compared to English words might increase attentional demands and thus responses to consonant strings, thus counteracting an inherent selectivity for words, this account cannot explain the lack of a higher response to words than consonant strings in the event-related data where, if anything, attention should favor words over consonant strings. Thus, the higher response to English words than consonant strings is task-dependent and is apparently only observed under conditions in which attention favors words.

If the increased response to words compared to consonant strings in the passive viewing condition really is the result of increased attention to words, then we would predict that this attentional modulation would be evident throughout extrastriate cortex and not just in the cLSSR. Consistent with this prediction, we found a similar pattern to that observed in the cLSSR in regions of extrastriate cortex that were responsive to English words (relative to fixation baseline) but not selective for words (relative to line drawings). Specifically we found a small but significantly greater response to words than consonant strings under passive viewing [$t(30) = 2.26$, $P < 0.031$], but no difference in response during performance of the one-back task [$t(30) = 1.73$, $P > 0.09$]. Further, a two-way ANOVA with stimulus (words vs. consonant strings) and task (passive viewing vs. one-back) as factors revealed no main effect of stimulus [$F(1, 30) = 0.08$, $P > 0.75$], but a main effect of task [$F(1,30) = 45.03$, $P < 0.0005$] and a significant task by stimulus interaction ($F(1, 30) = 10.63$, $P < 0.003$], just as we found in the cLSSR. Thus the increased response to English words versus consonant strings seen in the passive viewing task is not specific to the cLSSR, but is observed elsewhere in extrastriate cortex, consistent with an attentional account.

In summary, the cLSSR shows no stably higher response to English words than consonant strings when attentional biases are minimized. Previously reported differences in the ventral visual pathway to words and consonant strings may arise from attentional confounds.

Visual Similarity of Stimulus Classes Tested

The pattern of response observed in the WvD ROI matched ratings of the similarity (1, least similar; 10, most similar) of exemplars from each category to English words given by a new set of non-hebrew-reading participants ($n = 10$; consonant strings, 8.0; digit strings, 5.38; Hebrew words, 4.36; Chinese characters, 2.56; line drawings, 1.74). Similarity to English words was significantly higher for consonant strings than Hebrew words [$t(9) = 7.3$, $P < 0.001$], and significantly higher for Hebrew words than for Chinese characters [$t(9) = 5.7$, $P < 0.001$]. However, there was no difference in the ratings for Hebrew words and digit strings [$t(9) = 1.7$, $P > 0.1$].

Role of Attention in Higher Responses to Hebrew in Hebrew Readers

To test the hypothesis that increased responses to Hebrew stimuli might reflect increased attention to these stimuli, we measured responses during the one-back localizer task (SI Fig. 7) in which the opposite attentional bias would be expected (because the one-back task on Hebrew words is much more difficult, and hence attentionally engaging, for non-Hebrew than Hebrew readers). fMRI data were analyzed using a split half design to avoid selection bias. Half the data (one passive viewing and one one-back run) were used to identify the cLSSR and the pattern of response was determined using the remaining one-back run. One Hebrew reader had to be excluded from this analysis because the cLSSR could not be localized with only half the data set. A two-way ANOVA with stimulus and group as factors showed a main effect of stimulus [$F(1.8, 33.4) = 27.531$, $P < 0.013$], no main effect of group [$F(1, 19) = 2.39$, $P > 0.13$], and, critically, a stimulus-by-group

interaction [$F(1.8, 33.4) = 5.266$, $P < 0.013$]. Comparison of the response to the different stimuli across groups showed a significant difference in the response to Hebrew words [$t(19) = 3.477$, $P < 0.003$], but not to English words [$t(19) = 0.476$, $P > 0.6$], consonant strings [$t(19) = 1.435$, $P > 0.16$], or line drawings [$t(19) = 0.22$, $P > 0.8$]. Thus, differences in attentional engagement cannot account for the increased response to Hebrew words in Hebrew readers compared to non-Hebrew readers, indicating that experience indeed shapes the selectivity of voxels in the cLSSR.

Characteristics of Word and Consonant String Stimuli

In the blocked runs, words (60 in each run from a set of 120) and consonant strings (60 in each run from a set of 120) contained between four and seven characters (mean = 5). In the event-related runs, all words (50) and consonant strings (50) contained four characters (see SI Table 1 for a complete list of words and consonant strings used in the event-related runs). Consonant strings were formed by reversing the order of letters in each of the words and replacing vowels with matched consonants (a = v, e = w, i = x, o = b, u = l). Digit strings contained the digits 2 through 9 (0 and 1 were excluded because of similarity to individual letters) and were formed by mapping individual letters onto specific digits (a = 3, b = 4, c = 5, etc.).

Characteristics of the English word and consonant string stimuli were assessed using WordGen (7). Average word frequencies (per million) were 98 and 120 for the blocked-design and event-related runs, respectively. English words and consonant strings differed significantly both in terms of average number of orthographic neighbors (Blocked runs: words, 4.6; consonant strings, 0.1; Event-related runs: words, 8.5; consonant strings, 0.2) and average summed bigram frequencies (Blocked runs: words, 8,703; consonant strings, 669; Event-related runs: words, 5,490; consonant strings, 477).

Hebrew words were the closest possible translation of the English words, and thus Hebrew and English words were matched for meaning. The Hebrew words contained on average four characters in the blocked runs and 3.7 characters in the event-related runs.

1. Buchel C, Price C, Friston K (1998) Nature 394:274-277.

2. Price CJ, Wise RJ, Frackowiak RS (1996) Cereb Cortex 6:62-70.

3. Polk TA, Farah MJ (2002) J Exp Psychol Gen 131:65-72.

4. Cohen L, Lehericy S, Chochon F, Lemer C, Rivaud S, Dehaene S (2002) Brain 125:1054-1069.

5. Price CJ, Warburton EA, Moore CJ, Frackowiak RS, Friston KJ (2001) J Cogn Neurosci 13:419-429.

6. Nobre AC, Allison T, McCarthy G (1998) Brain 121:1357-1368.

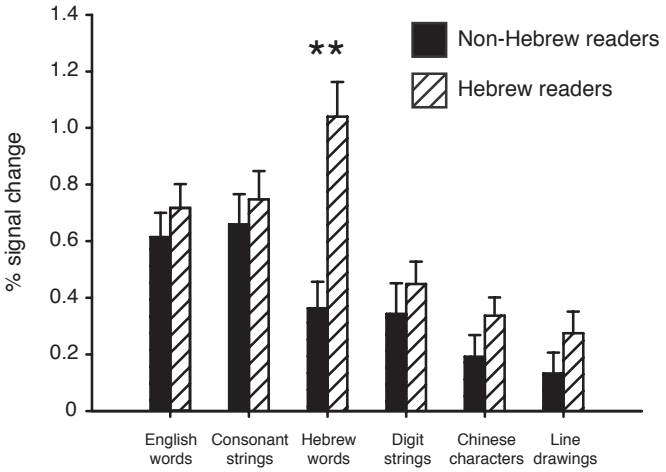
7. Duyck W, Desmet T, Verbeke LP, Brysbaert M (2004) Behav Res Methods Instrum Comput 36:488-499.

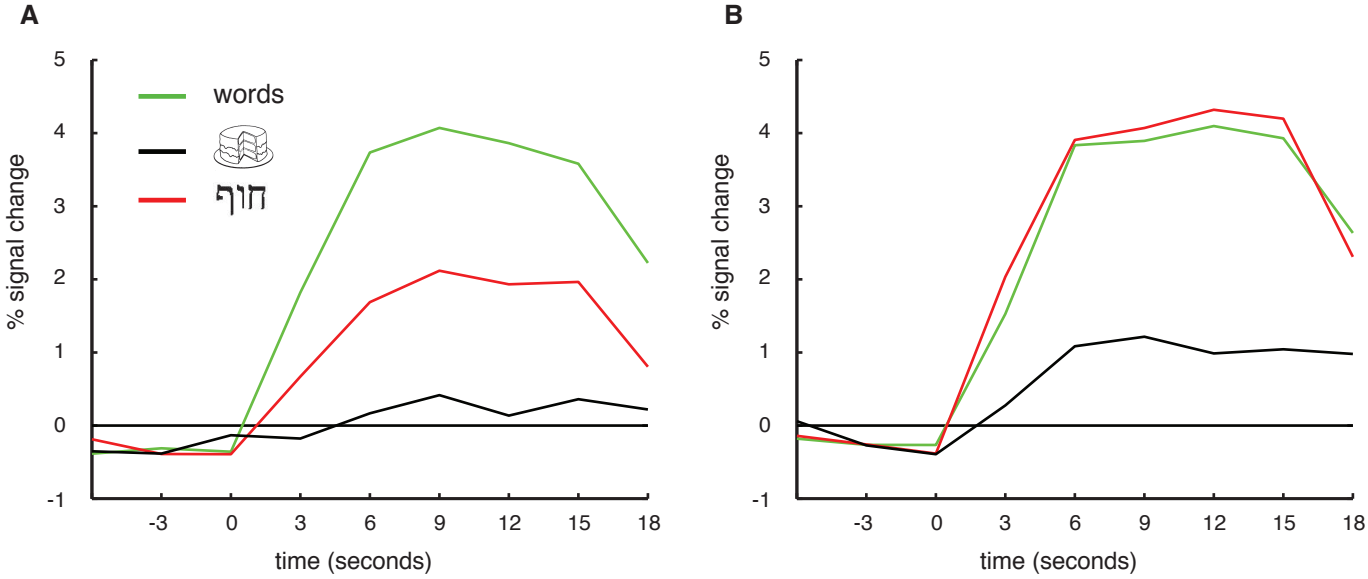
Fig. 5. Comparison of peak responses to different stimuli in non-Hebrew and Hebrew readers. The pattern of response was similar in both sets of readers except that the response to Hebrew words was much greater in the Hebrew readers than in the non-Hebrew readers ($P < 0.0001$). This increased response to Hebrew in participants with long-term experience of Hebrew words indicates that training shapes the selectivity of the cLSSR. There was no significant difference in response between groups to any of the other stimulus classes (all $P > 0.16$). Error bars show SE of the mean. (**, $P < 0.0001$).

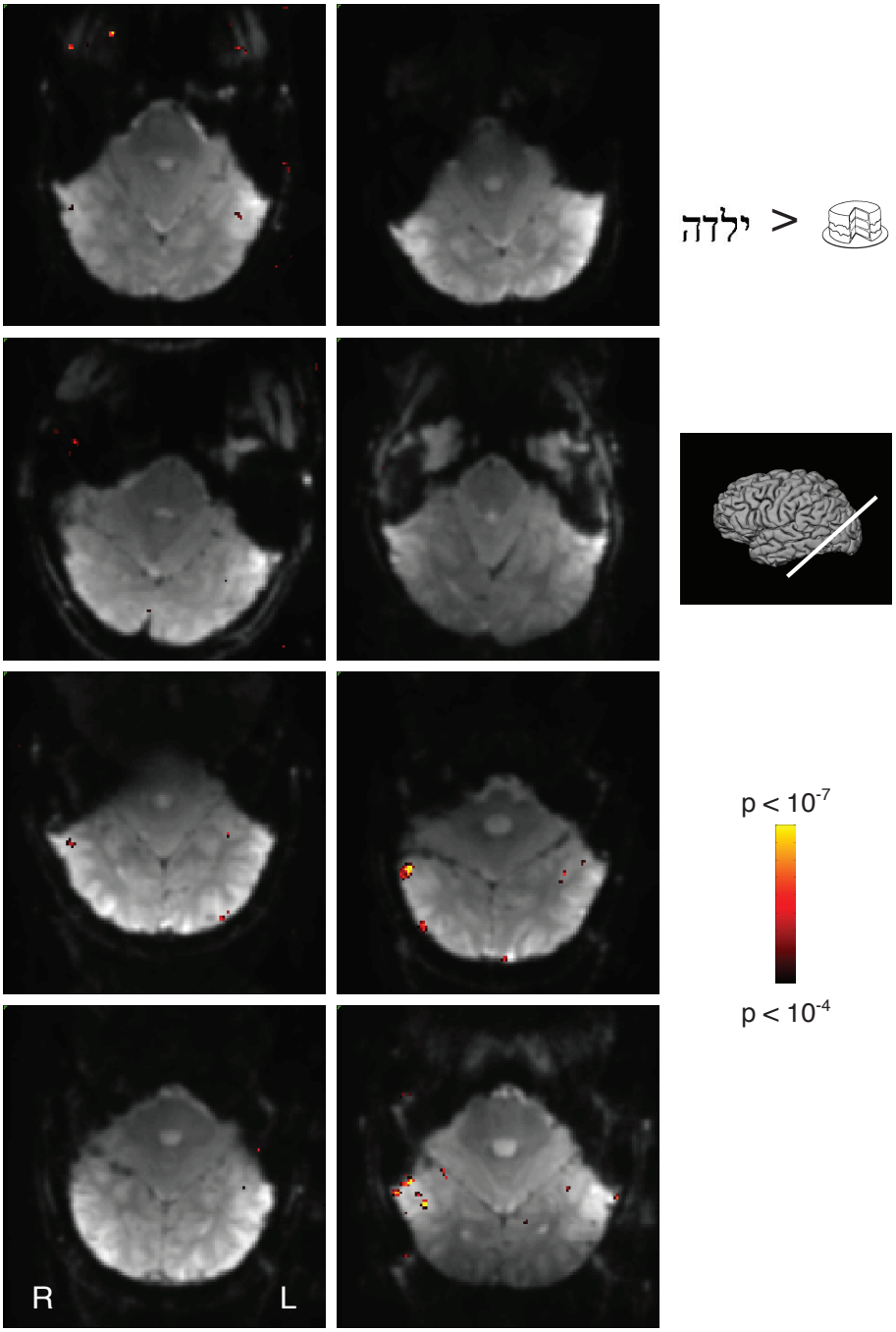
Fig. 6. Responses to English and Hebrew words and line drawings in the cLSSR during the one-back task in the blocked-design runs. Despite the increased difficulty for non-Hebrew readers (A) performing the one-back task on Hebrew words, there was nonetheless a much stronger response to English than Hebrew words in these participants, arguing against an attentional account of this effect. In Hebrew readers (B), however, there was no difference in response to Hebrew and English words, and the response to Hebrew words was significantly greater than that found in non-Hebrew readers ($P < 0.003$). The interaction between group and stimulus was highly significant ($P < 0.013$).

Fig. 7. Regions selectively responsive to Hebrew words in non-Hebrew readers. Data from the eight English readers in Fig. 2 showing voxels selectively responsive to Hebrew words compared with line drawings ($P < 10^{-4}$).

Fig. 8. Task-specific modulation of responses to words and consonant strings. A significant difference in response of the WvD ROI to words and consonant strings was only present during passive viewing and not when participants were required to perform a one-back task on all stimuli. A two-way ANOVA revealed a significant interaction between task and stimulus ($P < 0.035$). Error bars show SE of the mean (*, $P < 0.05$).


1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
% signal change
**
Non-Hebrew readers
Hebrew readers
English words
Consonant strings
Hebrew words
Digit strings
Chinese characters
Line drawings


A
B
words
חוף
% signal change
time (seconds)
5
4
3
2
1
0
-1
-3
0
3
6
9
12
15
18


ילדה >
R
L
p < 10-7
p < 10-4

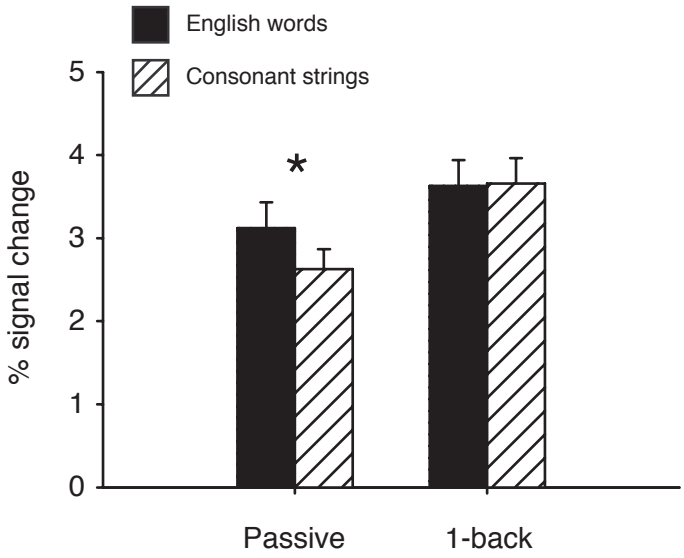

English words
Consonant strings
% signal change
5
4
3
2
1
0
*
Passive
1-back

Table S1. English word and consonant string stimuli used in the event-related runs.

English Words	Consonant strings
ARMY	YMRV
BABY	YBVB
BIRD	DRXB
BOAT	TVBB
BODY	YDBB
BUSH	HSLB
CALF	FLVC
CASH	HSVC
CITY	YTXC
COAT	TVBC
COIN	NXBC
DUCK	KCLD
FACE	WCVF
FARM	MRVF
FILM	MLXF
FIRE	WRXF
FIST	TSXF
FLAG	GVLF
FLEA	VWLF
FORK	KRBF
GIFT	TFXG
GIRL	LRXG
GOAT	TVBG
GOLD	DLBG
HAIR	RXVH
HAND	DNVH
HAWK	KWVH
HEEL	LWWH
HILL	LLXH
KITE	WTXK
KNEE	WWNK
LADY	YDVL
LAKE	WKVL
LAWN	NWVL
LEAF	FVWL
LOCK	KCBL
LUNG	GNLL
MAST	TSVM
MATE	WTVM
MILK	KLXM
NEST	TSWN
PARK	KRVP
POND	DNBP
RICE	WCXR
ROAD	DVBR
ROSE	WSBR
SHIP	PXHS
SIGN	NGXS
SNOW	WBNS
WINE	WNXW

Table S2. Behavioral performance in the 1-back task. Analyses of the behavioral data confirmed the hypothesized greater difficulty for Hebrew words in the 1-back task: only for percentage hits for Hebrew words was there a significant difference between the Hebrew and non-Hebrew readers ($t[20] = 4.3$, $p < 0.0005$; all other RT and accuracy comparisons, $t < 1.7$, $p > 0.1$).

Group	Hits*/RT (standard deviation)			
	English words	Consonant strings	Hebrew words	Line drawings
English Readers (n = 20)	77.8% (18.2) 646.0 ms (69)	68.8% (16.0) 648.9 ms (64)	62.8% (17.7) 674.8 ms (75)	84.1% (11.7) 635.9 ms (83)
Hebrew Readers (n = 11)	80.1% (10.4) 641.4 ms (59)	75.6% (14.6) 661.7 ms (60)	87.5% (9.3) 629.5 ms (68)	79.0% (14.9) 649.0 ms (102)

* False alarm rates for both groups for all conditions were less than 3%

Table S3. Behavioral performance on the motion task in the event-related runs.

Group	Accuracy/RT (standard deviation)					
	English words	Consonant strings	Digit strings	Chinese characters	Line drawings	Hebrew words
English Readers (n = 11)	96.7% (5.4) 521.1 ms (61)	95.1% (5.1) 521.0 ms (71)	95.9% (4.6) 517.5 ms (66)	96.2% (5.2) 522.6 ms (65)	95.3% (5.1) 542.3 ms (80)	97.2% (3.8) 527.4 ms (72)
Hebrew Readers (n = 11)	92.0% (12.2) 533.5 ms (97)	92.0% (13.1) 538.5 ms (108)	92.7% (12.9) 533.1 ms (95)	92.3% (11.2) 537.4 ms (104)	91.9% (10.9) 557.4 ms (115)	91.6% (13.5) 539.0 ms (98)