

Top-Down Processing of Symbolic Meanings Modulates the Visual Word Form Area

Yiyi Song,^{1*} Moqian Tian,^{1*} and Jia Liu^{1,2}

¹State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China, and ²Graduate University of Chinese Academy of Sciences, Beijing 100049, China

Functional magnetic resonance imaging (fMRI) studies on humans have identified a region in the left middle fusiform gyrus consistently activated by written words. This region is called the visual word form area (VWFA). Recently, a hypothesis, called the interactive account, is proposed that to effectively analyze the bottom-up visual properties of words, the VWFA receives predictive feedback from higher-order regions engaged in processing sounds, meanings, or actions associated with words. Further, this top-down influence on the VWFA is independent of stimulus formats. To test this hypothesis, we used fMRI to examine whether a symbolic nonword object (e.g., the Eiffel Tower) intended to represent something other than itself (i.e., Paris) could activate the VWFA. We found that scenes associated with symbolic meanings elicited a higher VWFA response than those not associated with symbolic meanings, and such top-down modulation on the VWFA can be established through short-term associative learning, even across modalities. In addition, the magnitude of the symbolic effect observed in the VWFA was positively correlated with the subjective experience on the strength of symbol–referent association across individuals. Therefore, the VWFA is likely a neural substrate for the interaction of the top-down processing of symbolic meanings with the analysis of bottom-up visual properties of sensory inputs, making the VWFA the location where the symbolic meaning of both words and nonword objects is represented.

Introduction

Previous fMRI studies have identified a region in the left middle fusiform gyrus of the human brain activated by written words. This region, called the visual word form area (VWFA) (Cohen et al., 2000, 2002; McCandliss et al., 2003), is shaped by reading experiences (Hashimoto and Sakai, 2004; Baker et al., 2007; Brem et al., 2010; Dehaene et al., 2010) and is not sensitive to low-level features of written words, such as size, position, font, or letter case (Dehaene et al., 2004; Binder et al., 2006; Vinckier et al., 2007; Glezer et al., 2009; Qiao et al., 2010; Braet et al., 2012). Neuropsychological studies have further revealed that the VWFA is necessary for reading, as the lesion of the VWFA is related to pure alexia with the hallmark feature of word-length effect (Mani et al., 2008; Pflugshaupt et al., 2009; Starrfelt et al., 2009). While extensive studies have focused on how the VWFA is tuned to visual properties of words, little is known about whether top-down influences modulate the VWFA. Recently, Price and Devlin (2011) proposed a hypothesis, called the interactive account, that proposes that the functionality of the VWFA depends on top-down

predictions mediated by feedback connections interacting with bottom-up sensory inputs. To test this hypothesis, we examined whether top-down influences generated from the processing of symbolic meanings associated with objects could interact with sensory inputs in the VWFA.

Words are a prototypical example of symbols, which are defined as something that denotes something other than themselves (DeLoache, 2000, 2004). Symbolization is thought a hallmark of human cognition: the use of symbols for acquiring knowledge distinguishes humans from other species (DeLoache, 2004) and symbolization plays a prominent role in cognitive development (DeLoache, 1987; Preissler and Carey, 2004). Although language is the most powerful and prevalent symbol system, we are exposed in daily life to all kinds of symbols, such as pictures, maps, and traffic lights. If, as proposed by the interactive account, the VWFA is modulated by top-down prediction representing associations between visual stimuli and speech sounds, meanings, or actions, we should expect the VWFA to be sensitive not only to symbolic objects (e.g., Eiffel Tower, a symbol of Paris) in particular (Fig. 1), but also to symbol–referent associations across modalities (e.g., vision–action association) in general.

On the other hand, top-down influences alone seem unable to fully account for the functional profile of the VWFA. For example, the lateral fusiform gyrus, where the VWFA is localized, prefers foveal stimuli (Hasson et al., 2002) and word-like line junctions (Szwed et al., 2011). Because the sensitivity to foveal line junctions is critical for word recognition, the development of the VWFA may occur through a neuronal recycling process in which reading encroaches on cortical regions with a prior preference for such visual properties (Dehaene and Cohen, 2011).

Received April 18, 2012; revised June 26, 2012; accepted July 16, 2012.

Author contributions: J.L. designed research; Y.S. and M.T. performed research; Y.S., M.T., and J.L. analyzed data; Y.S., M.T., and J.L. wrote the paper.

This work was funded by the 100 Talents Program of the Chinese Academy of Sciences, the National Natural Science Foundation of China (Grants 91132703 and 31100808) and the National Basic Research Program of China (Grants 2010CB833903 and 2011CB505402). We thank Nathan Witthoft, Jingguang Li, and Siyuan Hu for comments on the manuscript.

*Y.S. and M.T. contributed equally to this work.

Correspondence should be addressed to Dr. Jia Liu, Room 405, Yingdong Building, 19 Xijiekouwai St., Haidian District, Beijing 100875, China. E-mail: liujia@bnu.edu.cn.

DOI:10.1523/JNEUROSCI.1874-12.2012

Copyright © 2012 the authors 0270-6474/12/3212277-07\$15.00/0



Figure 1. Example stimuli for each stimulus condition in Experiment 1.

Therefore, our second prediction is that the VWFA should respond more strongly to images of chairs than of scenes, because the former contains more line junctions than the latter.

Materials and Methods

General procedure

Two experiments were conducted to examine how top-down processing of symbolic meanings associated with objects interacts with bottom-up visual properties of stimuli in the VWFA. In Experiment 1, we compared VWFA response for symbolic scenes versus nonsymbolic scenes, and then examined whether the symbolic effect observed in the VWFA was read out for subjective experiences on symbol–referent association. Chinese characters and pictures of chairs were also included to examine the sensitivity of the VWFA to bottom-up visual properties of the stimuli. In Experiment 2, we asked whether processing of symbolic meanings in the VWFA originated from top-down predictions and whether it could be established through associative learning. Specifically, in a cross-modality associative learning task, we trained participants to associate a set of novel objects with a set of button presses (i.e., vision–action association), and then examined whether VWFA response for objects denoting actions was higher than untrained but visually similar objects.

Experiment 1

Participants. Twelve college students (age 21–30, 3 males) participated in Experiment 1. All participants were native Chinese speakers, right-handed, and had normal or corrected-to-normal visual acuity. The fMRI protocol was approved by the Internal Review Board of Beijing Normal University. Informed written consent was obtained from all participants before the experiment.

Stimuli. Sixteen exemplars of symbolic scenes and nonsymbolic scenes were used in this experiment (Fig. 1). The symbolic scenes were landmarks of famous cities or places worldwide (e.g., Eiffel Tower, a symbol of Paris). Nonsymbolic scenes were those seen in daily life but bearing no specific symbolic meanings. Each exemplar of the symbolic scenes had a corresponding exemplar of the nonsymbolic scenes with approximately matched structure, contrast, and other visual properties. Both the symbolic and nonsymbolic scenes contained no words or word-like characters to which the VWFA is sensitive. Therefore, with these two sets of stimuli, we were able to manipulate symbolic meanings without confounding lexical contents. In addition, 16 exemplars of chair pictures were also included in the experiment, which contained nonsymbolic meanings but preferred visual properties for the VWFA (i.e., line junctions). Finally, 16 Chinese characters were included as a baseline. The Chinese characters were relatively high-frequency words selected from

500–2000 most frequently used Chinese characters, with intermediate visual complexity (each character consisting of 6–10 strokes and 2 logographemes).

fMRI scanning. Each participant attended a single session consisting of two blocked-design localizer runs and four blocked-design experimental runs. The localizer scan consisted of familiar Chinese characters and pictures of novel scenes to localize the VWFA and the parahippocampal place area (PPA) (Epstein and Kanwisher, 1998), respectively. The experimental scan consisted of Chinese characters (different from those used in the localizer scan), images of chairs, symbolic scenes, and nonsymbolic scenes. Each functional run lasted 315 s, consisted of 21 15 s blocks. Block 1, 6, 11, 16, and 21 were fixation-only baseline blocks, and each of the remaining blocks comprised 16 exemplars of an object category. The order of the experimental blocks was symmetrically counterbalanced so that the first half of a run was the mirror order of the second half. Therefore, the position of each condition was on average equated in a run.

Within an experimental block, each stimulus was presented for 400 ms with a 500 ms interstimulus interval (ISI). Participants performed an identity one-back task (i.e., pressing a button whenever two images presented in succession were identical), which is designed to ensure that participants allocated attention equally to all stimulus conditions regardless of stimulus familiarity. Scene stimuli subtended a visual angle of $\sim 8.7^\circ \times 5.8^\circ$ in the scanner, whereas other stimuli subtended a visual angle of $\sim 7.3^\circ \times 7.3^\circ$. The center-to-center distance between stimuli and the fixation point varied $\sim 0.36^\circ$ to discourage the use of low-level visual information for performing the identity task.

fMRI data acquisition. Scanning was performed on a Siemens 3T Trio scanner (Magnetom Trio, A Tim System) with a 12-channel phased-array head coil at Beijing Normal University Imaging Center for Brain Research. Twenty-five 2 mm thick (20% skip) axial slices were collected (in-plane resolution, 1.4×1.4 mm), covering the temporal cortex, occipital cortex, and part of the parietal cortex. T2*-weighted gradient-echo, echo-planar imaging procedures were used (TR, 3 s; TE, 30 ms; flip angle, 90°). In addition, MPRAGE, an inversion prepared gradient echo sequence (bandwidth, 190 Hz/pixel; flip angle, 7° ; TR/TE/TI, 2.53 s/3.45 ms/1.1 s; voxel size, $1 \times 1 \times 2$ mm), was used to acquire 3D structural images.

fMRI data analysis. Functional data were analyzed with FS-FAST (FreeSurfer functional analysis stream, Cortechs) (Dale et al., 1999; Fischl et al., 1999). After data preprocessing, including motion correction, intensity normalization, and spatial smoothing (Gaussian kernel, 6 mm FWHM), voxel time courses for each participant were fitted with a general linear model. Each condition was modeled by a boxcar regressor

Table 1. Talairach coordinates of ROIs averaged across participants (mean \pm SD) in Experiments 1 and 2

ROI	Hemisphere	Talairach coordinates			Voxel numbers
		x	y	z	
Experiment 1					
VWFA	Left	-42 \pm 6	-51 \pm 7	-15 \pm 2	81 \pm 30
PPA	Right	31 \pm 6	-55 \pm 8	-7 \pm 4	433 \pm 163
PPA	Left	-34 \pm 4	-58 \pm 11	-8 \pm 5	383 \pm 235
Experiment 2					
VWFA	Left	-47 \pm 4	-48 \pm 6	-14 \pm 6	212 \pm 146
LOC	Right	45 \pm 4	-71 \pm 10	1 \pm 3	194 \pm 153
LOC	Left	-43 \pm 5	-78 \pm 7	-4 \pm 5	218 \pm 164
FFA	Right	44 \pm 2	-42 \pm 5	-12 \pm 3	302 \pm 160
FFA	Left	-41 \pm 2	-47 \pm 4	-14 \pm 4	159 \pm 92

matching its time course, which was then convolved with a gamma function ($\delta = 2.25$, $\tau = 1.25$).

The region of interest (ROI) approach was used to analyze the fMRI data. First, we localized the VWFA and the PPA in the occipital-temporal cortex for each participant in the localizer scan. The VWFA was defined as a set of contiguous voxels in the left midfusiform gyrus that responded more strongly to Chinese characters than to natural landscapes ($p < 0.01$, uncorrected). The relatively modest threshold used here was to localize the VWFA in as many participants as possible with a decent number of voxels. The VWFA was successfully localized in the left hemisphere in 11 of 12 participants, with a mean Talairach coordinate of -42 , -51 , -15 (Table 1). The coordinate of the VWFA along the ventral occipitotemporal cortex ($y = -51$) was close to the widely accepted coordinate of the VWFA for English words ($y = -54$) (Cohen et al., 2000, 2002; Price and Devlin, 2003) and that for Chinese characters ($y = -52$) (Bolger et al., 2005). Critically, it was substantially posterior to a region engaged in processing lexical and semantic features of words in the occipitotemporal cortex ($y = -35$) (Binder et al., 2009). The PPA was defined in the parahippocampal gyrus in the same way but with the reverse contrast ($p < 0.01$, uncorrected), and the PPA was localized bilaterally in all participants. In the following analyses, we only included participants who showed both the VWFA and the PPA to enable a direct within-subject comparison between ROIs. In addition, because the response profiles of the left and right PPA were similar, the data were pooled across hemispheres.

Time courses were extracted and averaged across all experimental runs and all voxels within the predefined ROIs for each condition and for each individual. Because fMRI responses typically lagged 4–6 s after stimulus onset, the magnitude of ROI responses was measured as the averaged percentage MR signal changes at the latency of 6, 9, 12, 15, and 18 s (block duration, 15 s with TR of 3 s) compared with the fixation as a baseline. The magnitude, one per experimental condition per ROI per participant, was then submitted to further analyses. In addition to the analysis on the magnitude of ROI responses, we also calculated a t value for each voxel by comparing symbolic scenes with nonsymbolic scenes to measure the symbolic effect in the VWFA. Then the t values were averaged across voxels within the VWFA for each participant to serve as an index for the symbolic effect.

Subjective experiences on symbol–referent association. To evaluate the strength of symbolic meanings associated with the scenes used in the fMRI experiment, the participants were surveyed after the scan. They were instructed to indicate strength of the symbolic meanings associated with each scene picture based on their subjective experiences in a seven-point Likert scale from 1 (no symbolic meaning) to 7 (strong symbolic meaning). The participants were instructed to choose 1 if they could not associate the scene with any symbolic meaning beyond its literal content, and to choose 7 if the symbolic meaning emerged rapidly and effortlessly. Several exemplar pictures (not used in the experiment) were shown to the participants for practice before the evaluation. The subjective symbolic effect for each participant was calculated as a normalized difference score between the symbolic scenes and nonsymbolic scenes: $(S - N)/(S + N)$, where S is the score for symbolic scenes and N is the score for nonsymbolic scenes.

Experiment 2

Participants. Seven college students (age 20–23, 2 males) participated in Experiment 2. Three of the students also participated in Experiment 1. All participants were native Chinese speakers and had studied English for at least 10 years. They were right-handed, and had normal or corrected-to-normal visual acuity. Informed consent was obtained from all participants before the experiment.

Stimuli. Images of dumbbell-shaped objects used in the behavioral training and fMRI scanning were constructed by connecting two component figures with a connection bar (Song, 2010a,b) (see Fig. 3A, top). The component figures were made of simple shapes (e.g., arc, line, circle, square) to avoid resemblance to everyday objects; therefore, the participants were more likely to focus on their symbolic function than on the concrete objects themselves (DeLoache, 2000).

Sixteen to-be-trained objects were created from 16 component figures. In particular, the 16 component figures were arbitrarily divided into four groups, each of which contained four component figures. Of these four component figures (A, B, C, D), two (e.g., A and B) were selected as the left component of a dumbbell-shaped object; and the other two (e.g., C and D) were selected as the right component of a dumbbell-shaped object. By combining any one of the left component figures with any one of the right component figures, four dumbbell-shaped objects (A-C, A-D, B-C, B-D) were created.

To examine prediction errors from top-down influences, we also created 16 variant objects by randomly combining a left component figure in one group and a right component in another group of the to-be-trained objects (see Fig. 3A, middle). That is, the variants contained the same part components as the to-be-trained objects on average, but their configurations were novel. Finally, another set of 16 dumbbell-shaped objects were constructed in the same way as the to-be-trained objects, but from a new set of component figures, to serve as a nonexposed baseline condition (see Fig. 3A, bottom). The to-be-trained dumbbell figures and novel dumbbell figures were counterbalanced across participants.

Behavioral training. The training in this experiment is a cross-modality associative learning task; that is, the participants were trained to associate a visual stimulus with an action, rather than to associate a visual stimulus with another visual stimulus within a modality. In each trial, a dumbbell-shaped object was presented in the center of the screen for 1000 ms, with an ISI of 1000 ms. Each stimulus required a button press with either the left or right hand. For the left-hand response, participants pressed “c” in the keyboard with the left hand, whereas for the right-hand response, participants pressed “m” with the right hand. Participants were not explicitly informed of the association between stimuli and actions before the training; however, they managed to learn correct associations through auditory feedbacks, as when they made a wrong button press, an auditory feedback was presented.

Each training session consisted of 160 trials; that is, 10 trials per object. After the average reaction times (RTs) of each session stopped decreasing during the training, we continued the training for five or more sessions to reach an asymptote (i.e., no significant decrease in RT in at least 3 consecutive sessions). On average, subjects completed 30–35 training sessions, lasting ~ 3 –4 h in total in 3 consecutive days.

fMRI scanning. Each participant attended a single session consisting of two blocked-design localizer runs and three blocked-design experimental runs. Each run consisted of 21 15 s blocks, lasting 315 s. The localizer scan consisted of concrete English nouns (e.g., body, butter, carpet) with 4–7 letters, line drawings of unnamable objects, novel faces, and scrambled line-drawn objects. Each stimulus was presented for 200 ms, with an ISI of 550 ms. The experimental scan consisted of trained dumbbell-shaped objects (Trained), variants of trained objects (Variant), novel dumbbell-shaped objects (Novel), and line-drawn objects. Each stimulus was presented for 400 ms, with an ISI of 500 ms. During the experimental scan, participants did not perform the association task as they did in behavioral training; instead, they performed an identity one-back task, the same as that in Experiment 1.

MRI data acquisition and analysis. The MRI data acquisition and analysis were similar to those described by Experiment 1 except the definition of ROIs. Here, the VWFA was defined by contrasting English words with line-drawn objects ($p < 0.01$, uncorrected), the fusiform face area (FFA)

was defined by contrasting faces with line drawn objects ($p < 0.01$, uncorrected) (Kanwisher et al., 1997), and the lateral occipital cortex (LO) was defined by contrasting line-drawn objects with scrambled line-drawn objects ($p < 0.01$, uncorrected). The VWFA was localized in five of seven participants, whereas the FFA and the LO were localized bilaterally in all participants (Table 1). In the following analyses, we only included the participants who showed all ROIs for direct comparisons between ROIs. In addition, because the factor of hemisphere did not show any main effects or interact with other factors for the LO or the FFA, the data from left and right hemisphere were collapsed.

Results

Experiment 1

To examine the symbolic effect in the VWFA, a two-way ANOVA was performed with factors being symbolic meaning (symbolic vs nonsymbolic scenes) and cortical region (VWFA vs PPA) (Fig. 2A). We found a significant main effect of region ($F_{(1,10)} = 37.81, p < 0.001$), with BOLD response for both types of scenes being significantly higher in the PPA than in the VWFA. The main effect of the symbolic meaning did not reach significance ($F_{(1,10)} = 3.10, p = 0.11$). Importantly, we found a significant two-way interaction of symbolic meaning by region ($F_{(1,10)} = 8.01, p = 0.02$). *Post hoc* pairwise *t* tests further showed that VWFA response for the symbolic scenes was significantly higher than that for the nonsymbolic scenes ($t_{(10)} = 2.69, p = 0.02$), whereas the PPA showed no significant difference in symbolic meaning ($t_{(10)} < 1$). The same pattern was found in both left and right PPA as well, with the significant interaction of symbolic meaning by region being found for left PPA versus VWFA ($F_{(1,10)} = 7.14, p = 0.02$) and right PPA versus VWFA ($F_{(1,10)} = 7.18, p = 0.02$). Thus, although the PPA showed a stronger preference to the scene pictures, the difference in symbolic meaning between two types of the scene pictures was encoded in the VWFA instead.

Although the symbolic and nonsymbolic scenes were approximately matched in visual properties, the symbolic effect observed in the VWFA may instead reflect different types of object processing. For example, the symbolic scenes contained familiar objects that can be recognized at the individual level (e.g., Eiffel Tower), whereas objects in the nonsymbolic scenes can only be named at the category level (e.g., a tower). To rule out this alternative interpretation, we examined whether the LO, a region engaged in general object processing (Grill-Spector et al., 2001), was sensitive to symbolic meanings associated with objects. First, we defined the LO with the contrast of chairs versus words using the data from the experimental runs ($p < 0.01$, uncorrected); then, we examined the response magnitudes for the symbolic and nonsymbolic scenes in the independently defined LO. The LO was localized in 11 participants in the left hemisphere and 10 participants in right. We found that the LO in both hemispheres did not respond more strongly to the symbolic scenes than to the nonsymbolic scenes (all $t < 1$). Thus, the involvement of general object processing cannot account for the higher response for symbolic scenes in the VWFA; instead, the sensitivity to the symbolic meanings associated with objects was apparently unique to the VWFA.

To further examine whether the neural information on the symbolic meanings associated with the scenes in the VWFA was

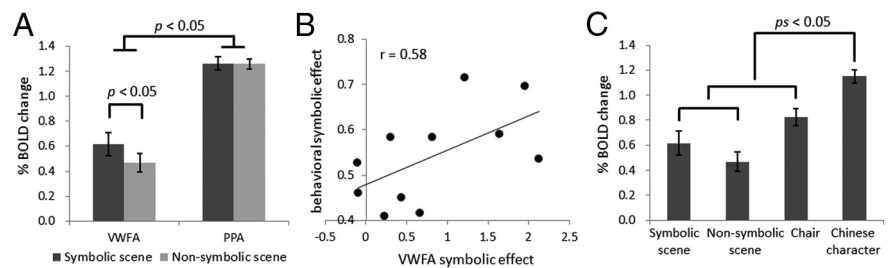


Figure 2. The symbolic effect in the VWFA. **A**, The response magnitudes for symbolic scenes and nonsymbolic scenes in the VWFA and PPA. *y*-axis indicates the percentage BOLD signal change, and error bars indicate ± 1 SEM. **B**, The correlation between the VWFA response to the symbolic scenes versus nonsymbolic scenes (VWFA symbolic effect) and the subjective experiences on the strength of symbolic meanings of the symbolic scenes versus nonsymbolic scenes (behavioral symbolic effect) across the participants. *x*-axis is the *t* value by comparing VWFA response to the symbolic scenes with the nonsymbolic scenes, whereas *y*-axis is the normalized difference in subjective evaluation of the strength of symbol–referent association in two types of scenes in a seven-point Likert scale. **C**, The response magnitude for symbolic scenes, nonsymbolic scenes, images of chairs, and Chinese characters in the VWFA. *y*-axis indicates the percentage BOLD signal change, and error bars indicate ± 1 SEM.

read out for subjective experiences, we instructed the participants to evaluate the strength of the symbolic meanings associated with the scenes after the scan, and then correlated the subjective experience with the symbolic effect observed in the VWFA. The average score for the symbolic scenes was 6.24 (SD, 0.41) in a seven-point Likert scale, which was significantly higher than that for the nonsymbolic scenes (1.88; SD, 0.57) ($t_{(10)} = 25.96, p < 0.001$), confirming the success in manipulating the symbolic meanings in the fMRI study. More importantly, the difference in subjective experiences between the symbolic and nonsymbolic scenes was positively correlated with the difference in VWFA response for these two types of scenes across individuals (Spearman $\rho = 0.60, p = 0.05$; Pearson $r = 0.58, p = 0.06$) (Fig. 2B). In contrast, no correlation between the behavioral symbolic effect and the neural symbolic effect was observed in the PPA (Spearman $\rho = 0.12, p = 0.73$; Pearson $r = -0.04, p = 0.92$).

In addition to showing the symbolic effect in the VWFA, we also found that the VWFA was sensitive to visual properties of the stimuli (Fig. 2C). As expected, the VWFA responded significantly more strongly to the Chinese characters than to the symbolic scenes ($t_{(10)} = 8.54, p < 0.001$), the nonsymbolic scenes ($t_{(10)} = 4.58, p < 0.001$), and the chairs ($t_{(10)} = 4.10, p < 0.01$). Interestingly, VWFA response for chairs, which contained numerous line junctions, was significantly higher than for both the symbolic ($t_{(10)} = 2.31, p = 0.04$) and nonsymbolic scenes ($t_{(10)} = 5.00, p < 0.001$). Thus, although the chairs contained as few symbolic meanings as the nonsymbolic scenes, the intrinsic visual properties of the chairs was still able to activate the VWFA effectively.

In short, the finding that the VWFA responded more strongly to the symbolic scenes than the nonsymbolic scenes suggests that the VWFA was engaged in processing of symbolic meanings. However, it is unclear whether this symbolic effect reflected top-down influences. If the influences are from higher-order regions engaged in processing sounds, meanings, or actions associated with objects, we should expect the VWFA to also be sensitive to symbol–referent associations across modalities (e.g., vision–action associations). In addition, the symbolic scenes are not inherently symbols. Only after they are used with the goal of referring do they become symbols. However, it is unclear how symbol–referent associations are established. To address these two questions, we trained participants in Experiment 2 to make a certain action when perceiving an object in an associative learning task. We examined the symbolic effect in the VWFA after participants

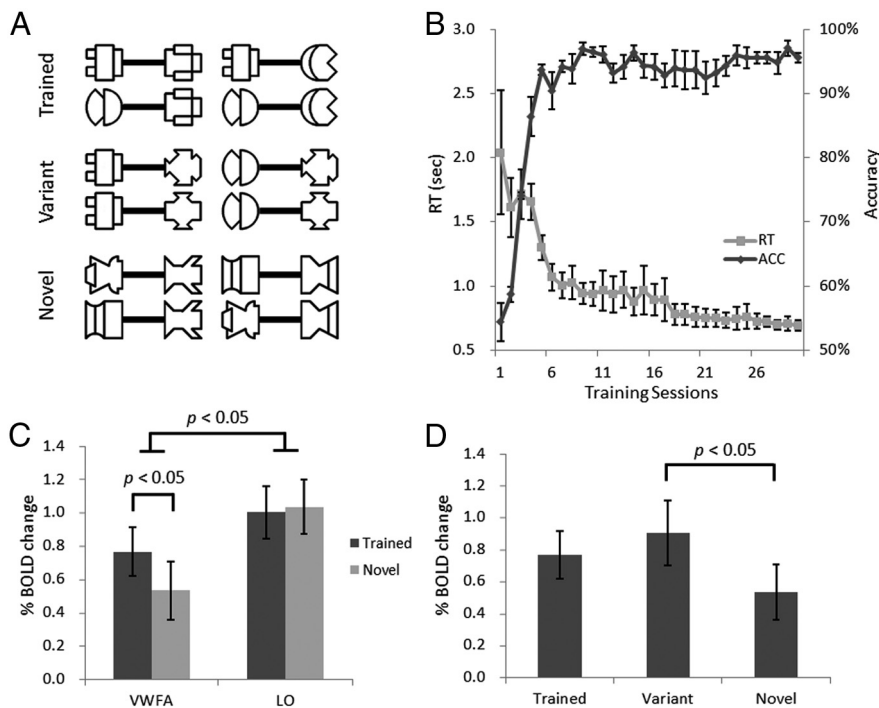


Figure 3. Associative experience in establishing the top-down influence on the VWFA. **A**, Four dumbbell-shaped exemplars for the Trained, Variant, and Novel stimuli respectively in fMRI scans in Experiment 2. Note that the Variant condition contains the same part components as those in the Trained condition, whereas their configurations are different. **B**, Behavioral results in associative learning. The accuracies and reaction times are shown as functions of training sessions. The left y-axis indicates the reaction time in making a correct action after seeing a dumbbell-shaped object, whereas the right y-axis indicates the percentage of correct responses in associative learning at the chance level of 50%. **C**, The response magnitudes for the Trained (i.e., symbolic) and Novel (i.e., nonsymbolic) stimuli in the VWFA and LO. **D**, The response magnitude for the variants of the trained stimuli (Variant) compared with the magnitude for the Trained and Novel condition respectively in the VWFA. y-axis indicates the percentage BOLD signal change, and error bars indicate ± 1 SEM.

had learned the cross-modality association (e.g., vision–action association).

Experiment 2

As expected, the behavioral training greatly improved the participants’ performance in associating visual stimuli to actions (Fig. 3B). The accuracy increased monotonically from session 1 to session 5 ($F_{(4,16)} = 65.51, p < 0.001$), and was $>90\%$ for the remaining sessions ($F_{(24,96)} = 1.11, p > 0.05$). The RT decreased monotonically from session 1 to session 17 ($F_{(16,64)} = 8.29, p < 0.001$), and then remained unchanged from session 18 to session 31 ($F_{(12,48)} = 1.23, p > 0.05$). Therefore, through the associative learning, the participants successfully learned to use the dumbbell-shaped objects (i.e., symbols) to denote button-press actions (i.e., referents).

To examine whether the symbolic effect established through the associative learning was observed in the VWFA, we compared VWFA response for the trained stimuli (i.e., symbolic objects) with that for the untrained stimuli (i.e., nonsymbolic objects) (Fig. 3C). We found a significantly higher response for the trained stimuli (vs the novel stimuli) in the VWFA ($t_{(4)} = 3.47, p = 0.03$), but not in the LO ($t < 1$). A two-way ANOVA of cortical region (VWFA vs LO) by symbolic meaning (Trained vs Novel) revealed a significant two-way interaction ($F_{(1,4)} = 7.78, p = 0.05$). In addition, the FFA showed a similar pattern as the LO; that is, no symbolic effect was observed in the FFA either ($t < 1$). Besides, a two-way interaction of cortical region (VWFA vs FFA) by symbolic meaning (Trained vs Novel) was significant ($F_{(1,4)} = 25.62,$

$p = 0.007$) as well. Together, these findings suggest top-down processing of symbolic meanings specifically modulates the VWFA, not the LO or the FFA.

Finally, we examined VWFA response for the variants of the trained stimuli, which contained the same parts as the trained stimuli but with novel configurations (Fig. 3D). We found that VWFA response for the variants was significantly higher than that for the untrained stimuli ($t_{(4)} = 4.91, p = 0.01$), suggesting that the variants were recognized as potentially meaningful but they could not be predicted efficiently (i.e., prediction errors).

Discussion

In daily life, we must master symbol systems to participate fully in society. In this study, we examined whether the VWFA, a region known to engage in reading, was modulated by top-down processing of symbolic meanings associated with objects. Two novel results were found. First, nonword objects, such as scenes and artificial objects, elicited a higher response in the VWFA when they were associated with symbolic meanings than when they were not. In addition, the symbolic effect reflected top-down influences on the VWFA, which can be established through short-term associative experiences even across modalities. Second, the magnitude of the symbolic effect observed in the VWFA was positively correlated with subjective experiences on symbol–referent association, suggesting that the symbolic meaning associated with stimuli is represented in the VWFA regardless of stimulus formats. In short, while previous findings show the sensitivity of the VWFA to visual properties of sensory inputs, our study clearly demonstrates the top-down influences on the VWFA from the processing of symbolic meanings associated with objects.

The finding that the VWFA represented symbolic meanings associated with stimuli fits perfectly with the interactive account on the functionality of the VWFA within a predictive coding framework (Price and Devlin, 2011). According to this hypothesis, cortical regions that engage in symbolic reasoning sent top-down predictions about the symbolic meanings associated with scenes and artificial objects to the VWFA when they were presented; therefore, the activation level of the VWFA was increased because of the interaction of the top-down prediction with the analysis of bottom-up visual properties of the stimuli. On the other hand, activation from the top-down predictions was precluded for the nonsymbolic scenes or untrained stimuli because there was no appropriate symbol–referent association. In addition, both symbolic scenes and artificial objects likely engaged processing of their associated meanings or actions automatically and unconsciously (Kherif et al., 2011), because we used a short presentation time (i.e., 400 ms) and a minimal task (i.e., the identity one-back task) that emphasized bottom-up processing of the stimuli during the scan. Finally, the difference in VWFA response to symbolic versus nonsymbolic scenes cannot be accounted for by whether or not the scenes could be named, because

the nonsymbolic scenes were also namable to some extent (e.g., buildings, bridges). Besides, previous studies have found that the VWFA even showed a decreased response for nameable objects (vs novel scripts) and no difference in response between nameable words and unnamable letter strings (Baker et al., 2007). Together, the results from our study suggests that the VWFA is sensitive to symbolic nonword objects through an automatic top-down processing of symbolic meanings.

Furthermore, our study extends the interactive account in two ways. First, our study not only provides direct evidence for the top-down influence on the VWFA, but also illustrates that one such top-down influence is the processing of symbolic meanings. Consistent with our findings based on scenes and artificial objects, previous studies have shown that nonword stimuli that demand processing of symbolic meanings can activate the VWFA effectively, despite their dramatic differences in visual appearance. For example, the VWFA has been shown to be sensitive to punctuation marks that refer to the manner of speaking (e.g., “?!”) (Reinke et al., 2008), body gestures that signify actions or social communication (Xu et al., 2009), and nonword objects in a category task requiring the analysis on the symbolic meanings of the objects (Starrfelt and Gerlach, 2007). Therefore, the preferred visual property (e.g., line junctions) is not a prerequisite for engaging the VWFA; instead, top-down processing of symbolic meanings is sufficient to modulate the VWFA. Second, the top-down influence on the VWFA can be established through short-term associative learning. In our previous studies, we showed that the VWFA responded more strongly to objects after they had been trained to denote English words (Song et al., 2010a,c) or line-drawn figures (Song et al., 2010b) through associative learning, which implies a top-down influence on the VWFA. However, one may argue that the activation in the VWFA did not reflect top-down processing; instead, the VWFA may simply become attuned to a newly learned language script. To rule out this alternative interpretation, we trained the participants in this study to use objects to denote actions across modalities. The cross-modality associative learning clearly demonstrates that the prediction signals must derive from higher-order cortical regions beyond either visual or motor cortex, which may be involved in processing abstract concepts in the middle temporal or frontal cortex (Binder et al., 2009; Wei et al., 2012).

In this study, scenes were chosen as the test stimuli because they do not contain many line conjunctions or require foveal vision, which were preferred visual properties of the VWFA (Hasson et al., 2002; Szwed et al., 2011). Although we observed the symbolic effect in the VWFA with the scenes, VWFA response for the scenes was lower than that for the chairs, which are associated with little symbolic meaning yet contain high-contrast line junctions. Therefore, the bottom-up visual properties suitable for reading also contribute significantly to VWFA response (Dehaene and Cohen, 2011). However, the magnitude of neural activation alone does not necessarily index the amount of information encoded in the VWFA. For example, we also found that VWFA response for the variants of the symbolic objects was higher than that for the untrained stimuli, yet both contained no symbolic meanings. The higher response for the variants may simply result from prediction errors when the top-down prediction failed to match the sensory inputs (Price and Devlin, 2011). In contrast, low responses can also carry critical information as the symbolic effect in the VWFA was read out for subjective experiences on symbol–referent association. Therefore, while the VWFA may have a weaker response to nonword symbolic objects

than to words, the VWFA nevertheless remains the place where the symbolic meaning of the stimuli is represented.

In sum, our study elucidates the role of top-down influences on the VWFA and further demonstrates that one of the top-down influences is the processing of symbolic meanings in particular (Price and Devlin, 2011). On the other hand, our results also showed that the VWFA was also sensitive to bottom-up visual properties, consistent with the neural recycling hypothesis (Dehaene and Cohen, 2011). Together, our study suggests that a model that consists of both the analysis of intrinsic visual properties of stimuli and the top-down processing of symbolic meanings associated with stimuli may provide a more comprehensive description on the nature the VWFA.

References

- Baker CI, Liu J, Wald LL, Kwong KK, Benner T, Kanwisher N (2007) Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proc Natl Acad Sci U S A* 104:9087–9092.
- Binder JR, Medler DA, Westbury CF, Liebenthal E, Buchanan L (2006) Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage* 33:739–748.
- Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex* 19:2767–2796.
- Bolger DJ, Perfetti CA, Schneider W (2005) Cross-cultural effect on the brain revisited: universal structures plus writing system variation. *Hum Brain Mapp* 25:92–104.
- Braet W, Wagemans J, Op de Beeck HP (2012) The visual word form area is organized according to orthography. *Neuroimage* 59:2751–2759.
- Brem S, Bach S, Kucian K, Guttorm TK, Martin E, Lyytinen H, Brandeis D, Richardson U (2010) Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *Proc Natl Acad Sci U S A* 107:7939–7944.
- Cohen L, Dehaene S, Naccache L, Lehéricy S, Dehaene-Lambertz G, Hénaff MA, Michel F (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.
- Cohen L, Lehéricy S, Chochon F, Lemer C, Rivaud S, Dehaene S (2002) Language-specific tuning of visual cortex? Functional properties of the visual word form area. *Brain* 125:1054–1069.
- Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9:179–194.
- Dehaene S, Cohen L (2011) The unique role of the visual word form area in reading. *Trends Cogn Sci* 15:254–262.
- Dehaene S, Jobert A, Naccache L, Ciuciu P, Poline JB, Le Bihan D, Cohen L (2004) Letter binding and invariant recognition of masked words: behavioral and neuroimaging evidence. *Psychol Sci* 15:307–313.
- Dehaene S, Pegado F, Braga LW, Ventura P, Nunes Filho G, Jobert A, Dehaene-Lambertz G, Kolinsky R, Morais J, Cohen L (2010) How learning to read changes the cortical networks for vision and language. *Science* 330:1359–1364.
- DeLoache JS (1987) Rapid change in the symbolic functioning of very young children. *Science* 238:1556–1557.
- DeLoache JS (2000) Dual representation and young children’s use of scale models. *Child Dev* 71:329–338.
- DeLoache JS (2004) Becoming symbol-minded. *Trends Cogn Sci* 8:66–70.
- Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598–601.
- Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9:195–207.
- Glezer LS, Jiang X, Riesenhuber M (2009) Evidence for highly selective neuronal tuning to whole words in the “visual word form area.” *Neuron* 62:199–204.
- Grill-Spector K, Kourtzi Z, Kanwisher N (2001) The lateral occipital complex and its role in object recognition. *Vision Res* 41:1409–1422.
- Hashimoto R, Sakai KL (2004) Learning letters in adulthood: direct visualization of cortical plasticity for forming a new link between orthography and phonology. *Neuron* 42:311–322.
- Hasson U, Levy I, Behrmann M, Hendler T, Malach R (2002) Eccentricity

- bias as an organizing principle for human high-order object areas. *Neuron* 34:479–490.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.
- Kherif F, Josse G, Price CJ (2011) Automatic top-down processing explains common left occipito-temporal responses to visual words and objects. *Cereb Cortex* 21:103–114.
- Mani J, Diehl B, Piao Z, Schuele SS, Lapresto E, Liu P, Nair DR, Dinner DS, Lüders HO (2008) Evidence for a basal temporal visual language center: cortical stimulation producing pure alexia. *Neurology* 71:1621–1627.
- McCandliss BD, Cohen L, Dehaene S (2003) The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn Sci* 7:293–299.
- Pflugshaupt T, Gutbrod K, Wurtz P, von Wartburg R, Nyffeler T, de Haan B, Karnath HO, Mueri RM (2009) About the role of visual field defects in pure alexia. *Brain* 132:1907–1917.
- Preissler MA, Carey S (2004) Do both pictures and words function as symbols for 18- and 24-month-old children? *J Cogn Dev* 5:185–212.
- Price CJ, Devlin JT (2003) The myth of the visual word form area. *Neuroimage* 19:473–481.
- Price CJ, Devlin JT (2011) The interactive account of ventral occipitotemporal contributions to reading. *Trends Cogn Sci* 15:246–253.
- Qiao E, Vinckier F, Szwed M, Naccache L, Valabrègue R, Dehaene S, Cohen L (2010) Unconsciously deciphering handwriting: subliminal invariance for handwritten words in the visual word form area. *Neuroimage* 49:1786–1799.
- Reinke K, Fernandes M, Schwindt G, O'Craven K, Grady CL (2008) Functional specificity of the visual word form area: general activation for words and symbols but specific network activation for words. *Brain Lang* 104:180–189.
- Song Y, Bu Y, Hu S, Luo Y, Liu J (2010a) Short-term language experience shapes the plasticity of the visual word form area. *Brain Res* 1316:83–91.
- Song Y, Bu Y, Liu J (2010b) General associative learning shapes the plasticity of the visual word form area. *Neuroreport* 21:333–337.
- Song Y, Hu S, Li X, Li W, Liu J (2010c) The role of top-down task context in learning to perceive objects. *J Neurosci* 30:9869–9876.
- Starrfelt R, Gerlach C (2007) The visual what for area: words and pictures in the left fusiform gyrus. *Neuroimage* 35:334–342.
- Starrfelt R, Habekost T, Leff AP (2009) Too little, too late: reduced visual span and speed characterize pure alexia. *Cereb Cortex* 19:2880–2890.
- Szwed M, Dehaene S, Kleinschmidt A, Eger E, Valabrègue R, Amadon A, Cohen L (2011) Specialization for written words over objects in the visual cortex. *Neuroimage* 56:330–344.
- Vinckier F, Dehaene S, Jobert A, Dubus JP, Sigman M, Cohen L (2007) Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. *Neuron* 55:143–156.
- Wei T, Liang X, He Y, Zang Y, Han Z, Caramazza A, Bi Y (2012) Predicting conceptual processing capacity from spontaneous neuronal activity of the left middle temporal gyrus. *J Neurosci* 32:481–489.
- Xu J, Gannon PJ, Emmorey K, Smith JF, Braun AR (2009) Symbolic gestures and spoken language are processed by a common neural system. *Proc Natl Acad Sci U S A* 106:20664–20669.