

Automatic Top-Down Processing Explains Common Left Occipito-Temporal Responses to Visual Words and Objects

Ferath Kherif, Goulven Josse and Cathy J. Price

Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London, WC1N 3BG, UK

Address correspondence to Dr Ferath Kherif, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, 12 Queen Square, London, WC1N 3BG, UK. Email: f.kherif@fil.ion.ucl.ac.uk.

Previous studies have demonstrated that a region in the left ventral occipito-temporal (LvOT) cortex is highly selective to the visual forms of written words and objects relative to closely matched visual stimuli. Here, we investigated why LvOT activation is not higher for reading than picture naming even though written words and pictures of objects have grossly different visual forms. To compare neuronal responses for words and pictures within the same LvOT area, we used functional magnetic resonance imaging adaptation and instructed participants to name target stimuli that followed briefly presented masked primes that were either presented in the same stimulus type as the target (word–word, picture–picture) or a different stimulus type (picture–word, word–picture). We found that activation throughout posterior and anterior parts of LvOT was reduced when the prime had the same name/response as the target irrespective of whether the prime–target relationship was within or between stimulus type. As posterior LvOT is a visual form processing area, and there was no visual form similarity between different stimulus types, we suggest that our results indicate automatic top-down influences from pictures to words and words to pictures. This novel perspective motivates further investigation of the functional properties of this intriguing region.

Keywords: language, priming, repetition suppression, visual system, word–object recognition

Introduction

Neuroimaging studies have shown that the visual processing of written words and pictures of objects invokes a gradient of responses along the left ventral occipito-temporal (LvOT) cortex with posterior LvOT regions processing visual forms and anterior LvOT regions processing higher level lexical or semantic features (Moore and Price 1999; Simons et al. 2003; Price and Mechelli 2005; Vinckier et al. 2007; Levy et al. 2008). The fact that written words are visually distinct from pictures of objects has led to expectations that neuronal activation in posterior LvOT will differ for written words and objects. However, although there is consistent evidence that bilateral fusiform activation is higher for pictures of objects than written words, the evidence that other parts of the LvOT cortex are more strongly activated by written words than pictures of objects is weak and inconsistent (Baker et al. 2007; Wright et al. 2008).

In this paper, we consider 2 hypotheses that might explain why there is currently no clear evidence for an LvOT region that is more selective for written words than pictures of objects during naming tasks. The first hypothesis (A) is that the visual forms of written words and objects are represented by different neuronal populations that are differentially selective to either written words or pictures. However, because these

words and picture-specific neuronal populations lie in close proximity to one another, they cannot be distinguished by functional magnetic resonance imaging (fMRI) (even when the spatial resolution is maximized, e.g., $<2\text{ mm}^3$ as in Wright et al. 2008). The second hypothesis (B) is that written words and pictures of objects activate the same neuronal LvOT populations because of common top-down influences carried by backward connections from higher level amodal areas. This could occur even when words and objects activate different neuronal populations in the bottom-up processing stream because backward connections are more abundant and divergent than forward connections (Friston et al. 2006), and their outputs are not limited to the neuronal populations that drove the higher level responses (Bar et al. 2006; Eger et al. 2007; Gilbert and Sigman 2007; Mahon et al. 2007; Williams et al. 2008; Hon et al. 2009). In contrast to hypothesis A, hypothesis B does not assume that the visual forms of written words and objects are represented by different neuronal populations. Instead, the differences between the visual forms of words and pictures could either be represented by neuronal populations that are specific to stimulus type in the bottom-up direction (Dehaene et al. 2002) or they could be represented by differences in the distributed pattern of responses across shared neuronal populations in LvOT (Price and Devlin 2004). In either case, the neuronal populations are nonspecific (in hypothesis B) because they respond to both written words and pictures of objects.

In Figure 1, schematic illustrations of visual processing are shown at 2 levels of the visual processing hierarchy. Models A1 and A2 are specific to stimulus type because neuronal populations for words and pictures respond independently, at both the lower and higher level in Model A1 and at the lower level in Model A2. In contrast, Models B1, B2, and B3 are not specific to stimulus type because neuronal populations receive common top-down connections in the context of either words or pictures. These 3 nonspecific models differ in whether bottom-up processing activates different neuronal populations for words and pictures (Models B1 and B2) or the same neuronal populations (Models B3). For example, the same neuronal populations could be involved in bottom-up visual processing when there are shared visual features in the word and picture (e.g., the tail of a pig and the letter *e*) or when the word and picture refer to the same concept or have the same name (e.g., CAT and a picture of a cat).

To evaluate evidence for hypothesis A and/or hypothesis B, we used fMRI and a repetition suppression technique that allows neuronal responses within the same voxel to be differentiated (Grill-Spector and Malach 2001; Naccache and Dehaene 2001; Henson and Rugg 2003; Eddy et al. 2006). The observation of interest is the reduction of brain activation at

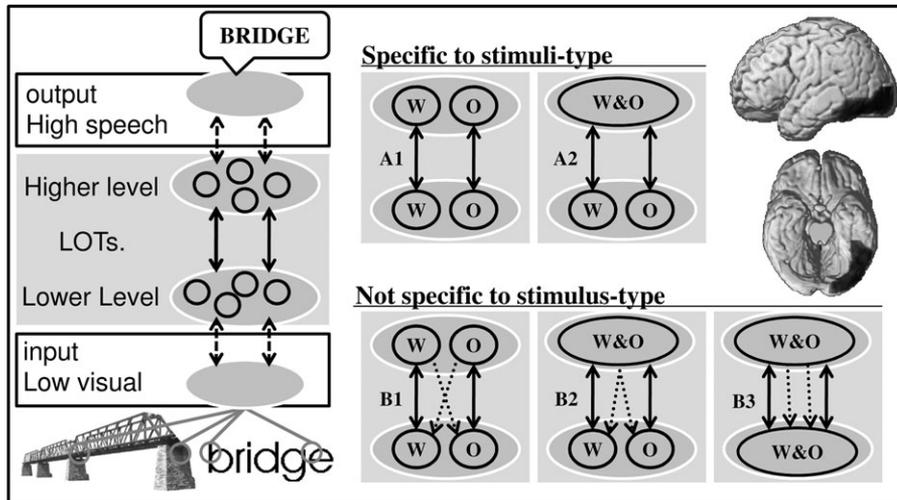


Figure 1. Hierarchical organization of visual processing and models of stimulus type specificity. The left side of the figure illustrates the order of bottom-up processing of written words and pictures during our picture-naming task starting with visual input (bottom row) via posterior LvOT and anterior LvOT to speech output (top row), with many stages between anterior LvOT and speech output that are not considered here. In both the left and right side of the figure, posterior LvOT is illustrated below anterior LvOT and bottom-up processing is illustrated with upward arrows while top-down processing is illustrated with downward arrows. The upper right part of the figure shows our left occipito-temporal region of interest (in black), defined as activation for written words relative to fixation projected on a 3D brain. The rest of figure illustrates 5 different models of stimulus type specificity/nonspecificity in the LvOT cortex. The stimulus type specific models (A1 and A2) respond differently to words and pictures. The stimulus type nonspecific models (B1, B2, and B3) show common responses for words and pictures. Neuronal populations are circled separately for words (W) and pictures (P) when there is stimulus type specificity in the bottom-up processing stream. In contrast, neuronal populations for words and pictures are linked together (W and P) when bottom-up processing (from lower to higher) is not specific to stimulus type. Within stimulus type, connectivity between lower and higher levels is always assumed to be bidirectional (bottom up and top down). Top-down processing across stimulus type is indicated by dotted lines in the 3 models that are not specific to stimulus type (B1, B2, and B3). To summarize the distinction between the models: In A1, neuronal populations are specific to stimulus type at both levels of the hierarchy. In A2, neuronal populations are specific to stimulus type at the lower but not the higher level. In B1, B2, and B3, neuronal populations are not specific to stimulus type because of top-down processing across stimulus type. Nevertheless, there is stimulus type specificity in the bottom-up processing at both levels in (B1) and at the lower but not higher level in (B2).

the voxel level when a target stimulus is preceded by another stimulus (the prime) that shares some level of processing with the target (i.e., the prime and target are related). The rationale for using this technique is that if neuronal populations are specific to either written words or pictures (hypothesis A), then we would expect repetition suppression when the prime and target were presented in the same stimulus type (i.e., word-word or picture-picture) but not when the prime and target were presented in different stimulus types (i.e., word-picture or picture-word).

In our study, we looked for repetition suppression when the prime and target were conceptually identical but not physically identical. In the conceptually identical conditions, the prime and target referred to the same object (e.g., *deer-DEER*), and in the unrelated conditions, the prime and target referred to different objects (e.g., *deer-CHAIR*). The primes and targets were never physically identical because 1) when the prime and target were both words (W-W), the prime was always presented with a different size, font and case to the target; 2) when the prime and target were both pictures (P-P), the prime was always a different exemplar in a different view; and 3) when the prime and target were different stimulus types (W-P and P-W), there was no greater physical similarity between conceptually identical primes (e.g., the word LION presented with a picture of a lion) than between unrelated pairs (e.g., the word LION presented with a picture of a table). Repetition suppression for conceptually identical relative to unrelated W-P or P-W trials could therefore not be explained in terms priming in the bottom-up visual processing stream. Instead, repetition suppression for W-P or P-W trials can only be explained by processing overlap in higher level semantic or phonological areas.

Irrespective of whether hypothesis A or hypothesis B is correct, we expected that repetition suppression would be observed in both posterior and anterior LvOT when word (W) targets were primed with conceptually identical words (W-W) or picture targets were primed by conceptually identical pictures (P-P). This prediction is based on prior literature that has shown repetition suppression in LvOT for conceptually identical W-W pairs when the prime and target were presented in different fonts and cases (Dehaene et al. 2004; Devlin et al. 2006) and for conceptually identical P-P pairs when the prime and target were different pictures of the same objects photographed in different views (Koutstaal et al. 2001; Vuilleumier et al. 2002; Simons et al. 2003). The critical evaluation of our hypotheses (A or B) depended on whether repetition suppression in posterior and/or anterior LvOT was observed when the prime and target were presented in different stimulus types (W-P, P-W).

Testing Hypothesis A

If neuronal populations in either posterior or anterior LvOT are specific to the type of stimulus, then repetition suppression for conceptually identical relative to nonidentical primes (e.g., *deer-DEER* vs. *deer-CHAIR*) should be greater when the prime and target are presented in the same stimulus type (W-W or P-P) relative to when they are presented in different stimulus types (P-W or W-P).

Testing Hypothesis B

If neuronal populations in LvOT are not specific to words or pictures, then we would expect repetition suppression for conceptually identical primes irrespective of the stimulus type

(W-P, P-W, W-W, and P-P). These effects would indicate that words and pictures had activated common semantic or phonological representations that were either in LvOT or in higher order areas that send top-down feedback to LvOT. On the basis of prior studies (Moore and Price 1999; Simons et al. 2003; Price and Mechelli 2005; Wheatley et al. 2005; Gold et al. 2006; Korsnes et al. 2008), we predicted that there might be common semantic representations for words and pictures in the more anterior part of LvOT (in the vicinity of $y = -45$ mm in Montreal Neurological Institute [MNI] space). In contrast, prior literature (Koutstaal et al. 2001; Vuilleumier et al. 2002; Simons et al. 2003; Dehaene et al. 2005; Devlin et al. 2006; Vinckier et al. 2007; Szwed et al. 2009) has shown that the posterior part of LvOT (posterior to $y = -60$ mm, MNI space) extracts visual form information that is not retinotopically bound but “abstracted” away from low level, location bound visual inputs. Repetition suppression in posterior LvOT for conceptually identical primes that were presented in a different stimulus type to the target (W-P or P-W) is therefore unlikely to reflect semantic and phonological processing but would be consistent with top-down feedback from semantic and phonological areas to bottom-up visual processing in posterior LvOT. In summary, repetition suppression in posterior LvOT for conceptually identical primes with visually different forms (i.e., W-P and P-W) would provide evidence for the influence of top-down processing in LvOT (see Fig. 2).

To investigate whether evidence for top-down processing was strategic or automatic, one group of participants were presented with word and picture targets in the context of unmasked (visible) primes (300-ms duration) and another group of participants were presented with the same word and picture targets in the context of masked (unconscious) primes (30-ms duration), see Results section for prime visibility test. The expectation was that masked primes would minimize strategic processing (i.e., when subjects use strategies to predict the target on the basis of the prime). Although visual masking can reduce feed-back processing (Lamme et al. 2002),

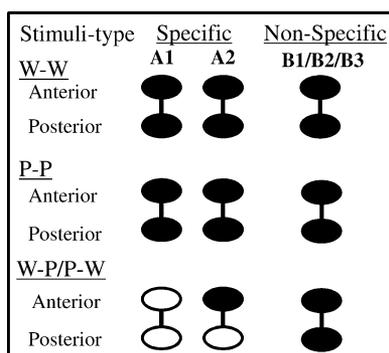


Figure 2. Predictions for activation in posterior and anterior LvOT. Predicted repetition suppression effects in posterior and anterior LvOT for each type of model and for prime-target pairs that are within stimulus type (W-W and P-P) or between stimulus type (W-P and P-W). This results in 3 different predictions for each model in both posterior and anterior LvOT. Black-filled circles indicate predicted repetition suppression. White, unfilled circles indicate no predicted repetition suppression. As can be seen, our predictions within stimulus type are the same for all 5 models. In contrast, our predictions for between stimulus type vary with the model. Model A1 (specific to stimulus type) does not predict any repetition suppression across stimulus type while Models A2, B1, B2, and B3 predict repetition suppression across stimulus type in the anterior LvOT. In addition, Models B1, B2, and B3 predict repetition suppression across stimulus type in posterior LvOT, which is a consequence by top-down processing from higher order language areas.

several studies (e.g., van Gaal et al. 2009) have demonstrated that top-down processing is a fundamental property of the brain irrespective of whether the stimuli are conscious or not conscious. This arises from the brains inherent interactive processing and the influence of abundant backward connections (Friston et al. 2006; Gilbert and Sigman 2007). There are also numerous studies that have shown that masked primes activate higher level semantic and motor processing regions in the feed forward direction (Bodner and Masson 2003; Gold et al. 2006; Sumner and Brandwood 2008).

All participants were instructed to name/read aloud the target stimulus highlighted in a red box (for illustration, see Fig. 3). The naming task ensured closely matched processing with identical output responses for word and picture targets. We could then compare repetition suppression effects for W-W, P-P, W-P, and P-W trials. This contrasts to previous studies that have either looked at repetition suppression in the same stimulus type (Koutstaal et al. 2001; Naccache and Dehaene 2001; Vuilleumier et al. 2002; Simons et al. 2003; Eddy et al. 2007) or focused on repetition suppression at the semantic association level (Kircher et al. 2009).

Materials and Methods

Participants

The study was approved by the National hospital and Institute of Neurology. Thirty volunteers (right-handed English native speakers with no history of neurological or psychiatric illness or language

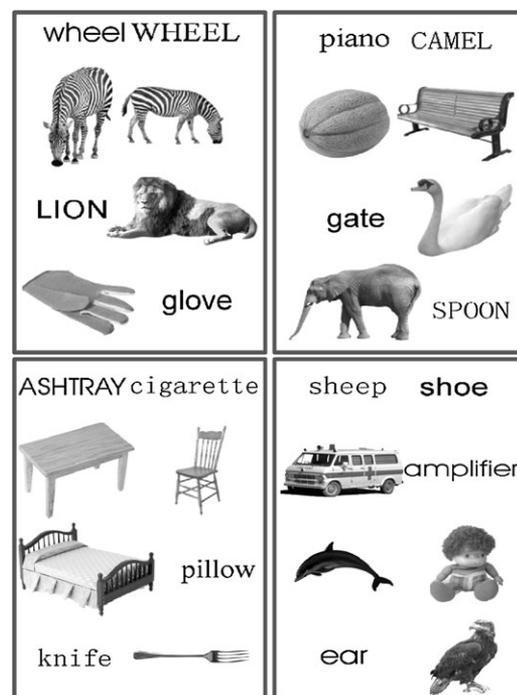


Figure 3. Examples of primes and targets. The 4 panels show examples of 4 types of relationship between the prime and the target and the 4 stimulus type pairings (W-W, P-P, W-P, P-W). Within each panel, the first row: prime = word, target = word, the second row: prime = picture, target = picture, third row: prime = word, target = picture, fourth row: prime = picture, target = word. Top-left panel = conceptual identity (same response): wheel-Wheel, zebra-ZEBRA, lion-LION, and glove-GLOVE. Bottom-left = semantic relationship: ashtray-CIGARETTE, table-CHAIR, bed-PILLOW, and knife-FORK. Top-right panel = unrelated: piano-CAMEL, melon-BENCH, gate-SWAN, and elephant-SPOON. Bottom-right = phonological relationship: sheep-shoe, ambulance-AMPLIFIER, dolphin-DOLL, and ear-EAGLE.

impairments) gave their informed consent to take part in this experiment. Six participants were excluded due to motion artifacts. The age of the remaining 24 participants (12 males and 12 females) ranged from 20 to 43 years (mean = 27). For 11 participants (5 females and 6 males), the prime was presented for 300 ms, and for 13 participants (7 females and 6 males), the prime was presented for 33 ms (2 frames) and preceded by forward and backward masks.

Material and Design

The stimulus set of words and pictures was identical to that reported in Mechelli et al. (2007) but the prime durations were much briefer, and participants were instructed to focus on the target stimulus only. Pictures were of real objects with the highest intersubject naming agreement. Words were the written names of the same objects and were therefore matched to the pictures for conceptual identity and naming response. Pairs of stimuli were created to form a prime and target pair according to a $2 \times 2 \times 4$ factorial design (i.e., 16 conditions) with 3 factors: stimulus type of the prime (word or picture), stimulus type of the target (word or picture), and the type of similarity between the prime and the target, see Figure 3. The 4 types of prime-target relationship were: unrelated (e.g., KITE-lobster), conceptually identical (e.g., BELL-bell), semantically related by association (e.g., ROBIN-Nest), or the same category membership (e.g., COW-bull) and phonologically related because the first phoneme was the same (e.g., BELL-belt). Because of the close relationship between orthography and phonology in English, phonologically related prime-target pairs were also orthographically related for word pairs. This might introduce differences between words and pictures in the phonological condition but would not be a confound if we observed common effects for words and pictures. In all prime-target pairs, we minimized physical similarity between the prime and target, see Figure 3. This was achieved for 1) W-W pairs by presenting the prime and target in different sizes, fonts and cases; 2) P-P pairs by presenting different object exemplars in different views; and 3) there was no physical similarity between any W-P and P-W pairs.

To avoid item-specific confounds and fully control for stimulus features across conditions, each object concept (prime or target) was fully counterbalanced across participants. For example, the target “crab” was presented to one set of participants with a prime that had an unrelated response (e.g., “slide”), to a second set of participants with a prime that had an identical response (i.e., crab), to a third set of participants with a semantic prime (e.g., “lobster”), and to a fourth set of participants with an orthographically/phonologically related prime (e.g., “crane”). Therefore, although all participants saw the same set of stimuli, different subsets of participants were exposed to different pairings of the stimuli.

Procedure

Stimulus presentation was via a video projector, a front-projection screen and a system of mirrors. Participants’ verbal responses were recorded and filtered using a noise cancellation procedure to monitor accuracy and response times.

The intertrial interval was held constant at 3200 ms, and the target duration was always 300 ms followed by 2400 ms fixation (see Fig. 4). However, the remaining 500 ms was occupied in 2 different ways. In 11 participants, the prime was presented for 300 ms, followed by a fixation cross for 200 ms. In 13 participants, the prime was presented for 33 ms (2 frames) and preceded by a forward mask for 200 ms and a backward mask for 267 ms. The duration of these forward and backward masks aimed to keep the interstimulus interval constant in the masked and unmasked conditions. The masked prime duration (33 ms) was chosen to minimize conscious awareness of the prime while being consistent with word and picture primes. We ran the risk that 33 ms is atypically short for a picture prime that may have reduced the chance of observing a priming effect. Fortunately, our fMRI results showed an effect of priming irrespective of whether the primes were words or pictures and irrespective of whether the primes were masked or unmasked. Therefore, the very short prime duration was not a disadvantage.

The target was presented within a red square, and the participants were instructed to read or name the target aloud as soon as it appeared on the screen while ignoring the prime in the unmasked conditions. In the masked conditions, participants were not informed that the stimuli to be named were preceded with a prime. The participants were trained to whisper their responses and to minimize jaw and head movements in the scanner. Participants’ verbal responses were recorded to monitor accuracy and response times.

Prime Visibility Tests

The visibility of the masked primes was investigated using the same stimulus set and settings as those used in the masked version of the fMRI experiment but with a different group of 16 participants (8 males, mean age 32 years). Participants were informed of the presence of the prime and they were asked to judge whether or not the prime and target referred to the same object. The prime visibility was assessed using d' (Van den Bussche et al. 2009).

fMRI Data Acquisition

Volunteers were scanned in a 3-T head scanner (Magnetom Allegra, Siemens Medical) operated with its standard head transmit-receive coil. A single-shot gradient-echo echo planar imaging (EPI) sequence was used with the following imaging parameters: 35 oblique transverse slices, slice thickness = 2 mm, gap between slices = 1 mm, time repetition

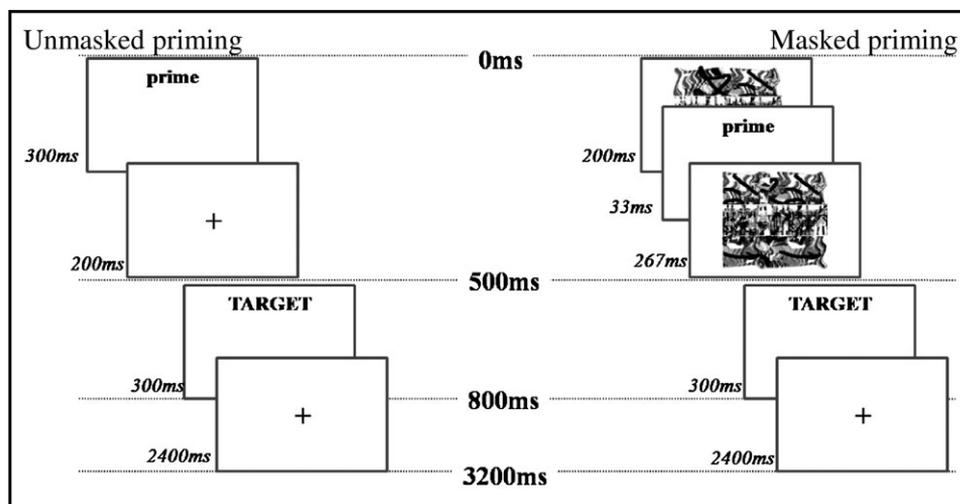


Figure 4. Timeline of the experiment. Procedure for unmasked (left column) and masked (right column) conditions. In both, the participants read or named the target presented for a duration of 300 ms. In the unmasked condition, the prime was presented for 300 ms, followed by a fixation cross for 200 ms. In the masked condition, the prime was presented for 33 ms (2 frames) and preceded by a forward mask for 200 ms and a backward mask for 267 ms.

(TR) = 2.275 s, $\alpha = 90^\circ$, time echo (TE) = 30 ms, matrix size = 64×64 . EPI data acquisition was monitored online using a real-time reconstruction and quality assurance system (Weiskopf et al. 2007). A total of 836 volumes were acquired in 2 separate runs, and the first 6 images (dummy scans) of each run were discarded to allow for T_1 -equilibration effects.

At the end of the functional runs, whole-brain structural scans were acquired using a Modified Driven Equilibrium Fourier Transform sequence. For each volunteer, 176 sagittal partitions were acquired with field of view = 256×240 , matrix = 256×240 , isotropic spatial resolution = 1^3 mm, TE = 2.4 ms, TR = 7.92 ms, flip angle $\alpha = 15^\circ$, time to inversion = 910 ms, 50% time to inversion ratio, fat sat angle = 190° , flow suppression angle = 160° , bandwidth = 195 Hz/pix. Acquisition time = 12 min.

Scanning Procedure

Data were acquired in 2 separate sessions, each including 200 trials (25 trials for each of the 16 conditions, over 2 scanning sessions) plus 100 null events (fixation cross). All conditions were presented within session in a randomized order so that the participants could not anticipate the type of prime-target relationship. In the first session, half the participants saw one set of objects as pictures and the other set of objects as written object names. In the second set, the same prime-target pairs were presented again but objects presented as written names in the first session were presented as pictures in the second session, whereas objects presented as pictures in the first session were presented as written names in the second session. The order of trials in the second session was different to that used in the first session, therefore participants could not predict what the target object would be. Although we expected long-term adaptation/priming effects to be consistent for targets with related and unrelated primes, there was also the possibility that the effect of priming might vary across session. We therefore ensured that our effects of interest were present during both the first and second session.

Data analysis

Data preprocessing and statistical analyses were performed with SPM5 (Wellcome Trust Centre for Neuroimaging, London, UK). Spatial transformations include realignment (movement artifacts correction), unwarping, normalization to the MNI space, and spatial smoothing (isotropic 6-mm full-width at half-maximum). A temporal high-pass filtering (1/128 Hz cutoff) was applied.

First-Level Statistical Analyses for Each Individual

Each stimulus pair (prime and target) was treated as a single trial because our focus was on neuronal responses to target stimuli that differed only as a function of the prime-target relationship. Correct trials were assigned to their specific experimental condition. One predictor variable for each of the experimental conditions was obtained by a convolution operation of the onset times of the trials with correct responses and the canonical hemodynamic response function. All incorrect responses (including null responses or unexpected responses) were modeled as an additional regressor. The contrast images for each first-level analysis are described below in the context in which they are reported in the Results section.

Group Level within Subjects ANOVAs

Our main effects of interest, at the group level, came from 2 second level within subjects analysis of variances (ANOVAs) that were computed separately for the 11 participants who were presented with unmasked primes and the 13 participants who were presented with masked primes. In each of these ANOVAs, there were 16 conditions that contained the first-level contrast images corresponding to each condition relative to fixation.

Preliminary investigation of these second-level ANOVAs demonstrated that LvOT activation did not differ for semantic, phonological, or unrelated prime-target pairs ($P > 0.05$ uncorrected). Likewise, Mechelli et al. (2007) found no evidence for semantic or phonological priming effects in LvOT during naming and reading aloud with the same prime-target stimuli. Therefore, we treated semantically and phonologically related items and all unrelated items as “nonidentical.” We then defined

our repetition suppression effect as activation that was reduced for conceptually identical relative to nonidentical primes.

The Results section reports the following effects from the group level within subjects ANOVAs

The Effect of Stimulus Type on the Targets

This involved a direct comparison of all reading aloud word targets and all picture naming targets (irrespective of prime).

The Effect of Stimulus Type on the Primes

This involved a direct comparison of stimuli with word primes and stimuli with picture primes (irrespective of target).

Repetition Suppression That Depends on Whether the Prime and Target Were Within Stimulus Type or Between Stimulus Type (to Test Hypothesis A)

This corresponded to the interaction of (repetition suppression) and (within vs. between stimulus type priming). We tested this analysis over all conditions and for words only (W-W) and pictures only (P-P).

Repetition Suppression That Does Not Depend on Whether the Prime and Stimulus Are Within or Between Stimulus Type (to Test Hypothesis B)

This corresponded to the main effect of repetition suppression where there was also an effect of repetition suppression within stimulus type (W-W and P-P) and between stimulus type (W-P and P-W).

Individual Results from the First-Level Analyses

Number of Participants Showing Repetition Suppression across Stimulus Type

This refers to the number of subjects whose first-level analysis showed a main effect of repetition suppression computed over stimulus type (i.e., summing over W-W, P-P, W-P, and P-W).

Number of Participants Showing Repetition Suppression That Was Specific to Stimulus Type

This refers to the number of subjects whose first-level analysis showed an interaction of (repetition suppression) and (within vs. between stimulus type priming).

Group Level between Session or Subject ANOVAs

The Effect of Masked Versus Unmasked Priming on Repetition Suppression

The between subjects ANOVA compared repetition suppression in the masked and unmasked priming conditions. For each subject, we included the contrast measuring repetition suppression for within stimulus type repetition suppression separately from between stimulus type repetition suppression. Thus, there were 4 conditions, one between subjects and one within subjects.

The Effect of First or Second Scanning Session on Repetition Suppression

A between session ANOVA compared repetition suppression in the first and second sessions.

The aim was to ensure that our results were not driven by the second session (when words from the first session were presented as pictures and pictures from the first session were presented as words).

Search Volume and Locations of Interest

We defined a search volume that included all left occipito-temporal voxels that were activated for words only (W-W summed over all prime-target relationships) relative to fixation ($P < 0.05$ corrected for multiple comparisons across the whole brain using family-wise error correction), see Figure 1. This is consistent with prior studies (Devlin et al. 2006; Duncan et al. 2009) and includes the region that has been labeled the “visual word form area” (Cohen et al. 2000, 2002). Nevertheless, the same search volume can also be redescribed as those voxels that were commonly activated for words and pictures. This is

because all the LvOT voxels that were activated for words were also equally activated for pictures. Within this search volume, we identified the coordinates of peak activation in the posterior occipito-temporal sulcus and also in the most anterior voxel and we did so separately for the masked and unmasked analysis (see Table 1). We then searched for repetition suppression effects at these coordinates or within 5 mm of these coordinates and report effects that were significant after correction for multiple comparisons in these small volume searches. Critically, the contrast used to identify the region of interest (W-W and P-P relative to fixation) was independent of the contrasts used to identify repetition suppression (i.e., conceptually identical vs. nonidentical).

Statistical Threshold

For identifying our LvOT region of interest, our statistical threshold ($P < 0.05$) was corrected for multiple comparisons across the whole brain. For repetition suppression effects, our correction for multiple comparisons was based on spherical search volumes (2- and 5-mm radius) centered on the coordinates from the peak activations identified in the posterior and anterior parts of the region of interest (see above and Table 1).

Statistical Power

To find evidence for hypothesis A ("stimulus type specificity in LvOT"), our analysis focused on evidence that repetition suppression in LvOT was greater within stimulus type (i.e., W-W or P-P) than between stimulus type (i.e., W-P or P-W). If this effect was not significant, then we have a null result which could result from a lack of power. Critically, however, evidence for hypothesis A was expected to be unidirectional (more repetition suppression in LvOT when primes were within than between stimulus type). As we observed a trend in the opposite direction (more repetition suppression in the LvOT region of interest for primes that were between than within stimulus type), we were not motivated to include more subjects. In addition, evidence for hypothesis B ("top-down processing irrespective of stimulus type") was based on repetition suppression in posterior LvOT when the primes and targets were presented in different stimulus types. A

significant effect in this context, during the masked priming conditions, indicated that there was sufficient power to detect top-down processing, particularly since this effect was replicated in 2 groups of participants (total = 24 subjects) and in 21/24 individuals. Therefore, there was no motivation to include more subjects.

Results

In-scanner Behavior

Mean accuracy for picture naming was 96.9% in the masked conditions and 96.2% in the unmasked presentations. Mean accuracy for reading aloud was 99.8% in both the masked and unmasked presentations. The difference between accuracy for naming and reading was significant in both the unmasked ($F_{1,10} = 231.1$, $P < 0.0001$) and masked ($F_{1,12} = 745.3$, $P < 0.0001$) conditions, but there was no effect of word versus picture primes (unmasked presentation: $F_{1,10} = 0.87$, $P < 0.44$; masked presentation: $F_{1,12} = 0.01$, $P = 0.918$) and no effect of prime-target relationship (unmasked: $F_{3,30} = 1.88$, $P < 0.11$; masked: $F_{3,36} = 1.7$, $P = 0.19$).

It was not possible to obtain an accurate estimation of the response times because the voice onset times in our auditory recordings were corrupted by the sound of the scanner and therefore did not have sufficient resolution to detect the 8- to 16-ms priming effects that were expected on the basis of prior literature (Ferrand et al. 1994; Alario et al. 2000; Van den Bussche et al. 2009). However, as explained in the discussion, our conclusions are based on robust neural repetition suppression effects. Although it would be interesting to add the corresponding behavioral effects, their presence or absence would not alter our conclusions.

Masked Prime Visibility Tests

In the same-different concept task, the mean d' values for the conditions were 0.031, 0.062, 0.14, and 0.078, respectively for W-W pairs, W-P pairs, P-W pairs, and P-P pairs (see Fig. 5). Although none of the participants scored better than chance, the mean d' value approached significance for the P-W pairs ($t_{15} = 1.9$, $P = 0.066$) indicating a greater than zero effect of the picture prime. Nevertheless, this value is similar to those reported in the subliminal priming literature for (e.g., Kouider et al. 2007) and below the values reported for supraliminal stimuli. None of the other mean d' values were significantly different from zero (W-W pairs, $t_{15} = 0.36$, $P > 0.7$; W-P pairs, $t_{15} = 0.77$, $P > 0.45$; P-P pairs: $t_{15} = 1.15$, $P > 0.15$). A repeated measure ANOVA with 2 within subject factors found no significant difference in d' for word versus picture primes ($F_{1,16} = 0.87$, $P = 0.36$), word versus picture targets ($F_{1,16} = 0.08$

Table 1
Peak activation in posterior and anterior LvOT

	Posterior LvOT		Anterior LvOT	
	Coordinates	Z score	Coordinates	Z score
(a) Activation for words and pictures (defining the regions of interest)				
Unmasked	-46, -62, -16	>8	-42, -36, -18	5.7
Masked	-38, -62, -18	>8	-40, -44, -14	5.9
(b) Repetition suppression in the unmasked conditions				
Main effect	-46, -62, -18	3.5	-40, -36, -18	2.7
	-48, -58, -18	3.9	-38, -38, -20	3.7
Within modality	-48, -62, -16	2.5	-40, -36, -18	1.7
	-48, -66, -18	2.7	-38, -38, -20	2.2
Between modality	-46, -60, -16	2.5	-40, -36, -18	2.0
	-46, -58, -18	3.0	-38, -38, -20	2.9
(c) Repetition suppression in the masked conditions				
Main effect	-38, -62, -16	3.1	-42, -44, -14	3.3
	-36, -60, -14	3.5	-42, -44, -14	3.3
Within modality	-40, -62, -18	2.7	-42, -44, -14	2.8
	-40, -58, -16	2.9	-44, -46, -16	3.1
Between modality	-38, -62, -16	1.9	-40, -44, -14	2.4
	-36, -60, -20	2.5	-40, -44, -12	2.7

Note: Peak coordinates (in MNI space) and Z scores are provided in posterior and anterior LvOT for the unmasked and masked conditions separately. (a) activation for words (defining the regions of interest) refers to the main effect of all word conditions (irrespective of prime) relative to fixation, with a correction for multiple comparisons across the entire brain. (b) and (c) show repetition suppression effects when the prime cued versus did not cue the correct response. They are reported summed over within and between stimulus type primes (i.e., the main effect of repetition suppression) and for the within and between stimulus type primes separately. In each of these contexts, 2 sets of coordinates and Z scores are provided. The top row reports those within 2 mm of the regions of interest, the second row are those within 5 mm of the regions of interest. All effects were significant after a correction for multiple comparisons in height, across the whole brain for the region of interest at the top of the table (a) or at the coordinates of the regions of interest for the repetition suppression results (b and c).

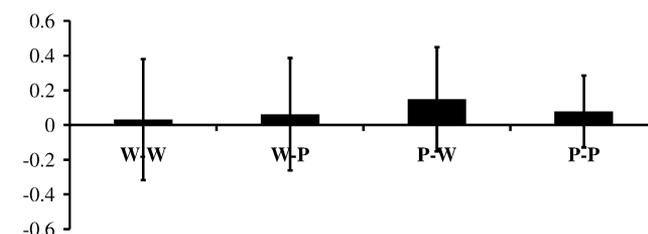


Figure 5. Prime visibility test. The figure shows the discrimination performance reported as d' (signal detection measure) for the prime visibility test. For each stimulus type pairing (W-W, P-P, W-P, P-W), the d' value was not significantly different from zero.

$P = 0.78$), or the interaction of these factors which tests for a difference in between versus within stimulus type ($F_{1,16} = 0.412$ $P = 0.53$). These results confirmed that even when aware of the presence of the prime, the participants were not able to match the prime stimuli at a conceptual level.

fMRI Results

The Effect of Stimulus Type on the Targets

As expected, reading (i.e., word targets) and naming (i.e., picture targets) resulted in common activation along the posterior to anterior hierarchy in the vicinity of the left occipito-temporal sulcus during both the masked and unmasked presentations. In accordance with previous reports (e.g., Wright et al. 2008), a direct comparison of reading aloud to picture naming (W-W and P-W > W-P and P-P) did not identify increased activation in any left occipito-temporal voxels, even when the threshold was reduced to $P < 0.05$ uncorrected and even when the contrast was within stimulus type (W-W > P-P). In contrast, there were 1901 and 1448 left occipito-temporal voxels (in the unmasked and masked conditions, respectively) that were more activated for picture naming than reading at $P < 0.001$. As shown previously (Wright et al. 2008), the peak coordinates for these picture selective effects were identified in the fusiform gyrus, which is medial to the left occipito-temporal sulcus (e.g., $x = -30$, $y = -62$, $z = -14$ / $x = +32$, $y = -60$, $z = -12$) or in a lateral occipital-temporal area (e.g., $x = -44$, $y = -76$, $z = 0$ / $x = +38$, $y = -80$, $z = 12$) which is in dorsal rather than ventral occipito-temporal cortex.

The Effect of Stimulus Type on the Primes

We also investigated the effect of stimulus type on the primes. In the unmasked conditions, activation increased ($P < 0.001$ uncorrected) for picture primes relative to word primes in: left and right lateral occipito-temporal cortex in the context of both picture targets (P-P-W-P: $x = +52$, $y = -70$, $z = 0$ / $x = -54$, $y = -66$, $z = 0$) and word targets (P-W-W-W: $x = +52$, $y = -66$, $z = -8$ / $x = -48$, $y = -68$, $z = -4$) and in the right ventral occipito-temporal cortex in the context of word targets (P-W-W-W: $x = +40$, $y = -52$, $z = -16$ / $x = -32$, $y = -54$, $z = -16$). Thus, the effect of stimulus type on unmasked primes was similar to the effect of stimulus type on the targets. Critically, however, we did not identify any voxels where there was an effect of stimulus type on masked primes, even when the threshold was lowered to $P < 0.05$ uncorrected. This is consistent with the prime visibility tests and suggests that our masking procedures were effective for both pictures and words.

Repetition Suppression That Depends on Whether the Prime and Target Were Within Stimulus Type or Between Stimulus Type (to Test Hypothesis A)

We found no evidence for either word or picture-specific repetition suppression in our LvOT region of interest or any other brain region ($P > 0.05$ corrected for multiple comparisons across the whole brain). Thus, there was no significant difference for repetition suppression within stimulus type than between stimulus type (W-W and P-P > W-P and P-W), even when the statistical threshold was lowered to $P < 0.05$ uncorrected in our LvOT region of interest or any fusiform or occipito-temporal voxels outside our region of interest. We also found no differences in the effect of repetition suppression within stimulus type for words only (W-W) and pictures only (P-P).

This is not surprising because all our primes and targets were physically different from one another even within stimulus type (for details, see Materials and Methods and Fig. 3).

Repetition Suppression That Does Not Depend on Whether the Prime and Stimulus Are Within or Between Stimulus Type (to Test Hypothesis B)

A main effect of repetition suppression was observed in both posterior and anterior LvOT, and this was consistent irrespective of whether the prime was within or between stimulus type and irrespective of whether the prime was 300 ms and unmasked or 33 ms and masked (see Figs 6 and 7 and Table 1). There was no significant interaction between repetition suppression and stimulus type. It is interesting to note, however, that when we reduced the statistical threshold to $P < 0.05$ uncorrected, we identified 94 voxels (with peak coordinates at $x = -38$, $y = -58$, $z = -18$) in the unmasked conditions, where repetition suppression was greater “between stimulus type” (W-P, P-W) than “within stimulus type” (W-W, P-P). Although the low Z score for this effect ($Z = 2.0$) does not allow us to distinguish it from noise, it is pertinent that repetition suppression was higher between than within stimulus type because this is not consistent with hypothesis A (stimulus specificity).

Evidence for the influence of top-down processing (hypothesis B) came from the observation that repetition suppression was observed, between stimulus type, throughout posterior and anterior LvOT for both the masked and unmasked

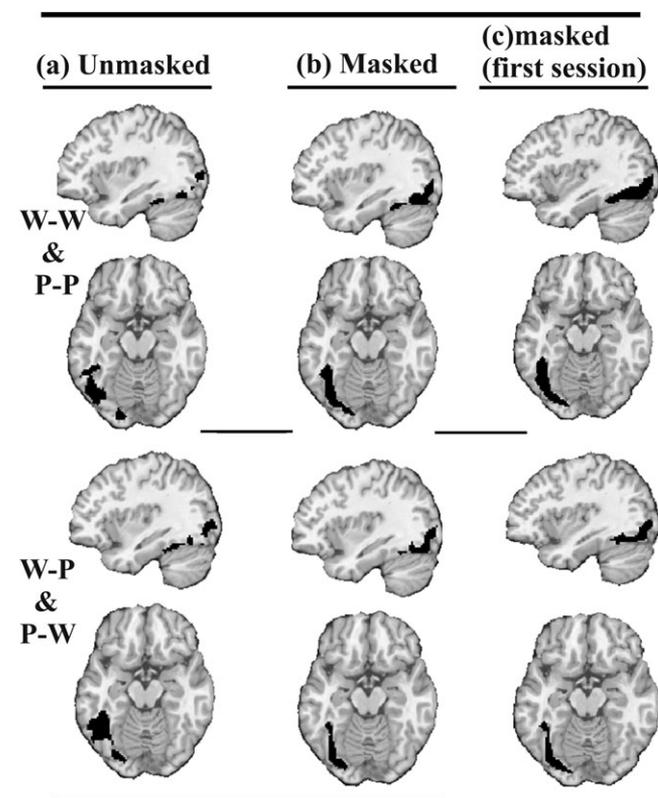


Figure 6. Reduced activation when the prime cued the response. Repetition suppression in LvOT ($P < 0.05$ uncorrected) illustrated in black on axial and sagittal slices with $z = -14$, $x = -44$ in MNI space, for within stimulus type pairs (W-W, P-P) and between stimulus type pairs (W-P, P-W) when the primes were (a) unmasked, (b) masked, and (c) masked first session only (for details of LvOT peaks, see Table 1).

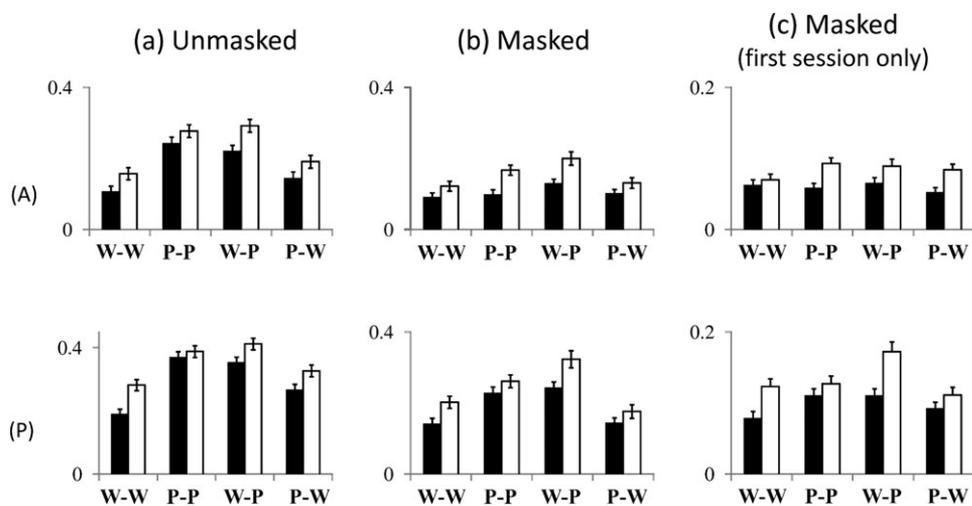


Figure 7. Contrast estimates. Repetition suppression in LvOT, irrespective of stimulus type. Contrast estimates (effect size on the y axis), at the maxima of Table 1 (A: anterior LvOT, P: posterior LvOT), for prime–targets that were conceptually identical (black bars) or non-identical (white bars) for each stimulus type pairing (W-W, P-P, W-P, P-W) when the primes were (a) unmasked across both sessions, (b) masked across both sessions, and (c) masked using data from the first session only. The black bars illustrate consistently reduced activation when the prime was conceptually identical to the target and therefore cued the same response. The white bars show activation when the primes were non-identical.

conditions (see Table 1 and Figs 6 and 7). This result cannot be explained by bottom-up access to abstract representations of the visual form of the stimulus because the primes and targets of W-P and P-W stimuli were not more visually similar for conceptually identical than non-identical pairs. The results are, however, consistent with Models B1, B2, and B3 in Figure 1.

Number of Participants Showing Repetition Suppression across Stimulus Type

To validate our findings further, we examined repetition suppression in each of our participants independently. This allowed us to exclude the possibility that our group findings were a consequence of poor spatial resolution when voxel values were averaged over different individuals. When the primes were unmasked, a main effect of repetition suppression in LvOT was observed in 10 out of the 11 participants with individual *Z* scores ranging from 4.0 in the most significant case to 2.0 in the least significant case with a mean *Z* score of 2.9. In the masked conditions, repetition suppression in LvOT was observed in 10 out of the 13 participants, with individual *Z* scores ranging from 5.3 in the most significant case to 1.7 in the least significant case with a mean *Z* score of 2.9.

Number of Participants Showing Repetition Suppression That Was Specific to Stimulus Type

In contrast to the relatively consistent and robust effects of repetition suppression across conditions, there was no evidence that this effect was specific to stimulus type. In the unmasked version of the experiment, only 2/11 participants showed more repetition suppression when the primes were within relative to between stimulus type and the *Z* scores for these effects were exceedingly low (*Z* = 1.7 and 1.9; *P* = 0.05 uncorrected). Moreover, a third participant showed the reverse effect (more repetition suppression between than within stimulus type, with a higher *Z* score of 3.0; *P* < 0.005 uncorrected). These individual effects are not interpretable because they are inconsistent across participants and do not survive a correction for multiple comparisons (see Wright et al.

2008). Likewise, in the masked version of the experiment, the comparison of within versus between stimulus type primes only revealed inconsistent and sparsely located voxels even when the threshold was lowered to *P* < 0.05 uncorrected: In 3/13 subjects, there was a trend for more repetition suppression within than between stimulus type with mean coordinates at (*x* = -42, *y* = -67, *z* = -18) and a mean *Z* score of 2.0 (range *Z* = 1.8–2.5). In contrast, 7/13 subjects showed more repetition suppression between than within stimulus type. In 3 of these 7 subjects, the mean coordinates were located at (*x* = -43, *y* = -67, *z* = -11) with *Z* scores ranging from 1.8 to 2.4. In the other 4 subjects, the mean coordinates were more anterior at (*x* = -48, *y* = -54, *z* = -16) with a mean *Z* score of 2.7 (range *Z* = 2.2–3.0). Therefore, as in the unmasked conditions, there was no consistent or interpretable evidence, at the single subject level, that repetition suppression was greater within than between stimulus type.

The Effect of Masked Versus Unmasked Priming on Repetition Suppression

The between subjects ANOVA comparing repetition suppression in the masked and unmasked priming conditions did not identify a significant interaction of LvOT repetition suppression and prime duration (*P* < 0.05 uncorrected). However, when averaged across conditions (related and unrelated), we found that the effect of stimulus type on primes (P-W + P-P) > (W-W + W-P) was greater in the unmasked condition than the masked condition (*x* = -30, *y* = -48, *z* = -16; *Z* score = 5.99; *x* = 52, *y* = -68, *z* = -2; *Z* score = 5.69; *x* = -38, *y* = -80, *z* = -16; *Z* score = 5.53). This effect of masking demonstrates that, as intended, the primes were more visible in the unmasked condition than the masked condition.

The Effect of First or Second Scanning Session on Repetition Suppression

The between session ANOVA comparing repetition suppression in the first and second sessions did not identify a significant interaction of LvOT repetition suppression and session (*P* > 0.05

uncorrected). More importantly, for the first session only, there was a main effect of repetition suppression ($x = -38$, $y = -50$, $z = -16$; Z score = 3.64) that was consistent both between and within stimulus type (see Figs 6c and 7c). This demonstrates that our results were not driven by concept repetition between the first and second session.

Discussion

In this study, we used an fMRI adaptation technique to examine whether LvOT activation was selective to written words relative to pictures of objects and, if not, whether common responses to words and pictures could be explained by the influence of top-down processing. At both the group and individual level, we found repetition suppression for conceptually identical relative to nonidentical prime-target pairs in both the posterior and anterior LvOT. This was observed irrespective of whether the prime-target pair was within stimulus type or between stimulus type and irrespective of whether the primes were masked or unmasked.

These results do not provide any evidence that LvOT neuronal populations are specific to stimulus type. Nor can repetition suppression in posterior LvOT be explained in terms of shared bottom-up processing of the visual stimulus because there was no enhanced visual similarity for our conceptually identical versus nonidentical trials when the primes and targets had different stimulus types. It is also highly unlikely that repetition suppression across words and pictures in posterior LvOT reflected bottom-up processing of conceptual or phonological information because this claim would be inconsistent with many years of prior investigation that has consistently shown posterior LvOT activation associated with visual processing of the stimulus rather than the concept it is representing (Koutstaal et al. 2001; Vuilleumier et al. 2002; Simons et al. 2003; Dehaene et al. 2005; Devlin et al. 2006; Xue et al. 2006; Vinckier et al. 2007; Xue and Poldrack 2007; Szwed et al. 2009). Most specifically, our data would be inconsistent with Devlin et al. (2006) who observed repetition suppression in posterior LvOT for visually similar word forms that were semantically dissimilar (corn-CORNER) and no significant suppression in the same region for semantic (notion-IDEA) or phonological (solst-SOLST) repetition. In short, the interpretation of our data requires a model that includes the influence of top-down modulations (Fig. 1B1, B2, and B3) and explains previous results (e.g., Devlin et al. 2006).

The role of top-down processing in repetition suppression has previously been described in what is referred to as the “predictive coding account.” This proposes that repetition suppression is a consequence of the prime improving the top-down expectation of the target and thereby reducing prediction error (Rao and Ballard 1999; Angelucci et al. 2002; Friston et al. 2006). Here, prediction error in LvOT refers to the mismatch between bottom-up inputs in the forward connections and top-down inputs from the backward connections. The mismatch arises because the presentation of the prime sets up a prediction for the response to subsequent inputs, and this prediction is carried by the backward connections from higher to lower levels in the hierarchy (see Fig. 1: B1, B2, B3). Predictions will be better when the prime has the same response as the target than a different response to the target. Conversely, prediction error will be higher when the prime is unrelated to the target than when the prime and the target have the same

response. Higher LvOT activation can therefore be explained by higher prediction error. This account explains why we observed repetition suppression throughout posterior and anterior parts of LvOT even when the primes had no perceptual similarity with the target (W-P, P-W). It is also consistent with the data and implications from Devlin et al. (2006) and Xue et al. (2006) and Xue and Poldrack (2007). Note that, in this context, prediction does not imply a conscious strategy. It refers to the brains inherent interactive processing and the influence of abundant backward connections (Friston et al. 2006; Gilbert and Sigman 2007).

Our results and the predictive coding account support the idea that top-down modulations, via backward connections, integrate bottom-up perceptual evidence from early visual cortex with prior knowledge and contextual information (from primes, task requirements, and attention). This perspective challenges the traditional view that visual recognition is purely the result of bottom-up processing where increasingly complex features are extracted from early visual areas to higher level brain regions. However, our explanation is consistent with that recently provided by Norris and Kinoshita (2008) who have proposed a Bayesian (top-down) framework for explaining unconscious cognition and visual masking.

The influence of top-down processing across stimulus types also offers a plausible explanation for why fMRI studies have not shown LvOT activation for reading relative to picture naming. However, our study does not indicate whether LvOT neuronal populations are specific to stimulus type in the bottom-up direction (Models B1 and B2 in Fig. 1) or whether differences between the visual forms of words and pictures are represented in LvOT by distributed patterns of responses across shared neuronal populations (Model B3 in Fig. 1). This could be investigated at the voxel level using repetition suppression if it was possible to manipulate the visual similarity between words and pictures (Szwed et al. 2009). An alternative approach would be to use multivariate pattern recognition (Haxby et al. 2001; Hanson et al. 2004; Simon et al. 2004) or the better temporal resolution as provided by magnetoencephalography or electroencephalography (Garrido et al. 2007). Evidence for Model B1 would entail stimulus type specific activation patterns or connectivity in posterior and anterior LvOT. Evidence for Model B2 would be stimulus type specific activation patterns or connectivity in posterior but not anterior LvOT. In the absence of evidence for Models B1 or B2, Model B3 becomes the most likely candidate but positive evidence for Model B3 may require intracellular recordings.

The nature of the top-down influences we observed is further informed by considering the spatial location of the repetition suppression effect and the duration of the prime. First, we note that repetition suppression across stimulus types extended all the way along the length of posterior and anterior parts of our LvOT region of interest (see Fig. 6). This suggests that top-down influences may be distributed throughout the left occipito-temporal hierarchy although we cannot exclude the possibility that across stimulus type repetition suppression in anterior LvOT was driven bottom up by amodal semantic processing. Second, repetition suppression across stimulus types was robust even in the context of brief and fully masked primes (see Figs 6 and 7). This suggests that the top-down influences across stimulus type did not depend on the participant identifying the prime or using an explicit strategy to predict the target.

Our results also demonstrate that the LvOT repetition suppression effects were specific to conceptually identical prime–target pairs with no significant differences between phonological, semantic, and unrelated primes. This suggests that the most influential factor is whether the prime predicts the response, in our case name retrieval, which requires a very precise conceptual identification. Semantically related primes may, however, have a more significant effect on LvOT activation during semantic categorization tasks (Gold et al. 2006) rather than our naming task. In brief, we are suggesting that the reason that we did not see evidence for neuronal adaptation for semantic association priming is that we used a naming task that requires concise conceptual identification whereas previous studies showing neuronal adaptation for semantic association primes have used semantic or lexical decisions that do not require a concise conceptual identification (e.g., Gold et al. 2006). The influence of the task on the type of priming has been demonstrated by others (e.g., Nakamura et al. 2007; Norris and Kinoshita 2008) and is consistent with the predictive coding explanation of our data.

Our overt speech task ensured that our participants engaged phonological retrieval processes for both written words and pictures. This reduces differences in the degree to which word or picture primes are processed implicitly at the phonological level. The disadvantage of using an overt speech task was that we were not able to extract accurate voice onset times because of the noise level in the scanner. We therefore do not know how response times varied across priming conditions. Nevertheless, we can report that previous studies using finger-press responses have shown that response times are reduced 1) in the context of repetition suppression in LvOT (Devlin et al. 2006; Gold et al. 2006) and 2) for related compared with unrelated masked primes presented within and between stimulus type during reading and picture naming tasks (Durso and Johnson 1979; McCauley et al. 1980; Ferrand et al. 1994; La Heij et al. 1999; Alario et al. 2000). More critically, if behavioral priming across stimulus type were asymmetrical (e.g., greater for word than picture primes), this would bias our results toward greater repetition suppression in the context of word than picture primes. However, it would not explain repetition suppression that was observed irrespective of whether words primed pictures or pictures primed words. Thus, our observation of bidirectional repetition suppression across stimulus type cannot be explained by (unknown) behavioral differences in word and picture priming.

Finally, future studies are required to identify the source or sources of the top-down influences on LvOT responses. This will require functional connectivity studies that distinguish forward and backward connections (e.g., dynamic causal modeling). However, these functional connectivity analyses are based on a priori region selection rather than a whole brain search and, at present, we do not know whether automatic repetition suppression across stimulus type in LvOT is driven by local backward connections (e.g., from anterior to posterior LvOT) or long-range backward connections from frontal, temporal, and parietal regions. Preliminary studies are therefore required to identify the candidate regions, for example, by using whole-brain analyses to explore how LvOT activation covaries positively and negatively with other brain regions.

We conclude that there is remarkable consistency in LvOT activation for written words and pictures of objects because, even if the neuronal populations for words and pictures are

distinct in the bottom-up processing stream, they are tightly interconnected by automatic backward connections and consequently they respond in common ways. Our results are consistent with many previous behavioral studies of priming and interference that have shown the very close link between word and object naming (Glaser 1992; Ferrand et al. 1994). The appreciation of the interactive and distributed nature of LvOT processing and the principals behind predictive coding, proposed here, allow us to integrate a wide range of previously disparate neuroimaging results. For example, greater left LvOT for low-frequency words and pseudowords relative to high-frequency words (Kuo et al. 2003; Mechelli et al. 2003; Kronbichler et al. 2004; Bruno et al. 2008) can be explained in terms of greater prediction error when the required response to a stimulus is less frequent or less familiar (i.e., frequency and familiarity influence the strength of the predictions from higher level regions). Our novel observations therefore have major implications for our understanding of reading, object recognition, and language. The tight interaction between these systems needs to be considered when studying each in isolation and in future studies that aim to decipher the neural code associated with visual word form processing in LvOT. In this context, our distinction between 5 models of response selectivity (Fig. 1) could provide a framework to guide future studies.

Funding

Wellcome Trust (WT082420) and the James S. MacDonnell Foundation (conducted as part of the Brain Network Recovery Group imitative).

Notes

Our thanks to Joseph Devlin, Karl Friston, Mohamed Seghier and Sue Ramsden for their helpful comments. *Conflict of Interest:* None declared.

References

- Alario F, Segui J, Ferrand L. 2000. Semantic and associative priming in picture naming. *Q J Exp Psychol A*. 53:741–764.
- Angelucci A, Levitt JB, Walton EJS, Hupe J, Bullier J, Lund JS. 2002. Circuits for local and global signal integration in primary visual cortex. *J Neurosci*. 22:8633–8646.
- 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.
- Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Schmidt AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, et al. 2006. Top-down facilitation of visual recognition. *Proc Natl Acad Sci U S A*. 103:449–454.
- Bodner GE, Masson ME. 2003. Beyond spreading activation: an influence of relatedness proportion on masked semantic priming. *Psychon Bull Rev*. 10(3):645–652.
- Bruno JL, Zumberge A, Manis FR, Lu Z, Goldman JG. 2008. Sensitivity to orthographic familiarity in the occipito-temporal region. *Neuroimage*. 39:1988–2001.
- 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(Pt 2):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(Pt 5):1054–1069.

- Dehaene S, Cohen L, Sigman M, Vinckier F. 2005. The neural code for written words: a proposal. *Trends Cogn Sci*. 9:335–341.
- 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(5):307–313.
- Dehaene S, Le Clec'h G, Poline JB, Le Bihan D, Cohen L. 2002. The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport*. 13:321.
- Devlin JT, Jamison HL, Gonnerman LM, Matthews PM. 2006. The role of the posterior fusiform gyrus in reading. *J Cogn Neurosci*. 18: 911–22.
- Duncan KJ, Pattamadilok C, Knierim I, Devlin JT. 2009. Consistency and variability in functional localisers. *Neuroimage*. 46(4):1018–1026.
- Durso FT, Johnson MK. 1979. Facilitation in naming and categorizing repeated pictures and words. *J Exp Psychol Hum Learn*. 5:449–459.
- Eddy M, Schmid A, Holcomb PJ. 2006. Masked repetition priming and event-related brain potentials: a new approach for tracking the time-course of object perception. *Psychophysiology*. 43:564–568.
- Eddy MD, Schnyer D, Schmid A, Holcomb PJ. 2007. Spatial dynamics of masked picture repetition effects. *Neuroimage*. 34:1723–1732.
- Eger E, Henson RN, Driver J, Dolan RJ. 2007. Mechanisms of top-down facilitation in perception of visual objects studied by fMRI. *Cereb Cortex*. 17:2123–2133.
- Ferrand L, Grainger J, Segui J. 1994. A study of masked form priming in picture and word naming. *Mem Cognit*. 22:431–441.
- Friston K, Kilner J, Harrison L. 2006. A free energy principle for the brain. *J Physiol (Paris)*. 100:70–87.
- Garrido MI, Kilner JM, Kiebel SJ, Friston KJ. 2007. Evoked brain responses are generated by feedback loops. *Proc Natl Acad Sci U S A*. 104:20961–20966.
- Gilbert CD, Sigman M. 2007. Brain states: top-down influences in sensory processing. *Neuron*. 54:677–696.
- Glaser WR. 1992. Picture naming. *Cognition*. 42:61–105.
- Gold BT, Balota DA, Jones SJ, Powell DK, Smith CD, Andersen AH. 2006. Dissociation of automatic and strategic lexical-semantics: functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *J Neurosci*. 26:6523–6532.
- Grill-Spector K, Malach R. 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol*. 107:293–321.
- Hanson SJ, Matsuka T, Haxby JV. 2004. Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a “face” area? *Neuroimage*. 23:156–166.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*. 293:2425–2430.
- Henson RNA, Rugg MD. 2003. Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia*. 41:263–270.
- Hon N, Thompson R, Sigala N, Duncan J. 2009. Evidence for long-range feedback in target detection: detection of semantic targets modulates activity in early visual areas. *Neuropsychologia*. 47:1721–1727.
- Kircher T, Sass K, Sachs O, Krach S. 2009. Priming words with pictures: neural correlates of semantic associations in a cross-modal priming task using fMRI. *Hum Brain Mapp*. 30(12):4116–4128.
- Korsnes MS, Wright AA, Gabrieli JDE. 2008. An fMRI analysis of object priming and workload in the precuneus complex. *Neuropsychologia*. 46:1454–1462.
- Kouider S, Dehaene S, Jobert A, Le Bihan D. 2007. Cerebral bases of subliminal and supraliminal priming during reading. *Cereb Cortex*. 17(9):2019–2029.
- Koutstaal W, Wagner AD, Rotte M, Maril A, Buckner RL, Schacter DL. 2001. Perceptual specificity in visual object priming: functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex. *Neuropsychologia*. 39:184–199.
- Kronbichler M, Hutzler F, Wimmer H, Mair A, Staffen W, Ladurner G. 2004. The visual word form area and the frequency with which words are encountered: evidence from a parametric fMRI study. *Neuroimage*. 21:946–953.
- Kuo W, Yeh T, Lee C, Wu YU, Chou C, Ho L, Hung DL, Tzeng OJL, Hsieh J. 2003. Frequency effects of Chinese character processing in the brain: an event-related fMRI study. *Neuroimage*. 18:720–730.
- La Heij W, Carmen Puerta-Melguizo M, Van Oostrum M, Starreveld P. 1999. Picture naming: identical priming and word frequency interact. *Acta Psychol*. 102:77–95.
- Lamme VA, Zipser K, Spekreijse HJ. 2002. Masking interrupts figure-ground signals in V1. *J Cogn Neurosci*. 14(7):1044–1053.
- Levy J, Pernet C, Treserras S, Boulanouar K, Berry I, Aubry F, Demonet J, Celsis P. 2008. Piecemeal recruitment of left-lateralized brain areas during reading: a spatio-functional account. *Neuroimage*. 43: 581–591.
- Mahon BZ, Milleville SC, Negri GAL, Rumiati RI, Caramazza A, Martin A. 2007. Action-related properties shape object representations in the ventral stream. *Neuron*. 55:507–520.
- McCauley C, Parmelee C, Sperber R, Carr T. 1980. Early extraction of meaning from pictures and its relation to conscious identification. *J Exp Psychol Hum Percept Perform*. 6:265–276.
- Mechelli A, Josephs O, Lambon Ralph MA, McClelland JL, Price CJ. 2007. Dissociating stimulus-driven semantic and phonological effect during reading and naming. *Hum Brain Mapp*. 28:205–217.
- Mechelli A, Gorno-Tempini ML, Price CJ. 2003. Neuroimaging studies of word and pseudoword reading: consistencies, inconsistencies, and limitations. *J Cogn Neurosci*. 15:260–271.
- Moore CJ, Price CJ. 1999. Three distinct ventral occipitotemporal regions for reading and object naming. *Neuroimage*. 10:181–192.
- Naccache L, Dehaene S. 2001. The priming method: imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cereb Cortex*. 11:966–974.
- Nakamura K, Dehaene S, Jobert A, Le Bihan D, Kouider S. 2007. Task-specific change of unconscious neural priming in the cerebral language network. *Proc Natl Acad Sci U S A*. 104(49):19643–19648.
- Norris D, Kinoshita S. 2008. Perception as evidence accumulation and Bayesian inference: insights from masked priming. *J Exp Psychol Gen*. 137(3):434–455.
- Price CJ, Devlin JT. 2004. The pro and cons of labelling a left occipitotemporal region: “the visual word form area”. *Neuroimage*. 22:477–479.
- Price CJ, Mechelli A. 2005. Reading and reading disturbance. *Curr Opin Neurobiol*. 15:231–238.
- Rao RPN, Ballard DH. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*. 2:79–87.
- Simon O, Kherif F, Flandin G, Poline JB, Rivière D, Mangin JF, Le Bihan D, Dehaene S. 2004. Automated clustering and functional geometry of human parietofrontal networks for language, space, and number. *Neuroimage*. 23:1192–1202.
- Simons JS, Koutstaal W, Prince S, Wagner AD, Schacter DL. 2003. Neural mechanisms of visual object priming: evidence for perceptual and semantic distinctions in fusiform cortex. *Neuroimage*. 19: 613–626.
- Sumner P, Brandwood T. 2008. Oscillations in motor priming: positive rebound follows the inhibitory phase in the masked prime paradigm. *J Mot Behav*. 40(6):484–489.
- Szwed M, Cohen L, Qiao E, Dehaene S. 2009. The role of invariant line junctions in object and visual word recognition. *Vision Res*. 49:718–725.
- Van den Bussche E, Van den Noortgate W, Reynvoet B. 2009. Mechanisms of masked priming: a meta-analysis. *Psychol Bull*. 135:452–477.
- van Gaal S, Ridderinkhof KR, van den Wildenberg W-P, Lamme VA. 2009. Dissociating consciousness from inhibitory control: evidence for unconsciously triggered response inhibition in the stop-signal task. *J Exp Psychol Hum Percept Perform*. 35(4):1129–1139.
- 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.
- Vuilleumier P, Henson RN, Driver J, Dolan RJ. 2002. Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat Neurosci*. 5:491–499.

- Weiskopf N, Hutton C, Josephs O, Deichmann R. Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: A whole-brain analysis at 3 T and 1.5 T. *Neuroimage*. 2006;33:493-504.
- Wheatley T, Weisberg J, Beauchamp MS, Martin A. 2005. Automatic priming of semantically related words reduces activity in the fusiform gyrus. *J Cogn Neurosci*. 17:1871-1885.
- Williams MA, Baker CI, Op de Beeck HP, Shim WM, Dang S, Triantafyllou C, Kanwisher N. 2008. Feedback of visual object information to foveal retinotopic cortex. *Nat Neurosci*. 11: 1439-1445.
- Wright ND, Mechelli A, Noppeney U, Veltman DJ, Rombouts SARB, Glensman J, Haynes J, Price CJ. 2008. Selective activation around the left occipito-temporal sulcus for words relative to pictures: individual variability or false positives? *Hum Brain Mapp*. 29:986-1000.
- Xue G, Chen C, Jin Z, Dong Q. 2006. Language experience shapes fusiform activation when processing a logographic artificial language: an fMRI training study. *Neuroimage*. 31:1315-1326.
- Xue G, Poldrack RA. 2007. The neural substrates of visual perceptual learning of words: implications for the visual word form area hypothesis. *J Cogn Neurosci*. 19:1643-1655.