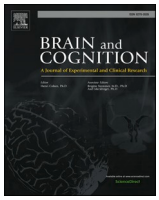
Brain and Cognition 152 (2021) 105757

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c



Neural correlates of the production effect: An fMRI study

Lyam M. Bailey [a], Glen E. Bodner [b], Heath E. Matheson [c], Brandie M. Stewart [a], Kyle Roddick [a], Kiera O'Neil [a], Maria Simmons [a], Angela M. Lambert [d], Olave E. Krigolson [e], Aaron J. Newman [a], Jonathan M. Fawcett [f,*]

[a] Dalhousie University, Department of Psychology and Neuroscience, Halifax, NS B3H 4R2, Canada
[b] Flinders University, College of Education, Psychology and Social Work, Adelaide, SA 5001, Australia
[c] University of Northern British Columbia, Psychology Department Prince George, BC V2N 4Z9, Canada
[d] University of Calgary, Department of Psychology, Calgary, AB T2N 1N4, Canada
[e] University of Victoria, School of Exercise Science, Victoria, BC V8W 2Y2, Canada
[f] Memorial University of Newfoundland, Department of Psychology, St. John's, NL A1B 3X9, Canada

ARTICLE INFO

Keywords:
Language
Speech
Memory
Brain
Human
Neuroimaging
Encoding
Recognition
Multivoxel pattern analysis
Representational similarity analysis

ABSTRACT

Recognition memory is improved for items produced at study (e.g., by reading them aloud) relative to a non-produced control condition (e.g., silent reading). This *production effect* is typically attributed to the extra elements in the production task (e.g., motor activation, auditory perception) enhancing item distinctiveness. To evaluate this claim, the present study examined the neural mechanisms underlying the production effect. Prior to a recognition memory test, different words within a study list were read either aloud, silently, or while saying "check" (as a sensorimotor control condition). Production improved recognition, and aloud words yielded higher rates of both recollection and familiarity judgments than either silent or control words. During encoding, fMRI revealed stronger activation in regions associated with motor, somatosensory, and auditory processing for aloud items than for either silent or control items. These activations were predictive of recollective success for aloud items at test. Together, our findings are compatible with a distinctiveness-based account of the production effect, while also pointing to the possible role of other processing differences during the aloud trials as compared to silent and control.

1. Introduction

A central issue in memory research is understanding how encoding strategies influence subsequent retention. An encoding strategy that has shown great promise for improving memory is the simple act of reading items aloud rather than silently. The memory advantage for reading aloud has recently been termed the "production effect" (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010; see also MacLeod & Bodner, 2017), and this advantage has been found for variants of production including mouthing, typing, writing, and spelling (e.g., Fawcett, Quinlan, & Taylor, 2012; Forrin, MacLeod, & Ozubko, 2012; MacLeod et al., 2010), singing (Quinlan & Taylor, 2013; Hassall, Quinlan, Turk, Taylor, & Krigolson, 2016), and even drawing (Wammes, Meade, & Fernandes, 2016). The production effect often scales up with the complexity of the productive act (e.g., Fawcett et al., 2012; Forrin et al., 2012; Quinlan & Taylor, 2013). Further, an influence of production is seen on various tests of explicit long-term memory including recognition and recall (e.g., Conway & Gathercole, 1987; Fawcett et al., 2012; Lin & MacLeod, 2012), but not on tests of implicit memory (MacLeod et al., 2010). This effect is typically larger when manipulated within-subjects as opposed to between-subjects (Bodner, Taikh, & Fawcett, 2014; Fawcett, 2013; Fawcett & Ozubko, 2016), suggesting that context modulates its influence (MacLeod et al., 2010; Ozubko, Gopie, & MacLeod, 2012).

The production effect has most often been attributed to production enhancing the distinctiveness of items in memory (e.g., Conway & Gathercole, 1987; Dodson & Schacter, 2001; MacLeod et al., 2010). Silent reading invokes both orthographic (visual) and semantic (meaning) processing but reading aloud necessitates additional productive elements — including engagement of the articulatory-motor system followed by auditory perception of the spoken item. The *production record* laid down at encoding thus includes these additional elements (Fawcett, 2013), which can later serve to help retrieve items from memory.

* Corresponding author.
E-mail address: jfawcett@mun.ca (J.M. Fawcett).

https://doi.org/10.1016/j.bandc.2021.105757
Received 11 May 2020; Received in revised form 16 May 2021; Accepted 18 May 2021
Available online 12 June 2021

By one version of a distinctiveness account, participants employ a distinctiveness heuristic at test (Dodson & Schacter, 2001; MacLeod et al., 2010) whereby access to the production record is used to discriminate studied from non-studied items ("If I can recollect saying it aloud at study, it was studied"). Although participants often report using this strategy (Fawcett & Ozubko, 2016), recent computational modeling of the production effect suggests that use of the production record need not be intentional/conscious (Jamieson, Mewhort, & Hockley, 2016). Instead, Jamieson et al. suggest that a distinctiveness account of the production effect may reflect intrinsic retrieval dynamics favouring recovery of items containing discriminative features.

Initial failures to observe a between-subjects production effect were taken as important evidence in favour of a distinctiveness account (e.g., MacLeod et al., 2010; Ozubko et al., 2012). To the extent that items are thought to be "distinctive" only in relation to other "non-distinctive" items from the same list (Hunt, 2006), a relative distinctiveness account predicts a production effect in within-subject designs but not in between-subject designs where there is no "backdrop" of non-produced items against which produced items stand out. However, meta-analyses and subsequent experiments have revealed a between-subject production effect in recognition memory (e.g., Bodner et al., 2014; Fawcett, 2013; Fawcett, Baldwin, Drakes, & Willoughby, submitted for publication; Forrin & MacLeod, 2016). Furthermore, Fawcett and Ozubko (2016) showed that the within-subject production effect reflects an increase in both familiarity (i.e., a sense of "knowing" that the item had been studied) and recollection (i.e., the ability to re-experience the episode in which the item had been studied) for produced items. In contrast, the between-subject production effect reflects only an increase in familiarity. These findings are difficult to explain with reference to distinctiveness alone and suggest that additional factors may contribute to the within-subject effect (Fawcett & Ozubko, 2016). For example, participants self-report paying more attention to the aloud items in post-experimental questionnaires (Fawcett & Ozubko, 2016) and are less likely to mind-wander when reading aloud than reading silently (Varao Sousa, Carriere, & Smilek, 2013).

1.1. The neural basis of speech production and human memory

The present study is the first to use functional magnetic resonance imaging (fMRI) to isolate the brain regions and processes contributing to the read-aloud version of the production effect. To inform this work, we first summarize research characterizing the neural networks involved in the processes invoked during a production task, including single-word speech production and memory retention.

Neuroimaging work in the area of speech production has implicated a left-lateralized network including the inferior frontal gyrus (IFG), middle frontal gyrus (MFG), superior frontal gyrus (SFG)—especially in the supplementary motor area (SMA), anterior cingulate cortex (ACC), insula, intra-parietal sulcus (IPS) and adjacent superior parietal lobule (SPL), angular gyrus (AG), and occipito-temporal cortex (including the fusiform, inferior occipital, middle occipital, and inferior temporal gyri; for meta-analyses, see Martin, Schurz, Kronbichler, & Richlan, 2015; Taylor, Rastle, & Davis, 2013; Vigneau et al., 2006; Vigneau et al., 2011; Wagner, Sebastian, Lieb, Tuscher, & Tadic, 2014). Subcortical regions have been similarly implicated, including the caudate nucleus, putamen, and thalamus. Right hemisphere brain regions are also consistently activated – albeit on a more restricted basis – including the IFG, pre-central gyrus (premotor cortex), middle temporal gyrus, and inferior parietal lobe.

Beyond production itself, participants also maintain – and further process – studied items in working memory (WM). Meta-analyses of studies involving such processes highlight a network of regions primarily in frontal and parietal cortices (Nee et al., 2012; Owen, McMillan, Laird, & Bullmore, 2005; Rottschy et al., 2012). Greatest convergence is observed in the superior frontal sulcus (separating the SFG and MFG) and superior parietal lobule, regions that relate most strongly to executive function. Additional regions implicated include the IFG and MFG (associated with maintenance of verbal information, and selection of information), ACC (e.g., task switching), and inferior parietal lobe (directing attention between items; Nee et al., 2012; Owen et al., 2005; Rottschy et al., 2012).

Work on longer-term memory representations has often focused on the medial temporal lobe (MTL), encompassing the perirhinal cortex, parahippocampal cortex, entorhinal cortex, and hippocampus. Activation levels in the hippocampus, left IFG/MFG, bilateral regions of premotor cortex, IPS/SPL, fusiform cortex, and hippocampus have been identified as predictive of later memory performance (Kim, 2011). Tasks involving item-specific encoding were associated with stronger activation of posterior IFG/MFG and premotor cortex, and the IPS/SPL. In contrast, associative encoding (i.e., memory for items in relation to one another) was more strongly related to activity in anterior IFG/frontal pole, insula, and hippocampal regions. Premotor and posterior parietal activation may relate to increased attention to individual items during such tasks (Kim, 2011), in line with the suggestion that parietal regions act as an "attentional circuit-breaker" that re-orients attention to relevant stimuli (Astafiev, Shulman, & Corbetta, 2006; Corbetta & Shulman, 2002; Vilberg & Rugg, 2008).

Recent work has highlighted the contrast between inferior and superior parietal regions in memory retrieval. For example, the Attention to Memory (AtoM) model (Ciaramelli, Grady, & Moscovitch, 2008) draws analogies between (1) bottom-up, alerting processes in attention and attentional capture by retrieved memory items, supported by the inferior parietal lobe (IPL); and (2) the SPL's involvement in top-down strategic orienting of attention and more effortful memory retrieval (e.g., of low-confidence items). A more recent model (Sestieri, Shulman, & Corbetta, 2017) takes a slightly different view, associating the SPL/IPS region with maintenance and attentional selection of task-relevant information retrieved from memory, and the IPL (primarily AG) region with recollecting specific details of an event or other retrieved information.

In addition to areas involved in encoding and retrieval generally, a popular view holds that successful retrieval should involve at least partial reinstatement (or recapitulation) of neural states that were present at encoding. Indeed there is evidence that activation during retrieval often reflects task-specific activation that was present during encoding. For example, a number of studies have shown that items studied alongside visual scenes, sounds, or scents during encoding differentially elicit greater levels of activation in visual association cortex, auditory association cortex, and olfactory cortices respectively during both encoding and test (e.g. Gottfried, Smith, Rugg, & Dolan, 2004; Vaidya, Zhao, Desmond, & Gabrieli, 2002; Wheeler, Petersen, & Buckner, 2000; see Danker & Anderson, 2010 for a review). More work advanced this area using multivoxel pattern analyses (MVPA), which permits examination of item-specific patterns of neural activation elicited by unique study items. In brief, a number of studies have shown that item-specific activity patterns elicited during encoding are often reinstated during successful recollection of those items, both with respect to task-relevant cortical activation (e.g. Ritchey, Wing, LaBar, & Cabeza, 2013; Wing, Ritchey & Cabeza, 2017) and activation in MTL structures (e.g. Danker, Tompary, & Davachi, 2017; Schultz et al., 2019; Staresina et al., 2012).

1.2. The present study

Participants studied a list of words, presented one at a time, in an event-related fMRI experiment. A cue indicated whether the word was to be read aloud, silently, or while saying aloud a control word ("check"). This control condition provided an estimate of baseline brain activation associated with articulating and hearing oneself produce a single word — but in a non-distinctive (i.e., non-item-specific) way. A similar manipulation was used by MacLeod et al. (2010), who showed that responding aloud with the same word ("yes" in that study) did not elicit

a production effect; hence, we did not expect our control condition to confer a memory benefit. After the study phase, participants completed a recognition test.

This design permitted a preliminary neuroimaging investigation into the mechanisms that underlie the production effect. Because distinctiveness-based accounts emphasize the additional sensorimotor processing of produced items (e.g., motor articulation), we expected stronger activation during encoding in both motor cortex (central sulcus/precentral gyrus) and auditory cortex (superior temporal gyrus) in the aloud condition than in the silent and control conditions. Presuming this information was used heuristically at test, activation in these regions was expected to correlate with the magnitude of the recollective component of the behavioural production effect. If participants differentially attended to the aloud items, stronger activation might also be expected for aloud items in areas associated with attention at encoding, such as the premotor and posterior parietal cortices (Kim, 2011). Insofar as manipulations of attention are most strongly associated with changes in recollection, activation in those regions should correlate with the magnitude of the production effect on recollection. Finally, participants might also demonstrate enhancement to other forms of encoding, such as semantic elaboration, resulting in greater activation of frontal or anterior superior temporal regions (e.g., Weber, Lau, Stillerman, & Kuperberg, 2016).

While our primary interest was in comparisons between conditions, we also examined item-specific patterns of neural activation with MVPA. One useful method of MVPA is representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008), which allows researchers to explore correspondence between item-specific neural activation and computational models that convey information about particular properties of experimental stimuli (for example, properties such as phonology, orthography, or semantic content). In the context of the current investigation, we reasoned that if produced items are indeed encoded more distinctively—owing to richer sensorimotor processing—then speaking words aloud ought to elicit more distinctive (i.e., dissimilar) activation patterns in sensorimotor regions, compared to reading them silently or saying "check" in response to every word. Moreover, activation patterns for produced items should better reflect the phonological properties of to-be-remembered words, given that phonology should approximately reflect sensory and motor information obtained through articulation. This claim would be evidenced by higher correspondence between neural data for produced items and a phonological model, compared to silent or control items.

Our primary interest was in brain activity during encoding, since the production effect is defined by how items are encoded. However, we also examined activity during the recognition test. To the extent that retrieval recapitulates encoding processes, similar regions may be recruited across study and test. In particular, if production results in relatively more elaborate memory representations, then we might see more activation in sensorimotor areas for aloud items at test. Models of the role of the parietal lobe in memory retrieval predict greater activation of the IPL for aloud items because they contain additional episodic details not present for items in the other two conditions.

2. Method

2.1. Subjects

Thirty-two healthy, English-speaking young adults (convenience sample; 20 females, 12 males; 20–32 years of age; $M = 24.1$ years) were recruited through on-campus advertising in exchange for $30 and an image of their brain. Previous work on the production effect has indicated that within-subjects designs typically elicit effect sizes of approximately Hedges $g = 0.6$ (Fawcett & Ozubko, 2016). A power analysis (implemented in *R* using the *pwr* package; Champely, 2020) indicated that a sample size of 26 (see below) had 80% power to detect a minimum effect size of 0.57. As such, our sample size was sufficient to detect the behavioural production effect. In the absence of prior published fMRI studies on this phenomenon, we assumed that the sample size appropriate for detecting the behavioural effect would be sufficient to identify an associated fMRI effect; as well, this number is consistent with typical sample sizes in fMRI studies published in recent years (Szucs & Ioannidis, 2020).

All participants were right-handed (Oldfield, 1971), with normal or corrected-to-normal vision, and reported no history of neurological conditions, attentional or language difficulties, current use of psychiatric medications, or contraindications to MRI scanning. The study was approved by the Dalhousie University Research Ethics Board. Participants provided informed consent according to the Declaration of Helsinki. Behavioural responses were not recorded for 4 subjects due to the response box malfunctioning; they were excluded from all behavioural and fMRI analyses. Two additional participants were excluded for appearing to confuse "know" and "no" responses at test (see Stimulus and Apparatus), or for not using each response category at least once. Thus, data from 26 participants were analyzed.

2.2. Stimuli and apparatus

Stimuli were presented using a custom script built in *PsychoPy2 1.84.2* (Peirce, 2009). The words were 120 nouns from MacDonald and MacLeod (1998), 5 letters to 10 letters in length, with frequencies greater than 30 per million (Thorndike & Lorge, 1944). For each subject, the script randomly assigned each word to one of four lists (30 words each), corresponding to the four experimental conditions (read aloud, read silently, sensorimotor control, and foils).

All words were presented in white, lowercase Courier size 20 font against a black background measuring 330 × 100 pixels superimposed in the centre of a complex visual scene that covered the remainder of the screen to reduce between-trial boredom. Study phase encoding instructions were provided using icons of a mouth (aloud condition), an eye (silent condition), or a check mark (sensorimotor control condition). These icons each measured 150 × 150 pixels, and were presented at the centre of the screen. All stimuli were presented on an LCD projector that was focused on a Mylar screen positioned in the bore behind the participants, viewed via an angled mirror. Throughout the test phase, participant responses were recorded by a fiber optic response pad (Current Designs Inc., Philadelphia, PA), using three buttons which were pressed by the index, middle, and ring fingers of one hand determined at random. Response hand was randomly determined for each subject to mitigate lateralized sensorimotor activation associated with operating the response box in the group-level contrasts. During the test phase, the mapping of the recollect, know, and no responses to these buttons was continuously presented as a reminder on-screen, centered and above the black background upon which stimulus words were presented.

2.3. Procedure

After providing informed consent and passing MRI safety screening, each participant was positioned in the MRI scanner. A brief scout scan determined head position, followed by four functional scans, and finally a structural scan. The study and test phase were both conducted in the MRI scanner. Each phase was further subdivided into two "runs" corresponding to separate fMRI scans, with a brief break in between. The first run of each phase was preceded by a practice version of the corresponding task, containing four replications of each condition. These practice phases used a unique set of words that did not appear in the main experimental fMRI runs, but were otherwise similar to their experimental counterparts.

2.3.1. Study phase

Participants were instructed to remember the studied items for an unspecified memory test. Each study phase trial began with a 250 ms fixation cross ("+") presented in the center of the screen to alert

participants to a new trial. The icon for that trial was then presented for 1000 ms, after which a word was presented for 2500 ms. During each study phase run, participants were presented with 15 trials of each condition (aloud, silent, sensorimotor control), presented in pseudo-random order. In addition to the 250 ms fixation period between trials, 30 additional "null" events lasting 2200 ms each were interspersed randomly to facilitate recovery of the event-related BOLD responses to each condition. These null events consisted of continuous display of the fixation cross. The placement of these null events, as well as the sequence of trials for each condition, was determined by the application *optseq2* (Dale, 1999; Dale, Greve, & Burock, 1999) to optimize recovery of the hemodynamic responses to individual stimuli (i.e., to improve estimation efficiency). The same trial order and timing were used for all participants, but the assignment of items to conditions and the order of specific items were randomized for each participant. Each study phase run lasted approximately 4 min.

2.3.2. Test phase

Recognition of the study items was tested using the remember-know recognition paradigm (Tulving, 1985), following a procedure detailed in Fawcett, Lawrence, and Taylor (2016). For each test item, the participant indicated whether they could *recollect* specific detail(s) about their studying of the item, *knew* they had studied the item but could not specifically recollect doing so, or *did not recognize* the item as one they had studied. Examples of each response type were provided. It was emphasized that recollect and know did not reflect differences in confidence, but rather reflected qualitative differences in how participants experienced their recognition of the items.

Each test phase trial began with a 1000 ms fixation cross ("+") presented in the center of the screen to alert participants of a new trial. A test item was then presented for 3000 ms, during which participants made their response using one of three buttons on the button box. The experiment continued after a period of 3000 ms, regardless of response.

During each test phase block, participants were presented with 15 items from each of the conditions used in the study phase (aloud, silent, sensorimotor control), as well as 15 "foil" items that were not presented during the study phase (i.e., the ratio of old items to new was 3:1). Each test phase run lasted approximately 5.5 min. As in the preceding phase, 40 null events (fixation lasting 2200 ms) were interspersed among the remaining trials according to an optimized sequence generated by *optseq2* (Dale, 1999). The same sequence of trial types and timing were used for all participants, but the order of the words was randomized for each participant.

2.4. MRI data acquisition

MRI scans were acquired on a 1.5 Tesla GE SIGNA LX MRI system (GE Medical Systems, Waukesha, WI) equipped with an 8 channel head coil. Each participant completed four functional scans followed by an anatomical scan. The fMRI scans used a gradient-echo, echo-planar pulse sequence with TR = 2 s, TE = 25 ms, flip angle = 90 deg, 64 × 64 matrix resulting in 3.75 × 3.75 mm in-plane voxel resolution with 34, 3.7 mm thick axial slices (no gap, interleaved slice acquisition). For each run, we obtained either 113 or 115 functional volumes (originally we specified 113 volumes, however after scanning 25 participants we realized that the response to the last stimulus item might be truncated, so added an additional two time points) during the study phase; 165 volumes during the test phase. Three additional volumes were acquired but automatically discarded from the start of every run immediately following acquisition. The T1-weighted anatomical image was obtained using a 3D fast spoiled gradient echo sequence (FSPGR BRAVO) with TR = 11.8 ms, TE = 4.69 ms, TI = 450 ms, flip angle = 12 deg, FOV = 202 mm, matrix = 224 × 224, 202 axial slices.

2.5. Data preprocessing and analysis

2.5.1. Behavioral data

Trials that did not contain a response (0.61%) were labelled as "missed" trials and were removed from analysis. The remaining behavioural data were analyzed as a function of condition (aloud, silent, control, foil) using multilevel logistic regression models (Baayen, Davidson, & Bates, 2008) implemented with the *brms* package (Bürkner, 2017a, 2017b) in R 3.5.1 (R Core Team, 2016).[1] These models were fit using a fully Bayesian approach with weakly informative priors[2]. Results are summarized on the back-transformed response (i.e., percentage) scale rather than the logit scale. Models included random intercepts and slopes for both subject and item, representing the "maximal" random structure corresponding to our design (Barr, Levy, Scheepers, & Tily, 2013).

2.5.2. fMRI data

The fMRI data were processed using FEAT (fMRI Expert Analysis Tool) Version 5.98, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). To mitigate potential task-related motion artefact (particularly during the aloud and sensorimotor control conditions, in which speaking aloud might cause head motion), motion correction using MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002) was applied; visual inspection of the results was used to exclude any data where head motion across time points exceeded 2 mm. This resulted in the removal of 5 runs from further analysis; no more than one run per participant was removed. Additionally, 4 runs had excessive head movement only later in the runs, and so were trimmed by removing time points from the onset of excessive head motion to the end of the run (anywhere from 50 to 100 time points). Additional preprocessing steps included: non-brain removal using BET (Smith, 2002); spatial smoothing using a Gaussian kernel of FWHM 6 mm; grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor; and high pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 50.0 s). Prior to statistical analyses we excluded all trials that corresponded to an incorrect response at test from both study- and test-phase fMRI analyses.

Spatial registration and normalization was carried out using FLIRT (Jenkinson & Smith, 2001; Jenkinson et al., 2002), with each individual's EPI volumes registered to their respective high-resolution structural image (using rigid body transformation), and the high-resolution structural in turn registered to the MNI152 template using first linear affine, and then nonlinear methods (the latter implemented in FNIRT; Andersson et al., 2007a, 2007a). The outputs of first-level statistical analysis (see next paragraph) were transformed to standard space using the combined EPI-to-structural and structural-to-MNI152 transforms, and resliced to 2 mm isotropic resolution.

Statistical analyses of the fMRI data proceeded over three levels. The

[1] Each model was fit using 10,000 iterations with 5000 warm-up samples; convergence was verified through visual inspection and using standard convergence metrics such as R-hat ≈ 1 (Gelman & Hill, 2006). There were no divergent transitions. For more detail see Fawcett and Ozubko (2016) and Fawcett et al. (2016).

[2] For the "old" and independence know data, priors for the intercept (false alarm rate) and slopes were represented by *Normal*(-1, 2) and *Normal*(0, 4), respectively. Priors for the intercept were broad but acknowledged that false alarms were likely to be rare (i.e., < 50%); priors for the slopes were effectively uniform. For the recollect data, priors for the intercept (false alarm rate) and slopes were represented by *Normal*(-4, 2) and *Normal*(0, 4), respectively. These changes reflected our knowledge that false alarms would be less common for recollect responses. Priors for the SD of each random effect were represented by *Normal*(0, 2) with a regularizing prior on the correlation matrix equivalent to *LKJ*(4). All priors are reported on the logit scale. Models fit instead using (less principled) default priors provided by the *brms* package produced similar results.

first level was performed on each run individually, and involved multiple linear regression using FSL's FILM with local autocorrelation correction (Woolrich, Ripley, Brady, & Smith, 2001). Regressors included time series for each stimulus type (aloud, silent, sensorimotor control; foils for test phase runs) convolved with a model of the hemodynamic response (a gamma function), as well as the six parameters derived from the motion correction step as covariates of no interest. The regressors of interest were orthogonalized with respect to the motion parameters to eliminate issues of collinearity. Contrasts of interest included each condition relative to baseline, as well as the pairwise contrasts aloud-silent, aloud-control, and control-silent. For the test phase, we also contrasted each "old" condition (aloud, control, silent) against the foil items.

The parameter estimates and associated variances from first-level analyses were combined in the second-level analysis, separately for each participant and phase (study/test) using fixed-effects linear regression to estimate the mean effect across runs for each contrast for each participant.

Finally, the third-level analysis was performed using nonparametric permutation inference with FSL's *randomise* (Winkler, Ridgway, Webster, Smith, & Nichols, 2014). Correction for multiple comparisons of the resulting statistical maps was applied using threshold-free cluster enhancement (Smith & Nichols, 2009) with family-wise error correction set at $p < .05$. For between-condition contrasts, the data were masked during the nonparametric inference procedure to restrict the analyses to voxels that were significantly ($p \leq 0.05$) activated in the nonparametric analysis of the minuend for that contrast relative to baseline. For example, the aloud-control contrast was restricted to voxels that were significant in the aloud-baseline contrast. Tables of the resulting activations were generated using FSL's *cluster* routine to identify clusters of contiguous voxels (with a minimum spatial extent of 25 adjacent voxels, to exclude small clusters that were likely spillover from another ROI) and the location of the peak *z* score within each cluster, and *atlasquery* to identify the anatomical label of the voxel having the peak *z* value within each cluster.

We conducted additional analyses in which we correlated fMRI data from the study phase either with test-phase performance in each condition corrected for false alarms[3], or with the behavioural production effect[4]. Both behavioural measures were calculated three ways: (1) overall 'old' performance (percentage of items correctly identified as old at test), (2) 'independent know' performance (percentage of old items to which participants made a 'know' response after excluding trials with a 'recollect' response), and (3) recollection performance (percentage of old items to which participants made a 'recollect' response). This resulted in six behavioral scores for each participant: three accuracy scores and three production effect scores, computed separately for old, know, and recollect performance.

We correlated the BOLD response for each condition during the study phase relative to baseline with the six behavioural scores. Here, the first- and second-level analyses included *all* trials from the study phase (i.e., we did not exclude items that were incorrectly identified as 'new' at test). We then conducted three sets of third-level analyses in which the BOLD response for each condition (aloud, silent, control) was correlated with behavioural scores: either overall performance in the corresponding condition (e.g., BOLD responses to aloud trials were correlated with accuracy for aloud items), or the behavioural production effect. This was achieved by conducting third-level analyses as described earlier, but in this case the behavioural score of interest was included as a covariate. Finally, the results from these correlation analyses were masked with activation maps from our main analyses (described above) to ensure that correlations with behaviour were restricted to areas showing task-related activation. (see Table 1)

This procedure was repeated for each condition, resulting in 18 separate correlation analyses, each including one behavioural score as a single covariate. Finally, we replicated this procedure to correlate the BOLD response derived from the aloud-silent and aloud-control contrasts during the study phase with the behavioural production effect (aloud-silent and aloud-control respectively) separately for old, know, and recollect judgments. Having conducted these 24 correlation analyses, we corrected for multiple comparisons by applying a Bonferroni adjustment (i.e., alpha level / 24) to the results of any correlation that yielded significant results.

Most of the fMRI analyses resulted in very large clusters spanning multiple brain regions; in many cases, all activations were subsumed in a single contiguous cluster. To produce tables that accurately represented the brain regions included in these large activation clusters, we performed clustering for each activation map of interest, within each brain region defined in the Oxford-Harvard cortical and subcortical atlases (e.g., Newman et al., 2010a, 2010b, 2015). In all cases, the statistical analyses for each contrast were performed on the entire brain and corrected using threshold-free cluster enhancement; this segmentation into regions of interest (ROIs) was performed only for the purpose of generating Tables 2–6.

With respect to results from the test phase, because we were primarily interested in activation that reflected encoding processes (i.e., reinstatement), activation derived from each contrast at test was spatially constrained to areas that were activated for the same contrast at study (e.g., results from the aloud-baseline contrast at test were masked with the aloud-baseline contrast at study, and so on). Contrasts involving foils were constrained to the contrast of the minuend condition relative to baseline at study (e.g., the aloud–foil contrast was masked with aloud–baseline from the study phase, and so on). The only test phase contrast that was not spatially constrained in this manner was foil–baseline. Results for non-masked test phase contrasts are reported in supplementary material.

2.5.3. Representational similarity analysis (RSA)

Procedures for our multivariate analyses are detailed in supplementary materials, so we will describe them only briefly here. We first obtained single-trial estimates of activation for every item presented during the study phase using an iterative modelling procedure proposed by Mumford, Turner, Ashby, and Poldrack (2012). Mathematically, single-trial estimates provide a pattern vector for every trial, whereby each value in the vector indicates the level of activation in a particular voxel. We performed RSA to assess correspondence between these neural pattern vectors and a formal phonological model, using a whole brain searchlight analysis. More specifically, at the center of every searchlight sphere (3 mm radius) we constructed a neural dissimilarity matrix (DSM) comprising pairwise correlation distances between the activation patterns to each item, for each condition and subject separately. These neural DSMs were correlated with a phonological model: a DSM comprising pairwise phonological edit distances reflecting phonological dissimilarity between study items. Our searchlight analysis used functions from the CoSMoMVPA toolbox (Oosterhof, Connolly, & Haxby, 2016) implemented in MATLAB (The MathWorks, Inc., Natick) and additional custom code.

To avoid biasing the MVPA results with the results from our univariate analysis, the results from our searchlight analysis were constrained to a set of independent, a priori ROIs relevant to word reading identified by Murphy, Jogia, and Talcott (2019; see Table S1 for the list of ROIs and their MNI coordinates). Within these ROIs, we used random-

[3] This correction involved subtracting the percentage of false alarms (i.e. "remember" or "know" responses to foil items) from the percentage of hits to items that were present during the study phase. This correction was intended to control for participants guessing at test.

[4] For correlations involving neural data from the aloud or silent conditions, the production effect was defined as performance on aloud trials minus performance on silent trials. For correlations involving neural data from the control condition, the production effect was defined as performance on aloud trials minus performance on control trials.

Table 1
Activation coordinates for each condition relative to baseline during the study phase.

Study Phase			Aloud-Baseline					Silent-Baseline					Control-Baseline				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
Frontal	Cingulate Gyrus anterior	LH	943	5.43	0	18	36						357	4.25	−2	6	44
		RH	518	5.23	2	18	34						39	3.74	2	4	46
	Frontal Operculum Cortex	LH	426	5.42	−50	10	−2						388	5.04	−50	10	−2
		RH	264	4.31	38	12	8						278	4.28	50	10	−2
	Frontal Orbital Cortex	LH	507	5.07	−38	28	−6						702	4.84	−38	32	−6
			103	4.32	−22	4	−14										
		RH	122	4.18	24	6	−10						35	2.98	34	26	0
			90	3.44	38	28	−2										
	Frontal Pole	LH	96	3.9	−48	38	6	161	2.96	−40	38	10	2540	4.82	−32	48	28
		RH	152	3.18	50	38	10										
	Inferior Frontal Gyrus pars opercularis	LH	551	5.31	−54	10	−2	387	3.73	−44	14	18	1189	5.76	−60	14	−4
		RH	254	3.77	48	8	14						98	3.66	40	20	12
			46	3.59	60	12	0						93	3.79	52	14	−2
	Inferior Frontal Gyrus pars triangularis	LH	327	4.09	−44	30	−2	390	3.89	−46	28	22	991	4.58	−44	24	20
		RH	74	2.92	40	32	4						77	3.6	56	20	−8
													30	3.4	42	20	14
	Middle Frontal Gyrus	LH	237	5.67	−40	−2	58	611	4.21	−46	28	24	2585	5.5	−32	−2	58
								535	4.14	−38	−2	60					
		RH	139	4.8	56	6	48						356	4.57	40	0	62
	Paracingulate Gyrus	LH	359	5.01	−2	18	38	435	3.83	−6	10	52	646	5.61	0	10	52
		RH	125	4.52	2	18	38	29	2.88	2	8	50	126	5.3	2	10	54
	Precentral Gyrus	LH	4447	7.03	−44	−14	36	1163	4.44	−48	2	50	4308	6.11	−50	−6	48
		RH	3536	7.08	46	−10	34						3140	6.45	54	−2	44
	Subcallosal Cortex	LH	30	3.88	0	6	2										
	Superior Frontal Gyrus	LH	948	4.89	0	10	60	395	3.7	−4	10	56	1798	5.63	−4	10	60
		RH	645	5.62	6	10	62						626	5.28	2	10	56
	Supplementary Motor Area	LH	927	5.9	0	4	62	338	4.72	−6	6	54	889	6.44	−2	2	62
		RH	735	6.51	2	4	62	66	3.42	2	6	54	747	6.21	2	4	60
Parietal	Angular Gyrus	LH						187	3.52	−44	−52	42	409	5.09	−42	−58	54
													158	4.56	−62	−50	12
	Central Opercular Cortex	LH	1084	7.08	−58	−10	6						911	5.97	−62	−10	6
		RH	840	6.84	60	−4	10						762	6.04	60	−8	12
	Cingulate Gyrus posterior	LH	102	4.33	−2	−44	0						155	3.76	−4	−48	−4
		RH	80	4.08	4	−42	0						58	3.37	2	−44	2
	Parietal Operculum Cortex	LH	360	4.6	−62	−26	16						153	4.24	−64	−38	20
		RH	145	4.47	64	−28	20						44	3.05	56	−30	18
	Postcentral Gyrus	LH	2496	7.1	−46	−14	30						2330	6.15	−50	−12	24
			198	3.48	−2	−50	74										
		RH	1776	7.43	46	−12	34						1506	6.13	48	−12	30
	Precuneous Cortex	LH	216	3.92	−2	−80	52						689	5.03	−2	−82	48
		RH	74	3.56	2	−68	64						474	4.75	2	−58	72
			49	3.38	2	−82	50										
	Superior Parietal Lobule	LH	32	3.06	−22	−44	48						489	4.61	−38	−56	52
	Supramarginal Gyrus anterior	LH	116	4.33	−64	−24	16						95	3.48	−52	−40	46
													41	3.82	−64	−24	16
		RH	116	4.24	72	−16	12						78	4.36	72	−16	12
	Supramarginal Gyrus posterior	LH	391	4.73	−52	−44	10	42	3.35	−42	−52	42	1193	5.59	−64	−44	10
		RH	264	4.91	52	−38	6						304	5.58	52	−38	6
Temporal	Heschls Gyrus	LH	343	6.77	−56	−10	4						359	5.2	−56	−18	8
		RH	344	6.47	54	−14	2						295	5.17	56	−12	6
	Inferior Temporal Gyrus posterior	LH	111	4.44	−46	−44	−16						299	4.27	−48	−46	−8

(*continued on next page*)

Table 1 (*continued*)

Study Phase			Aloud-Baseline					Silent-Baseline					Control-Baseline				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
	Inferior Temporal Gyrus temporooccipital	LH	764	5.55	−48	−54	−18	394	4.41	−46	−64	−16	1053	6.35	−48	−54	−24
		RH	616	6.04	56	−62	−20	199	4.32	46	−60	−16	789	5.38	48	−60	−16
	Lingual Gyrus	LH	1744	6.7	−14	−64	−18	215	5.95	−16	−88	−6	1801	6.08	−16	−88	−6
		RH	1519	6.95	20	−64	−20	159	6.14	16	−86	−8	1488	6.85	14	−88	−8
	Middle Temporal Gyrus anterior	LH	240	4.8	−62	2	−12						125	4.25	−62	2	−12
		RH	128	5.71	62	−6	−10						31	3.67	64	−4	−10
	Middle Temporal Gyrus posterior	LH	416	4.77	−68	−20	−6						964	5.09	−66	−42	4
		RH	680	5.21	72	−28	0						764	5.48	52	−34	−2
	Middle Temporal Gyrus temporooccipital	LH	217	4.58	−52	−44	6						957	5.55	−58	−58	8
		RH	160	5.22	52	−38	4						635	5.63	46	−38	4
	Planum Polare	LH	494	7.03	−56	0	−4						244	5.31	−52	4	−4
		RH	349	6.53	58	−2	−2						163	4.99	58	−6	4
	Planum Temporale	LH	765	7.2	−58	−10	4						729	5.83	−62	−10	4
		RH	552	7.07	62	−14	2						543	5.48	64	−8	6
	Superior Temporal Gyrus anterior	LH	404	7.05	−56	−2	−4						375	5.69	−64	−10	4
		RH	388	6.81	60	−6	−4						309	4.84	60	2	−2
	Superior Temporal Gyrus posterior	LH	1348	7.07	−62	−20	2						1331	5.84	−70	−10	4
		RH	1418	7.24	62	−14	0						1371	5.6	66	−18	0
	Temporal Fusiform Cortex anterior	LH	236	3.87	−32	−2	−34										
		RH	114	3.21	24	−2	−44										
	Temporal Fusiform Cortex posterior	LH	575	4.94	−34	−52	−26	73	3.55	−38	−42	−24	596	6.41	−40	−50	−28
		RH	93	3.6	36	−42	−30						103	4.56	36	−42	−30
			29	2.83	38	−8	−24						31	3.23	38	−16	−18
	Temporal Occipital Fusiform Cortex	LH	719	5.84	−18	−64	−18	488	4.94	−38	−62	−16	729	6.56	−44	−52	−28
		RH	735	6.58	36	−62	−20	547	5.52	42	−58	−18	768	5.74	42	−60	−26
	Temporal Pole	LH	1147	6.78	−56	4	−6						508	6	−52	8	−6
			77	3.49	−34	2	−36										
		RH	1040	6.18	54	8	−12						433	5.16	54	10	−6
			245	4.23	26	0	−18										
Occipital	Cuneal Cortex	LH	302	4.53	−6	−86	40						333	4.86	0	−88	44
		RH	384	4.66	2	−86	36						386	5.26	4	−86	34
	Intracalcarine Cortex	LH	672	5.86	−4	−72	8	81	4.46	−8	−90	−2	712	4.92	−18	−86	10
		RH	913	5.1	2	−74	10	30	3.57	12	−88	0	740	5	12	−88	0
	Lateral Occipital Cortex inferior	LH	2345	6.36	−40	−84	−10	1700	5.95	−36	−88	−4	2258	6.05	−24	−90	4
			35	2.96	−52	−62	8						60	4.4	−60	−62	6
		RH	2628	6.65	44	−78	−16	1970	6.92	30	−86	−6	2277	6.14	44	−78	−16
	Lateral Occipital Cortex superior	LH	454	5.22	−30	−84	8	802	3.78	−32	−64	40	2664	5.29	−40	−58	52
								195	4.92	−20	−90	8					
		RH	512	4.87	40	−88	8	796	4.96	26	−88	6	562	5.01	34	−86	16
													266	3.84	6	−74	60
	Occipital Fusiform Gyrus	LH	1515	6.73	−24	−82	−12	1371	6.61	−22	−82	−12	1652	6.63	−34	−74	−20
		RH	1481	6.89	16	−68	−20	1159	6.49	30	−84	−8	1495	6.61	26	−86	−12
	Occipital Pole	LH	2999	6.63	−14	−94	−10	2330	6.3	−16	−90	−8	2783	6.47	−20	−94	2
		RH	2678	6.28	18	−94	−14	1644	5.99	24	−90	−6	2360	6.45	14	−90	−8
	Supracalcarine Cortex	LH	80	5.07	0	−74	12						78	3.87	0	−74	12
		RH	91	4.94	2	−74	12						88	3.97	2	−74	12
Medial	Amygdala	LH	327	5.04	−20	−2	−18						128	3.26	−16	−4	−20
		RH	353	4.45	26	−2	−16						26	3.21	28	−2	−14
	Hippocampus	LH	474	5.26	−20	−26	−8						361	3.73	−22	−26	−10
		RH	297	3.38	22	−12	−22						140	3.51	34	−14	−18
	Insular Cortex	LH	1201	4.94	−38	2	2						930	4.61	−34	20	−2

(*continued on next page*)

Table 1 (*continued*)

Study Phase			Aloud-Baseline					Silent-Baseline					Control-Baseline				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
		RH	1220	5.4	36	8	2						835	4.35	32	6	2
	Parahippocampal Gyrus anterior	LH	610	5.04	−20	−2	−18						106	3.26	−16	−4	−20
													52	3.73	−36	−20	−18
		RH	586	4.22	28	0	−18						71	3.51	34	−14	−18
	Parahippocampal Gyrus posterior	LH	469	5.26	−20	−26	−8						252	4.35	−2	−44	−4
			27	3.22	−18	−36	−24										
		RH	143	3.67	10	−30	−8						230	3.85	−10	−2	12
Basal Ganglia	Caudate	LH	389	5.5	−10	0	10										
		RH	231	4.5	8	2	8						37	3.04	18	4	18
	Pallidum	LH	243	4.62	−18	4	2						139	4.21	−20	2	0
		RH	234	4.78	20	4	2						168	3.88	26	−10	−2
	Putamen	LH	740	5.08	−26	6	4						664	4.57	−22	4	0
		RH	696	5.05	22	6	6						624	4.41	26	6	2
Midbrain	Thalamus	LH	1065	5.75	−10	−2	8						699	4.21	−10	−4	10
		RH	941	5.09	6	−2	8						549	4.28	6	−2	4

effect cluster statistics to identify any clusters which significantly differed between conditions.

3. Results

3.1. Behavioural

Recollect and know responses during the recognition test were initially aggregated into "old" responses and analyzed as a function of item type (aloud, silent, control, foil). Mean proportions corresponding to each condition are plotted in Fig. 1. Contrasts for differences between the individual conditions were calculated using the *emmeans* package (Lenth, 2018). Participants were more likely to respond "old" to silent items than to foil items, *difference* = 43.4%, *95% HDI* [37.0%, 50.0%], but there was little difference in the recognition of silent and control items, *difference* = 1.9%, *95% HDI* [−5.7%, 9.8%]. The latter finding supports previous work showing that producing a non-unique response to items does not reliably improve memory (MacLeod et al., 2010). Importantly, we also replicated the standard production effect: Recognition was greater for aloud items compared to either silent items, *difference* = 21.3%, *95% HDI* [13.9%, 28.5%], or control items, *difference* = 19.4%, *95% HDI* [10.6%, 27.9%].

We next evaluated the effect of production on recollection and familiarity. For familiarity, we first excluded all trials for which a recollect response had been made, to correct for nonindependence of remember/know judgments that can underestimate familiarity (e.g., Jacoby, Yonelinas, & Jennings, 1997; Mangels, Picton, & Craik, 2001; Ochsner, 2000; Ozubko et al., 2012; Yonelinas, 2002; Yonelinas & Jacoby, 1995). In the context of a logistic regression model, fitting the model for "know" responses after excluding "recollect" responses produces an estimate comparable to standard independent remember/know calculations (for statistical proof, see Fawcett et al., 2016; Fawcett & Ozubko, 2016).

Complementary models were fit to the proportion of recollect and independent know responses made for each item type. Separate models were fit to each judgment because recollect and know judgments were mutually exclusive and hence dependent. Mean proportions are also reported in Fig. 1. For the recollection model, participants were more likely to recollect silent items than foil items, *difference* = 11.6%, *95% HDI* [6.8%, 16.9%], whereas recollection of the silent items again failed to differ from the recollection of the control items, *difference* = 2.3%, *95% HDI* [−3.0%, 7.6%]. Recollection was greater for aloud items compared to either silent items, *difference* = 14.1%, *95% HDI* [7.7%, 21.3%], or control items, *difference* = 11.9%, *95% HDI* [5.0%, 19.2%].

For the familiarity model, silent items were more familiar than foil items, *difference* = 35.2%, *95% HDI* [27.8%, 42.7%], but not control items, *difference* = 0.0%, *95% HDI* [9.4%, 9.2%]. Familiarity was again greater for aloud items compared to either silent items, *difference* = 20.9%, *95% HDI* [12.3%, 29.7%], or control items, *difference* = 20.9%, *95% HDI* [10.2%, 31.4%]. These findings replicate the production effects on recollection and familiarity in past behavioural studies (Fawcett & Ozubko, 2016; Ozubko et al., 2012).

We calculated the Cohen's *d* effect size of the behavioural production effect, defined as the standardised difference between produced and silent items, for each measure separately (using the R package *rstatix*; Kassambara, 2021). This yielded effect sizes of $d = 1.36$, 1.00, and 1.29 for the combined "old", recollection, and familiarity judgments respectively. Finally, we assessed the split-half reliability for the observed production effect using a permutation-based approach (implemented in *R* using the *splithalf* package; Parsons, 2020) with 5000 random splits. In brief, this entailed iteratively splitting the data from each participant into two halves (without replacement) such that, for each permutation, the production effect was calculated twice. Production effect scores were

Table 2
Activation coordinates for contrasts between conditions during the study phase.

Study Phase			Aloud-Silent					Aloud-Control					Control-Silent				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
Frontal	Cingulate Gyrus anterior	LH	787	4.64	−2	16	32						243	4.04	−8	2	42
		RH	210	4.2	2	16	32										
			65	3.85	2	−8	46										
	Frontal Operculum Cortex	LH											155	3.16	−48	16	−6
		RH	52	3.49	52	10	0										
	Frontal Orbital Cortex	LH	47	3.5	−20	4	−16						170	3.89	−54	20	−8
		RH	77	3.36	22	4	−16										
	Inferior Frontal Gyrus pars opercularis	LH	75	4.39	−60	12	−4						298	4.78	−60	12	−4
		RH	45	3.35	58	10	14										
	Inferior Frontal Gyrus pars triangularis	LH											54	3.63	−52	20	−8
	Middle Frontal Gyrus	LH											383	3.94	−38	0	56
	Paracingulate Gyrus	LH											42	3.52	0	10	54
	Precentral Gyrus	LH	2520	6.77	−48	−10	30						1640	6.22	−52	−6	24
													312	4.7	0	−14	68
		RH	2350	6.9	50	−8	26	65	2.9	60	2	18	1825	5.97	48	−8	30
			121	4.59	2	−14	68						206	4.35	2	−14	68
	Superior Frontal Gyrus	LH	356	4.12	−16	−2	66						399	4.31	−8	−6	72
													80	3.31	−24	4	46
		RH	278	4.18	6	10	64						191	3.96	8	6	68
	Supplementary Motor Area	LH	561	4.74	0	−10	66						646	5.31	−2	−8	68
		RH	537	4.76	2	2	62						466	5.17	2	2	64
Parietal	Angular Gyrus	LH											114	4.09	−56	−60	14
													44	3.77	−40	−54	50
	Central Opercular Cortex	LH	921	6.66	−58	−10	6	124	3.5	−60	−10	6	733	6.7	−62	−10	6
		RH	670	6.48	62	−4	12	62	3.18	62	−6	6	684	6.07	58	−4	12
	Cingulate Gyrus posterior	LH	82	4.4	−2	−44	0						119	3.75	−4	−48	−4
		RH	65	4.59	4	−38	0						38	3.34	2	−44	2
	Parietal Operculum Cortex	LH	302	5.41	−62	−26	16						142	4.75	−62	−24	16
		RH	111	4.11	62	−26	18						37	3.44	64	−24	18
	Postcentral Gyrus	LH	2082	6.78	−44	−12	28	57	3.36	−44	−14	30	1334	6.72	−52	−10	24
													243	3.34	−4	−44	70
													27	2.61	−44	−22	62
		RH	1151	7.05	52	−10	30	75	3.1	66	−10	12	927	6.2	50	−10	30
			204	4.19	26	−28	70						105	3.26	4	−40	72
													87	3.34	24	−30	60
	Precuneous Cortex	LH											550	3.98	0	−78	36
		RH											233	3.67	2	−78	38
	Superior Parietal Lobule	LH											150	3.99	−38	−54	52
	Supramarginal Gyrus anterior	LH	116	5.18	−64	−24	16						41	5.21	−64	−24	16
		RH	93	3.72	64	−20	22						75	4.02	66	−22	18
	Supramarginal Gyrus posterior	LH	207	4.22	−46	−42	12						481	4.6	−62	−42	14
		RH	177	4.35	60	−36	12						207	5.2	48	−38	8
Temporal	Heschls Gyrus	LH	280	6.55	−56	−10	4	52	4.25	−56	−12	2	317	5.06	−56	−10	4
		RH	296	6.2	50	−18	4	80	4.24	58	−6	0	272	5.9	54	−14	8
	Inferior Temporal Gyrus posterior	LH											48	3.98	−48	−42	−8
	Inferior Temporal Gyrus temporooccipital	LH	125	3.1	−46	−46	−14						231	3.3	−52	−62	−14
													37	3.57	−46	−46	−8
		RH	91	3.06	54	−52	−22						159	3.32	64	−50	−20
	Lingual Gyrus	LH	1110	5.61	−16	−64	−18						1185	5.25	−16	−64	−18
		RH	1043	6.3	16	−64	−18						1025	5.77	16	−64	−18

(*continued on next page*)

Table 2 (*continued*)

Study Phase			Aloud-Silent					Aloud-Control					Control-Silent				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
	Middle Temporal Gyrus anterior	LH	223	5.17	−64	−8	−8						87	4.21	−68	−10	−6
		RH	128	4.63	60	4	−16	69	3.3	66	−2	−10	31	3.66	62	2	−14
	Middle Temporal Gyrus posterior	LH	204	4.61	−64	−12	−8						617	4.53	−66	−26	−4
		RH	635	5.83	54	−20	−8						673	6.19	56	−20	−6
	Middle Temporal Gyrus temporooccipital	LH	29	3.11	−48	−44	6						510	4.46	−56	−58	8
													31	2.85	−60	−60	−8
		RH	97	3.98	50	−38	4						437	4.99	46	−40	6
	Planum Polare	LH	460	6.66	−60	−8	4	53	4.05	−56	0	−4	230	5.56	−60	−8	4
		RH	315	6.27	58	−2	−2	77	4.3	62	−4	2	162	4.88	66	−4	6
	Planum Temporale	LH	752	7.17	−58	−10	4	159	5.23	−62	−14	2	705	6.82	−64	−10	6
		RH	471	6.21	62	−10	0	156	4.68	62	−8	0	484	5.95	56	−14	6
	Superior Temporal Gyrus anterior	LH	401	6.93	−60	−10	2	205	5.14	−60	−12	0	374	6.64	−64	−10	4
		RH	387	6.14	58	−2	−4	299	4.56	62	−6	0	309	4.84	62	−6	0
	Superior Temporal Gyrus posterior	LH	1254	6.93	−70	−10	4	323	5.46	−64	−14	0	1292	6.77	−66	−10	4
		RH	1347	6.54	58	−28	4	375	4.52	64	−8	0	1321	6.17	56	−18	−6
	Temporal Fusiform Cortex anterior	LH	82	3.28	−40	−6	−28										
	Temporal Fusiform Cortex posterior	LH	161	3.7	−30	−34	−32						118	3.87	−46	−44	−8
	Temporal Occipital Fusiform Cortex	LH	211	5.42	−22	−64	−20						99	5.13	−18	−64	−18
		RH	120	4.84	20	−60	−18						64	4.98	22	−62	−18
													38	3.29	48	−52	−30
	Temporal Pole	LH	827	5.54	−54	6	−10						446	5.21	−56	14	−8
			33	2.97	−28	2	−32										
		RH	880	5.9	54	6	−8	248	3.81	58	6	−8	378	4.66	54	10	−6
			152	3.69	34	4	−26										
Occipital	Cuneal Cortex	LH	167	3.8	−2	−76	22						279	4.2	−4	−84	32
		RH	161	3.68	2	−74	22						223	3.86	2	−78	36
	Intracalcarine Cortex	LH	365	4.14	−6	−76	10						353	3.15	−8	−78	10
		RH	533	4.36	24	−64	8						397	4.1	18	−62	6
	Lateral Occipital Cortex inferior	LH	215	3.47	−42	−84	−24						194	3.72	−56	−66	−10
													57	3.97	−56	−64	12
		RH	52	3.08	60	−68	−10										
	Lateral Occipital Cortex superior	LH											516	3.95	−12	−62	56
	Occipital Fusiform Gyrus	LH	204	5.34	−16	−68	−20						225	5.07	−16	−68	−20
			29	3.43	−10	−92	−24										
			27	2.65	−46	−68	−24										
		RH	270	6.19	16	−68	−20						328	5.19	22	−62	−20
	Occipital Pole	LH	41	3.28	−8	−94	−24						144	3.6	−2	−92	0
													59	3.56	0	−94	−16
		RH	35	3.14	10	−92	−24						271	2.96	8	−92	6
													76	4.22	6	−94	−20
	Supracalcarine Cortex	LH	53	3.55	0	−74	20						44	2.99	0	−74	20
		RH	45	3.15	2	−74	18						26	2.9	2	−78	10
Medial	Amygdala	LH	296	4.1	−22	−4	−12						55	3.16	−20	−14	−12
		RH	261	4.28	22	−2	−14										
	Hippocampus	LH	226	3.97	−18	−24	−12						116	3.6	−20	−24	−10
	Insular Cortex	LH	634	4.66	−36	−10	12						149	3.91	−36	−10	10
													51	3.04	−42	6	0
		RH	630	4.02	36	8	2						26	2.81	32	14	8
													25	3.38	46	8	−4
	Parahippocampal Gyrus anterior	LH	303	3.76	−20	2	−16										
		RH	142	4.12	22	−2	−16										
	Parahippocampal Gyrus posterior	LH	251	4.57	−2	−44	−2						81	3.45	−20	−24	−12

(*continued on next page*)

Table 2 (*continued*)

Study Phase			Aloud-Silent					Aloud-Control					Control-Silent				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
													44	3.28	−36	−32	−12
													26	4.45	−2	−44	−4
		RH	69	3.88	8	−36	−2										
Basal Ganglia	Caudate	LH	60	3.35	−14	12	10										
			25	3.45	−8	0	10										
	Pallidum	LH	194	3.79	−26	−16	−4						51	3.19	−24	−4	−4
		RH	152	4.05	22	−2	−4						84	3.66	20	−2	−2
	Putamen	LH	536	4.22	−30	−16	−4						139	3.4	−30	−10	−6
		RH	466	4.01	22	8	6						229	3.21	26	8	−6
Midbrain	Thalamus	LH	722	4.44	−8	−4	8						302	3.63	−14	−16	0
		RH	771	4.2	10	−4	8						381	3.96	8	−2	2

Table 3
Activation coordinates for each condition relative to baseline during the test phase, constrained to areas that were active for the same contrasts during the study phase (* but note that the Foil-baseline contrast was not constrained to study phase activation).

Conjunction Study & Test Phase			Aloud-baseline					Silent-baseline					Control-baseline					Foil-baseline *				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
Frontal	Cingulate Gyrus anterior	LH	504	6.07	−4	16	38						346	5.46	−4	16	38	130	4.23	−4	12	40
		RH	175	5.18	4	20	36						37	4.44	2	18	36	103	3.45	6	8	44
	Frontal Operculum Cortex	LH	332	5.16	−46	16	0						296	5.55	−50	12	−4					
		RH	63	4.24	38	24	0						45	4.12	48	16	2					
	Frontal Orbital Cortex	LH	464	5.68	−32	26	−2						611	5.12	−32	24	−4	54	3.08	−46	16	−8
			41	3.17	−20	10	−10															
		RH	85	4.72	34	28	−4															
			49	4.25	18	4	−12															
	Frontal Pole	LH	89	3.96	−46	40	8	27	3.37	−42	44	8	1882	5.01	−42	48	8					
	Inferior Frontal Gyrus pars opercularis	LH	551	5.72	−60	12	20	220	3.95	−36	18	22	1143	5.83	−50	14	−4	366	3.86	−50	10	28
		RH	95	3.75	60	12	16						48	3.93	50	16	0					
			27	3.6	52	16	−2															
	Inferior Frontal Gyrus pars triangularis	LH	245	4.31	−50	18	−6	264	4.66	−50	28	22	923	4.73	−50	26	24	126	3.33	−46	28	18

(*continued on next page*)

Table 3 (*continued*)

Conjunction Study & Test Phase			Aloud-baseline					Silent-baseline					Control-baseline					Foil-baseline *				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
		RH											46	3.94	52	18	−4					
	Middle Frontal Gyrus	LH	210	5.87	−34	−4	58	546	4.89	−46	26	26	2157	5.38	−34	−4	58	431	4.77	−32	−4	56
								357	5.03	−36	−4	58						76	3.09	−46	24	24
		RH	42	4.21	34	−2	62						206	4.7	34	−2	50	60	3.19	36	−2	60
	Paracingulate Gyrus	LH	342	6.69	−2	14	52	379	5.69	−2	18	40	589	6.57	−4	20	42	379	4.91	−2	14	48
		RH	116	5.91	2	12	52	29	4.36	2	18	42	103	5.29	2	20	38	316	4.52	2	10	48
	Precentral Gyrus	LH	2697	6.41	−50	2	34	784	5.52	−36	−8	58	2439	5.41	−46	4	42	2862	5.24	−52	−2	36
			230	3.96	−6	−14	70						139	4.01	−6	−14	70					
		RH	586	4.75	58	6	28						501	4.72	30	−4	48	305	3.56	40	−8	60
			311	4.77	38	−10	60															
	Superior Frontal Gyrus	LH	805	6.39	−2	14	54	223	4.32	−6	22	48	892	5.45	−4	24	48	165	3.64	−28	−4	62
																		74	4.24	−2	16	52
																		48	3.15	−8	−6	68
		RH	239	5.65	2	14	54						78	4.33	2	18	52	162	3.38	10	−4	70
			34	4.04	30	−4	62						31	3.49	30	−4	60					
	Supplementary Motor Area	LH	804	5.84	−2	2	62	246	4.25	−4	−2	66	706	4.72	−2	−2	64	503	4.36	0	6	50
		RH	494	5.45	2	6	58	45	3.25	4	6	50	320	3.95	2	8	52	379	4.42	2	6	50
Parietal	Angular Gyrus	LH						156	5.04	−38	−56	38	334	5.85	−36	−56	36					
	Central Opercular Cortex	LH	282	4.57	−46	2	14						346	4.6	−50	8	−2	214	4.33	−54	−20	20
			90	4.36	−56	−22	20											28	2.6	−40	−6	12
	Cingulate Gyrus posterior	LH	79	4.03	0	−40	2						47	3.34	0	−32	6					
		RH	29	3.57	2	−40	2															
	Parietal Operculum Cortex	LH											48	3.95	−48	−28	14	240	4.28	−50	−24	18
	Postcentral Gyrus	LH	880	5.03	−46	−20	52						774	5.3	−48	−24	36	2304	4.6	−38	−28	64
			184	3.92	−2	−40	70						199	3.74	−6	−46	66					
		RH	105	4.17	56	−14	28						107	3.95	52	−16	40	61	3.33	42	−36	52
			28	3.24	2	−38	70						49	3.93	42	−18	50					
	Precuneous Cortex	LH	187	3.94	−2	−82	56						563	5.16	−6	−72	48					
		RH											59	3.74	2	−76	42					
	Superior Parietal Lobule	LH											486	6.4	−42	−44	46	1038	4.79	−28	−58	48
		RH																377	3.75	42	−38	52
	Supramarginal Gyrus anterior	LH											94	6.44	−46	−40	46	839	4.74	−48	−36	46
		RH																55	3.11	46	−32	46
	Supramarginal Gyrus posterior	LH	83	3.57	−46	−42	30	41	5.04	−46	−48	46	538	6.52	−44	−42	46	226	4.26	−46	−40	48
													45	3.3	−48	−50	12	81	3.63	−56	−42	18
		RH																106	3.82	44	−38	52
Temporal	Inferior Temporal Gyrus posterior	LH	71	4.8	−46	−44	−20						113	4.39	−46	−44	−20	25	3.03	−46	−42	−20
	Inferior Temporal Gyrus temporooccipital	LH	724	6.28	−46	−62	−18	340	4.66	−50	−62	−24	646	6.21	−42	−62	−10	437	5.88	−46	−62	−18
		RH	187	4.62	48	−62	−20						129	4.61	46	−60	−16	225	4.31	46	−60	−16
	Lingual Gyrus	LH	750	5.5	−12	−90	−12	176	5.18	−12	−90	−12	389	5.18	−12	−90	−8	472	5.74	−12	−90	−12
		RH	760	6.85	8	−78	−26	142	5.9	14	−88	−6	497	6.45	16	−88	−8	345	5.82	16	−88	−8

(*continued on next page*)

Table 3 (*continued*)

Conjunction Study & Test Phase			Aloud-baseline					Silent-baseline					Control-baseline					Foil-baseline *				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
	Middle Temporal Gyrus posterior	LH	154	4.17	−66	−34	−2						430	3.82	−68	−38	−4					
	Middle Temporal Gyrus temporooccipital	LH	170	4.72	−56	−52	6						471	3.66	−56	−54	4	50	3.96	−44	−60	0
	Planum Temporale	LH																29	3.66	−54	−42	18
	Superior Temporal Gyrus posterior	LH	121	4.07	−62	−38	2						38	3.28	−66	−34	0					
	Temporal Fusiform Cortex posterior	LH	330	6.76	−36	−50	−28	56	3.65	−44	−44	−18	295	5.69	−36	−50	−28	257	5.75	−36	−48	−28
		RH	81	5.22	36	−42	−30						80	4.97	34	−40	−30	70	4.05	36	−42	−30
	Temporal Occipital Fusiform Cortex	LH	686	6.56	−36	−50	−26	466	5	−42	−58	−26	636	6.3	−42	−60	−16	677	6.23	−42	−62	−22
		RH	671	6.52	34	−50	−26	358	5.34	40	−58	−26	627	5.82	36	−52	−26	655	5.45	40	−60	−22
	Temporal Pole	LH	419	5.02	−54	14	−6						401	5.68	−52	12	−4	277	3.99	−54	12	−10
		RH	176	3.93	58	14	−8						134	3.92	52	16	−4					
Occipital	Cuneal Cortex	RH	43	3.85	2	−88	46															
	Intracalcarine Cortex	LH	164	4.4	−18	−88	2	41	4	−18	−88	2	66	4.35	−18	−88	2	268	4.82	−18	−88	2
		RH	84	2.93	10	−72	4	30	4.38	12	−88	0	57	4.51	12	−88	0	185	4.21	12	−88	0
			61	3.91	12	−88	0															
	Lateral Occipital Cortex inferior	LH	1983	7.21	−36	−86	−8	1437	6.2	−30	−86	−8	1886	6.84	−30	−86	−10	2058	7.21	−30	−84	−8
		RH	1966	7.09	32	−86	−14	1474	6.15	30	−86	−6	1833	6.46	32	−84	−6	1839	6.33	38	−86	−8
	Lateral Occipital Cortex superior	LH	220	4.32	−32	−88	8	617	4.57	−30	−66	36	1867	5.41	−28	−66	40	656	4.93	−28	−60	50
			25	3.98	−4	−84	52	63	3.81	−26	−88	8	170	4.4	−30	−88	10	161	4.4	−22	−90	8
		RH	218	4.93	26	−88	6	129	4.81	26	−88	6	161	5.1	26	−88	6	280	4.94	26	−88	6
								114	3.71	30	−58	50	37	4.1	14	−70	52	127	3.29	30	−58	52
	Occipital Fusiform Gyrus	LH	1421	6.81	−30	−86	−18	1246	6.41	−20	−90	−10	1417	6.69	−28	−86	−10	1504	7.08	−28	−84	−8
		RH	1380	7.48	24	−88	−12	1054	5.96	30	−84	−6	1350	6.47	20	−88	−10	1382	6.3	22	−88	−12
	Occipital Pole	LH	1918	6.92	−16	−94	−8	1783	6.53	−18	−92	−10	1680	7.31	−26	−92	−2	2103	7.06	−18	−94	−4
		RH	1891	6.83	24	−90	−6	1460	6.88	14	−92	−8	1322	6.28	22	−94	0	1834	6.9	24	−92	−4
Medial	Amygdala	LH	156	4.22	−20	−10	−12						71	3.36	−16	−10	−14					
		RH	50	3.45	14	0	−16															
	Hippocampus	LH	239	4.25	−22	−26	−8						78	3.83	−22	−30	−6					
		RH	39	3.78	20	−28	−8															
	Insular Cortex	LH	566	5.77	−34	22	0						368	5.31	−32	22	−2	204	3.45	−32	−6	12
													144	3.8	−38	−4	12					
		RH	326	5.22	32	20	0						288	4.74	40	18	−4					
	Parahippocampal Gyrus anterior	LH	51	3.39	−16	−6	−18															
	Parahippocampal Gyrus posterior	LH	317	4.28	−10	−32	−8						103	3.83	−24	−32	−6					
		RH	132	3.78	20	−28	−8															
Basal Ganglia	Caudate	LH	364	5.74	−14	2	14						202	4.73	−10	−2	12					
		RH	189	5.01	12	12	4															
	Pallidum	LH	173	4.72	−12	4	−2						103	3.84	−16	6	0	34	3.05	−24	−16	0
		RH	88	3.91	16	8	−2															
	Putamen	LH	600	4.59	−16	10	−2						576	4.45	−26	4	−4	116	3.55	−26	−10	12
		RH	338	4.45	18	6	−10						384	3.64	24	10	4					
Midbrain	Thalamus	LH	1039	5.31	−10	−2	10						684	4.84	−10	−4	10	194	3.54	−10	−18	4
		RH	663	5.1	2	−20	10						396	4.39	2	−14	14					

Table 4
Activation coordinates for contrasts between aloud, silent, and sensorimotor control conditions during the test phase, constrained to areas that were active for the same contrasts during the study phase.

Conjunction Study & Test Phase			Aloud-Silent					Aloud-Control					Control-Silent				
Lobe	ROI	Hemisphere	Cluster size (mm3)	Max z	x	y	z	Cluster size (mm3)	Max z	x	y	z	Cluster size (mm3)	Max z	x	y	z
Frontal	Middle Frontal Gyrus	LH											25	3.31	−32	0	46
	Precentral Gyrus	LH	81	3.56	−48	4	16						25	2.78	−46	−6	40
Parietal	Central Opercular Cortex	LH	60	3.5	−42	2	6										
	Postcentral Gyrus	LH											25	3.06	−44	−20	46
Temporal	Lingual Gyrus	LH	60	2.79	0	−80	−22										
		RH	137	3.16	4	−80	−20										
			92	3.4	14	−60	−14										
	Superior Temporal Gyrus posterior	LH	41	3.22	−60	−38	2										
Occipital	Lateral Occipital Cortex superior	LH											27	2.88	−30	−64	46
	Occipital Fusiform Gyrus	RH	34	3.28	30	−72	−26										
			25	3.05	14	−84	−24										
Medial	Insular Cortex	LH	40	3.65	−40	2	6										

then compared between the two halves to gain an estimate of internal consistency. The (Spearman-Brown corrected) mean split-half internal consistency for combined "old" judgments was r_{SB} = . 0.35, 95% CI [−0.06, 0.65]; for recollection r_{SB} = 0.44, 95% CI [0.07, 0.72].[5]

3.2. Neuroimaging

3.2.1. Study phase

Activation for each condition was first contrasted with fixation baseline, to identify the broad networks engaged in each condition. Details of these results are provided in Table 1. All three conditions—aloud, silent, and sensorimotor control—were associated with extensive bilateral cortical activation (more extensive in the left hemisphere) in all lobes of the cortex as well as the midbrain, but the silent condition was associated with relatively less extensive activity. Similar regions were activated in all three conditions but an absence of activation in the silent condition was notable in the following regions: auditory processing regions of the superior and middle temporal gyri (including Heschl's gyrus and the planum polare); the hippocampus, parahippocampal gyri, basal ganglia and amygdalae; posterior medial cortical regions including the supra-calcarine cortex, parietal operculum, and posterior cingulate gyri.

The contrasts between conditions, shown in Fig. 2 and Table 2, focused on our research questions. We first consider brain areas activated significantly more in the aloud condition than in the control condition (which was similar in terms of motor activity and auditory perception of self-generated speech). Reading the study words aloud, compared to saying the word "check" while reading them, yielded greater activation along the superior temporal cortex bilaterally, including areas consistent with primary and secondary auditory cortices (including Heschl's gyrus and the planum temporale) as well as more anterior regions often associated with speech processing. Activation was also found in the inferior motor cortex region (pre- and post-central gyri and central opercula), consistent with areas associated with speech articulators; this activation was bilateral but more extensive in the right hemisphere.

The contrasts for the aloud and control conditions against the silent condition yielded relatively more extensive differences, but these two contrasts yielded generally similar patterns of activation as shown in Fig. 2, with details provided in Table 2. For both contrasts, the central sulci (primary motor/sensory cortices) and superior temporal gyri were more strongly and extensively activated bilaterally, relative to the aloud versus control contrast. Activation in these contrasts also extended into additional areas including the frontal lobes (bilateral IFG, SFG, SMA, left MFG), inferior and superior parietal regions (bilateral SMG; left SPL and left AG were activated only for the control-silent contrast), and temporal-occipital areas including the middle and inferior temporal gyri, lateral occipital cortex, fusiform and lingual gyri. Extensive medial and subcortical activation was also obtained, including in the hippocampi, parahippocampal gyri, amygdalae, cingulate gyri (anterior and posterior), and basal ganglia (putamen, pallidum; caudate for the aloud-silent contrast only).

Given that the behavioral production effect was characterized by better memory for aloud items than either silent or control items, we performed a conjunction analysis to identify the brain regions that were significantly activated in the contrasts of the aloud condition against each of the other conditions; this analysis is shown in the bottom right panel of Fig. 2. The brain areas consistently associated with the "production effect" contrasts during study were those identified in the aloud versus control contrast—motor and auditory cortices—confirming that these were a subset of the regions identified in the control versus silent contrast.

[5] We did not compute split-half reliability for familiarity because calculating the familiarity-based production effect requires the exclusion of all trials with a "recollect" response. Therefore, reliability calculations for familiarity would be based on an extremely low number of trials (as low as 6 trials per half for some participants). More generally, our reported estimates may be imprecise, and should therefore be viewed with caution, again owing to the low number of trials involved in these calculations (15 trials for each condition in each half). Indeed, this experiment was not designed to accurately assess reliability and split-half calculations were undertaken only because no other reports of reliability for this measure are available within the literature. The robustness of our behavioural results is further supported by the fact that the experiment was sufficiently powered to detect expected effect sizes, and that our results are entirely consistent with prior literature.

Table 5
Activation coordinates for aloud, silent, and sensorimotor control conditions relative to foil items during the test phase, constrained to areas that were activated by the minuend condition relative to baseline at study.

Conjunction Study & Test Phase				Aloud-Foil				Silent-Foil					Control-Foil				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
Frontal	Cingulate Gyrus anterior	LH	229	3.8	−2	22	34						109	3.45	−4	24	32
		RH	42	3.44	2	22	34										
	Frontal Operculum Cortex	LH	231	4.06	−42	16	4						87	3.3	−50	12	−4
	Frontal Orbital Cortex	LH	256	4.05	−38	22	−4						322	3.8	−32	24	−4
	Frontal Pole	LH	83	4.57	−44	40	6	32	3.25	−42	44	8	1459	4.85	−34	60	4
	Inferior Frontal Gyrus pars opercularis	LH	491	4.27	−54	14	22						771	4.09	−34	16	22
	Inferior Frontal Gyrus pars triangularis	LH	252	3.92	−44	36	8	30	2.92	−48	28	22	726	4.07	−46	34	16
	Middle Frontal Gyrus	LH	130	4.56	−38	0	50	431	4.49	−44	22	34	1742	4.65	−48	32	24
		RH											58	2.81	32	4	64
	Paracingulate Gyrus	LH	221	4.28	0	22	40						374	4.22	−4	28	38
		RH	50	3.9	2	20	40										
	Precentral Gyrus	LH	1080	4.36	−40	−2	50						848	4.01	−36	4	32
			81	3.13	−4	−14	76										
	Superior Frontal Gyrus	LH	400	4.51	−2	10	66						118	3.97	−6	24	48
		RH	93	4.07	2	10	62										
	Supplementary Motor Area	LH	374	4.59	−2	8	66										
		RH	126	4.06	2	8	62										
Parietal	Angular Gyrus	LH						171	4.26	−40	−56	40	387	5.2	−38	−56	38
	Central Opercular Cortex	LH	70	3.59	−44	4	14										
	Cingulate Gyrus posterior	LH	67	4.12	0	−40	2						50	3.49	0	−32	6
		RH	42	3.77	2	−40	2										
	Postcentral Gyrus	LH	125	3.5	−4	−50	72						73	3.3	−42	−22	48
			29	3.35	−38	−30	46						60	3.15	−8	−48	68
	Precuneous Cortex	LH	183	3.49	−2	−80	58						483	4.09	−8	−60	60
		RH											26	3.12	6	−64	58
	Superior Parietal Lobule	LH											315	4.61	−34	−58	40
													26	3.19	−12	−58	60
	Supramarginal Gyrus anterior	LH											90	4.24	−50	−40	48
	Supramarginal Gyrus posterior	LH						40	3.53	−46	−48	46	494	4.53	−42	−52	42
Temporal	Inferior Temporal Gyrus posterior	LH											52	3.06	−68	−44	−18
	Inferior Temporal Gyrus temporooccipital	LH	419	4.1	−56	−54	−20						30	3.07	−64	−48	−14
		RH	55	3.38	54	−58	−28										
	Lingual Gyrus	LH	132	3.79	−2	−82	−16										
		RH	337	4.75	6	−78	−24						47	3.96	6	−78	−26
	Middle Temporal Gyrus posterior	LH	127	4.26	−62	−34	−2						444	4.01	−68	−40	−4
	Middle Temporal Gyrus temporooccipital	LH	94	3.7	−60	−46	4						187	3.53	−58	−44	−8
	Superior Temporal Gyrus posterior	LH	59	4.04	−64	−34	0										
	Temporal Fusiform Cortex posterior	LH	26	3.06	−26	−38	−28										
	Temporal Occipital Fusiform Cortex	LH	126	3.44	−48	−58	−24										
		RH	148	4.17	28	−48	−22										
	Temporal Pole	LH	90	3.91	−56	18	−8						146	3.41	−54	20	−14
Occipital	Cuneal Cortex	RH	31	3.2	4	−92	40										
	Lateral Occipital Cortex inferior	LH	153	3.13	−56	−78	−10						25	3.1	−62	−62	−12
		RH	398	4.3	38	−74	−28						103	3.62	40	−72	−28
	Lateral Occipital Cortex superior	LH						505	3.74	−38	−60	42	1350	5.32	−38	−66	46
		RH											30	3.23	14	−72	56
	Occipital Fusiform Gyrus	RH	479	5.01	10	−82	−24						210	3.92	10	−74	−24
	Occipital Pole	RH	60	3.41	2	−92	36										
Medial	Hippocampus	LH	120	3.45	−22	−26	−12										
	Insular Cortex	LH	234	4.39	−36	20	−2						260	4.33	−32	20	−4

(*continued on next page*)

Table 5 (*continued*)

Conjunction Study & Test Phase				Aloud-Foil				Silent-Foil				Control-Foil				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
		RH	38	3	32	20	4						26	2.92	28	16	4
	Parahippocampal Gyrus posterior	LH	231	3.83	−2	−40	0										
		RH	62	3.16	12	−34	−4										
Basal Ganglia	Caudate	LH	330	5.26	−10	4	12						156	4.68	−10	4	8
		RH	166	4.26	10	10	6							3.97	−14	4	2
	Pallidum	LH	64	4.74	−12	4	0						39	3.55	−16	8	−2
	Putamen	LH	111	3.84	−16	8	−2						106	3.19	−32	−14	−8
													59	2.77	24	12	2
		RH											26	4.41	−10	−2	10
Midbrain	Thalamus	LH	720	4.84	−2	−18	12						338	3.2	−16	−20	10
													40	3.73	2	−16	14
		RH	345	4.44	2	−18	12						215				

Table 6
Activation coordinates from the Aloud-baseline contrast during the study phase correlating with behavioural performance (recollection success) at test.

Study Phase Lobe	ROI	Hemi	Aloud-baseline vs "recollect" accuracy				
			Cluster size (mm3)	Max z	x	y	z
Frontal	Precentral Gyrus	LH	41	3.62	−56	2	8
Parietal	Central Opercular Cortex	LH	27	3.44	−54	2	6
	Parietal Operculum Cortex	LH	27	3.56	−48	−30	18

3.2.2. *Representational similarity analysis (RSA)*

Our RSA investigation of study phase data did not reveal any significant differences between conditions for individual study item activation patterns. However, we did observe some interesting non-significant trends, detailed in Table S2. Aloud items exhibited higher correlations with a phonological model in frontal ROIs (left SMA, right IFG and right precentral gyrus) when compared to silent items, and in temporal ROIs (left posterior ITG and MTG) as well as the occipital pole when compared to control items.

3.2.3. *Test phase*

Because our primary interest with respect to the test phase concerned activation reflecting encoding processes (i.e., reinstatement), activation from all contrasts at test was masked with the same contrasts at study (with the exception of the foil–baseline contrast). Non-constrained results from the test phase are described in supplementary materials; activation coordinates are reported in Tables S3, S4, and S5.

Activation for each type of item at test relative to fixation is reported in Table 3. Relative to baseline, all item types elicited extensive activation across all lobes of the cerebral cortex and subcortical regions. Notably, much less activation was obtained in the test phase, relative to the study phase, in superior temporal lobe regions associated with auditory processing.

Contrasts between conditions are shown in Fig. 3 and Table 4. No differences were observed for the aloud-control contrast. For the aloud-silent contrast, activation was present in inferior motor cortex (left precentral gyrus and central operculum) and temporo-occipital (bilateral lingual gyrus, left posterior STG) cortices associated with vocalization and auditory processing. This contrast also yielded focal activation in the fusiform gyrus on the inferior temporal lobe. The control-silent yielded a similar, though less extensive pattern of activation, including clusters in left MFG, pre- and postcentral gyri, and superior lateral occipital cortex.

Additional contrasts were made for each study condition relative to the recognition test foils. The contrasts are shown in Fig. 4 and Table 5. Notably, activation elicited by aloud and control conditions relative to foil yielded bilateral activation (though more extensive in the left hemisphere) in frontal (SFG, MFG, IFG), sensorimotor (pre- and postcentral gyri), and temporal cortices (MFG, temporal pole). Moreover, the aloud-foil contrast elicited more extensive activation in frontal, temporal and parietal cortices whereas activation for the control-foil contrast was more extensive in the parietal lobe (IPL, SMG). The silent-foil contrast yielded more focal activation, with significant clusters in frontal (IFG and MFG), parietal (AG, SMG), and occipital (superior lateral occipital cortex) cortices in the left hemisphere. No areas were commonly activated by all three studied conditions relative to foil items (but see supplementary materials for a description of commonly activated areas when test phase results were not constrained to areas activated in the study phase).

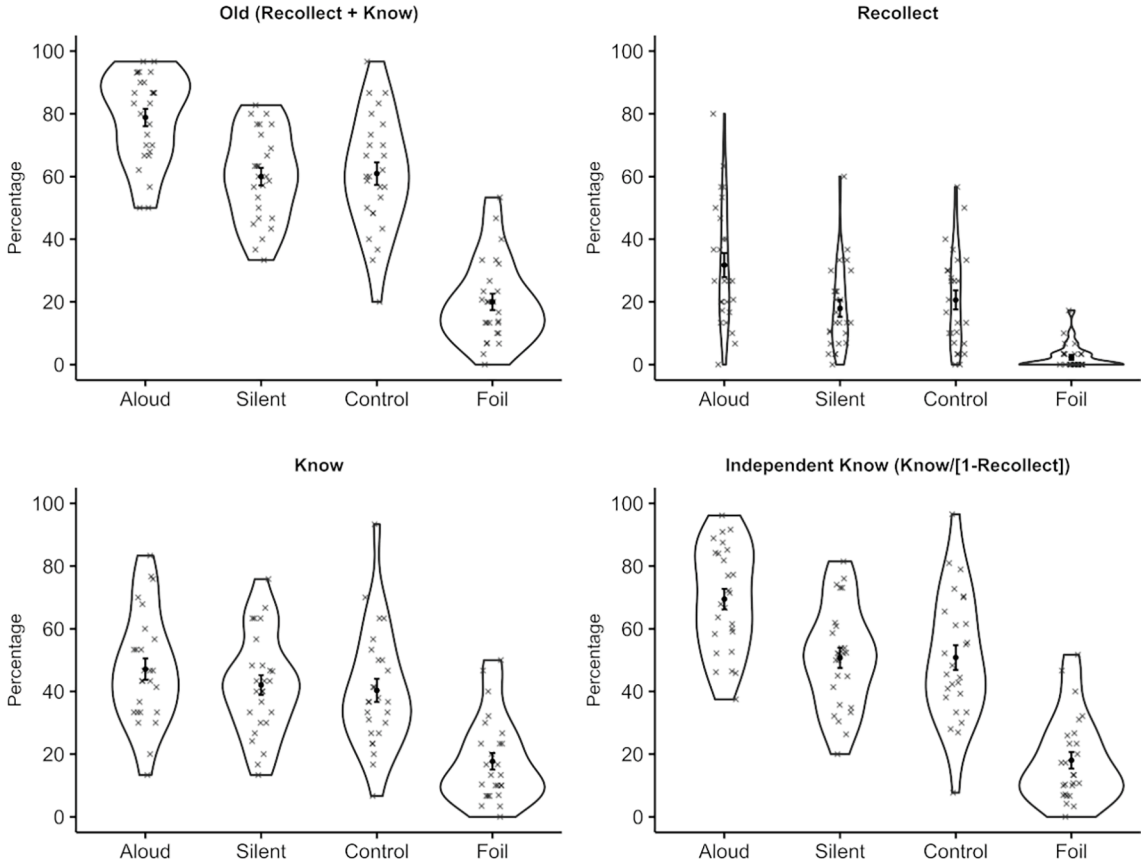


Fig. 1. Mean old responses and separate recollect, know, and independent know responses (%) as a function of item type (aloud, silent, control, foil). Violin plots and X's indicate the distribution of individual participant means. Fitted circles reflect the empirical means; error bars reflect the standard error of the mean. X's have been jittered in the horizontal plane to make them more easily distinguishable.

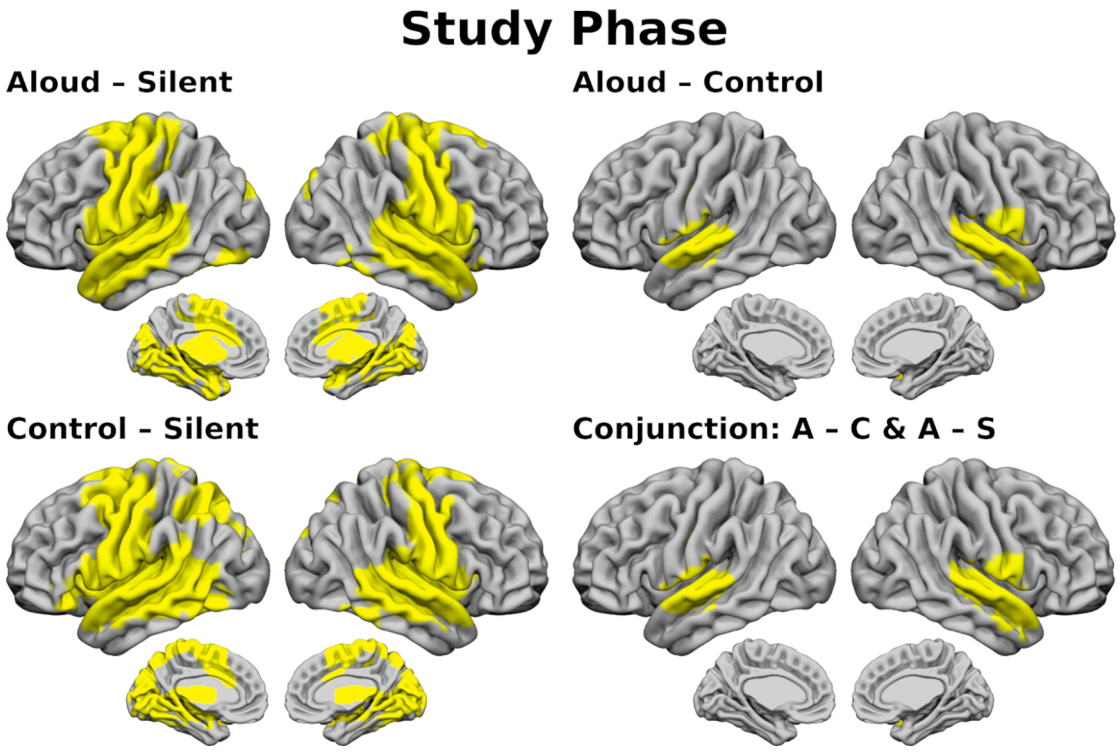


Fig. 2. Differences in activation between conditions in the study phase. Contrasts show aloud items relative to silent items (top left panel), aloud items relative to sensorimotor control items (top right), sensorimotor control items relative to silent items (bottom left), and conjunction of aloud relative to sensorimotor control and aloud relative to silent (bottom right).

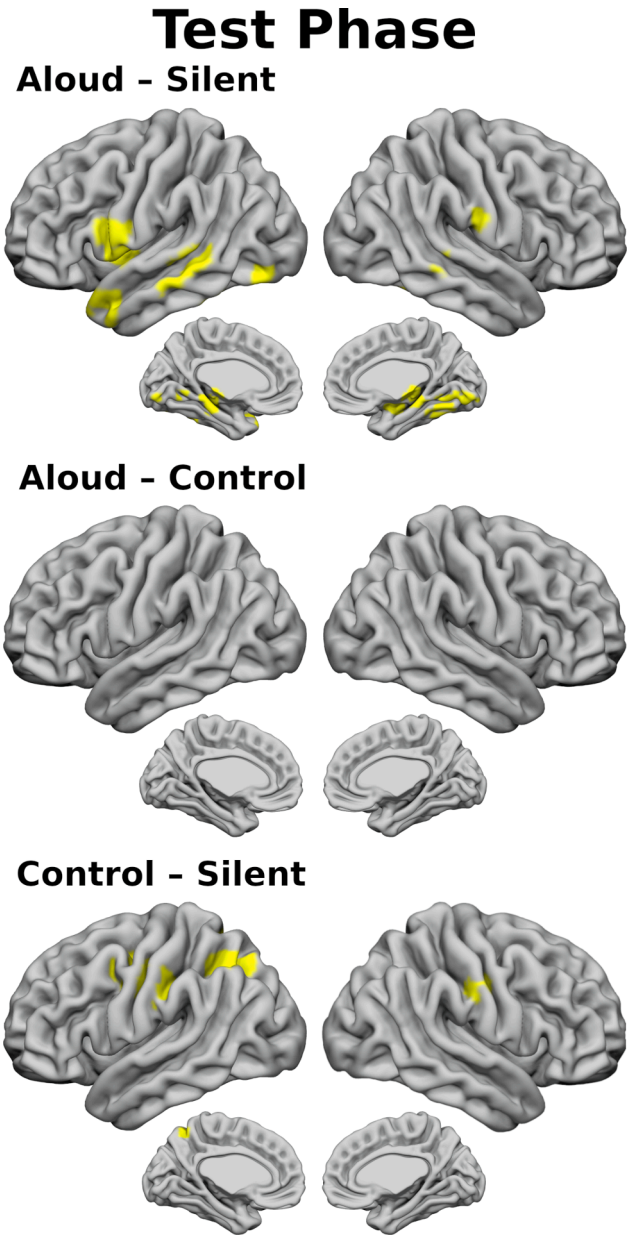


Fig. 3. Differences in activation between conditions in the test phase, constrained to areas showing activation for the same contrasts in the study phase. Contrasts show aloud items relative to silent items (top panel), aloud items relative to sensorimotor control items (middle panel), sensorimotor control items relative to silent items (bottom panel).

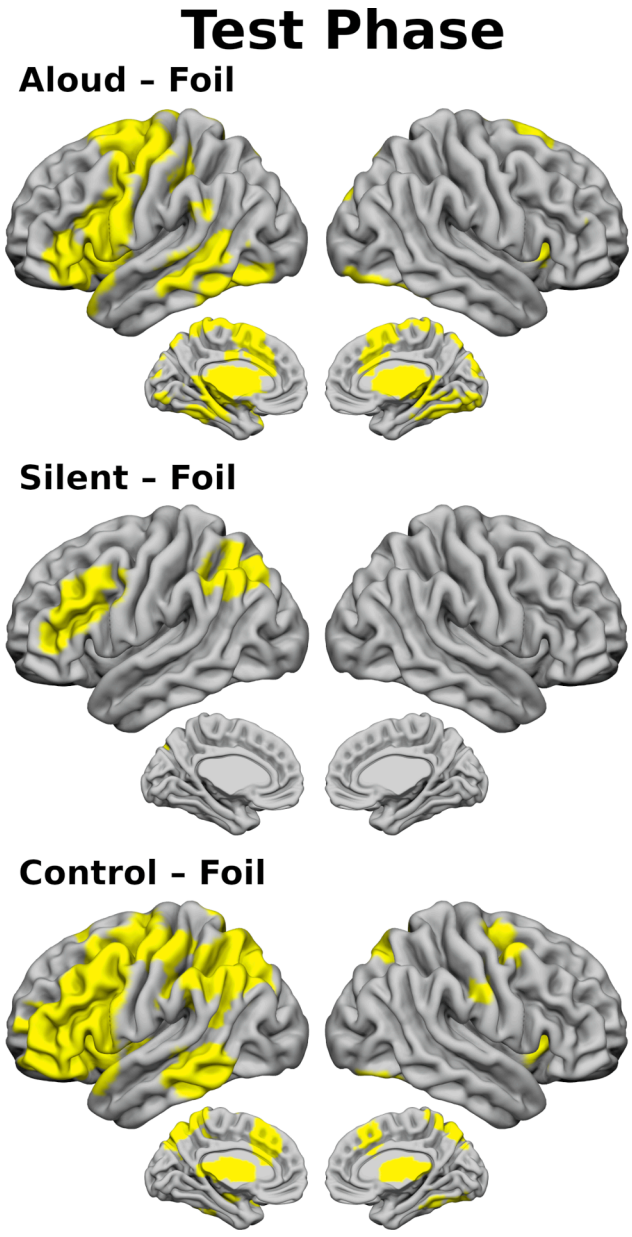


Fig. 4. Differences in activation between foil items and all other conditions in the test phase. Contrasts show aloud items relative to foil (top panel), silent items relative to foil (middle panel), and sensorimotor control items relative to foil (bottom panel).

3.2.4. Brain-Behavior correlations

We also investigated correlations between brain activation and both (1) overall performance in each condition (aloud, control, silent) as indexed by recollect and know judgments (separately), as well as for the combined 'old' (recollect + know) judgments, each corrected for false alarms; and, (2) the behavioural production effect as indexed by recollect and know judgments (separately), as well as for the combined 'old' (recollect + know) judgments. Significant correlations from these analyses (surviving multiple comparison correction) are shown in Fig. 5 and Table 6.

With respect to successful recollection, activation in the aloud-baseline contrast during the study phase significantly correlated with recollect judgments in left inferior motor cortex regions (precentral gyrus, central and parietal opercular cortices). No other significant correlations were present with respect to the aloud or silent conditions relative to baseline. Moreover, there were no significant correlations with respect to either the aloud-silent or aloud-control contrast.

4. Discussion

The present study is the first to use fMRI to characterize the neural mechanisms giving rise to the production effect. Participants studied subsets of words aloud, silently, or by making a non-unique verbal "check" response (sensorimotor control condition), followed by a recognition memory test. With respect to behavioural findings, a production effect was obtained, with greater recollection and familiarity ratings for the aloud items than for the silent items. These findings replicate earlier work in this area (Fawcett & Ozubko, 2016; Ozubko et al., 2012). Our primary focus, however, was in understanding the neural correlates of the production effect. In this respect, a distinctiveness-based account predicted that reading aloud, relative to

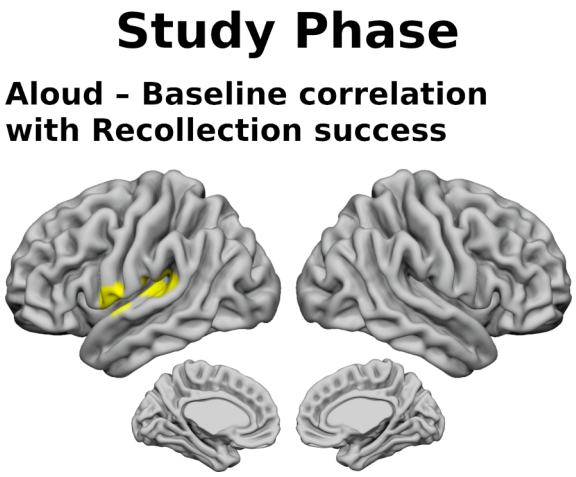


Fig. 5. Coloring indicates activation from the aloud–baseline contrast during the study phase that significantly correlated with recollection success for aloud items at test.

the other conditions, would result in stronger activation of sensorimotor and phonological regions both at study (reflecting the encoding of these features) and test (reflecting their retrieval). This account was supported by the data, as we summarize below according to the experimental phase.

4.1. Study phase

Contrasting the three encoding conditions against the fixation baseline revealed activation in brain regions typically associated with encoding in verbal memory tasks, along with regions reflecting the relative sensorimotor demands of the tasks. These activations included inferior frontal (including IFG and MFG), premotor, and parietal (IPS and SMG) regions. Occipital and inferior temporal cortex were also activated (consistent with visual presentation of words), as was the posterior/middle superior temporal sulcus region (consistent with lexical processing). The conditions involving motor speech output (aloud and control) activated motor and somatosensory cortex along the central sulcus, basal ganglia (also associated with sensorimotor control), and auditory processing regions in the superior temporal gyrus and extending into inferior parietal and middle temporal regions. Hippocampal and parahippocampal gyri were also more activated in the aloud and control conditions, which we attribute to greater allocation of attention in those conditions (see our discussion of study contrasts).

Our primary goal was to identify brain regions that showed differential activation for the contrasts between the aloud condition and the two other conditions – those revealing the neural basis of the production effect. The aloud–silent contrast elicited more extensive activation than the aloud–control contrast, but recognition did not differ credibly between the silent and control conditions. Therefore, differences between these contrasts are not likely due to factors driving the production effect. In particular, the stronger activation in motor and auditory cortices for the aloud and control conditions likely reflected these tasks' recruitment of motor speech production and auditory perception by both of these tasks.

For this reason, we focus on areas that were consistent in the contrasts of the aloud condition with the other two conditions. Most extensively, the superior temporal lobe bilaterally was most activated in the aloud condition, from the planum temporale through Heschl's gyrus (primary auditory cortex) to anterior regions commonly associated with speech processing (Hickok & Poeppel, 2016; Venezia et al., 2017), including the superior temporal sulcus and part of the right middle temporal gyrus. Stronger activation was also obtained for aloud items in the inferior parts of the precentral and postcentral gyri (bilaterally although more extensive in the right hemisphere) – areas involved in the motor control of speech. This pattern is consistent with a distinctiveness account in which the articulatory and sensory (auditory and somatosensory) experiences that occur during production are incorporated into the production record. Supporting this interpretation, activation in inferior motor regions correlated with recollection success for the aloud condition (relative to baseline).

Surprisingly, we did not observe correlations with respect to either contrast (aloud–silent or aloud–control) or the behavioural production effect. This may be because activation in the silent and control conditions are poorly correlated with test performance (indeed, neither the silent–baseline nor control–baseline contrasts yielded significant correlations with test performance on trials of their respective condition). Given that activation in the silent condition contributes to the aloud–silent contrast, this may have introduced noise (uncorrelated variance) which masked potential correlations with the behavioural production effect. A similar account may explain why the aloud–control contrast did not correlate with a behavioural production effect defined as performance on aloud trials minus control trials. Alternatively, the absence of correlations with between-conditions contrasts may be due to a lack of sensitivity, given that the magnitude of between-condition differences are inherently lower than for contrasts of conditions involving stimulus presentation and/or motor responses relative to a resting baseline. More generally, these correlation analyses should be viewed with some caution; they were largely exploratory, and it is also possible that the absence of correlations with respect to the production effect and/or either contrast was due to limited power for an fMRI investigation of individual differences (Dubois & Adolphs, 2016). As such, the significant correlation between aloud–baseline activation and recollection accuracy might be viewed as providing greater confidence in a distinctiveness account of our main findings (whereby sensorimotor activation facilitates later retrieval of aloud items), but may not be immediately informative as to the neural mechanisms underlying the production effect. This finding warrants replication with a larger sample size.

Finally, our multivariate analysis (RSA) indicated some interesting non-significant trends whereby activation for aloud items was more distinctive—evidenced by higher correlations with a phonological model—in areas associated with articulation (SMA, IFG, precentral gyrus) when compared to silent items; and areas associated with lexical processing (ITG, MTG) when compared to control items. Although these between-condition differences were non-significant, it is important to recognize that the study was not designed with this analysis in mind, and that these were exploratory post hoc analyses. As such, they provide further — however tentative — support for a distinctiveness account and warrant further investigation by future studies, perhaps with designs specifically tailored to MVPA (e.g., Zeithamova, de Araujo Sanchez, & Adke, 2017).

Our results are also consistent with the possibility that the production effect arises in part from increased attentional engagement or supplementary processing during aloud trials. Task-relevant activation in sensorimotor areas was greater for the aloud relative to the control condition, congruent with attentional up-regulation on aloud trials (e.g., Johansen-Berg & Matthews, 2002; Rinne et al., 2005; Rowe, Friston, Frackowiak, & Passingham, 2002). However, we cannot rule out the possibility that sensorimotor activation in the control condition was muted due to its repetitive nature. Moreover, activation was present in IFG and superior temporal gyrus in both the aloud and control conditions relative to silent. With respect to semantic processing, a meta-analysis of brain networks related to semantic comprehension of spoken and written language implicated both the IFG and superior temporal gyrus (Rodd, Vitello, Woollams, & Adank, 2015). Therefore, enhanced encoding and semantic processing of aloud items may generate more stable memory representations, facilitating later recollection.

Medial temporal regions such as the hippocampus were more active

during aloud than silent trials but, interestingly, they were also more active during control trials—despite recognition being similar in the control and silent conditions. Moreover, activation in medial temporal regions at study did not correlate with later recognition. Given that the hippocampus is often associated with successful encoding, and is also known to be modulated by attentional manipulations, this lack of correlation was unexpected. This suggests that encoding was enhanced for both aloud and control trials, but proved to be of little benefit for the control trials because the dominant feature in that episode (i.e., having said "check") was not deemed to be as diagnostic of prior study as retrieval of having said the actual test item aloud. Indeed, hippocampal activation in the aloud condition might reflect greater attention to and encoding of the stimuli, whereas hippocampal activation in the control condition might reflect greater attention to and encoding of the response. This possibility warrants further exploration.

4.2. Test phase

Surprisingly, we did not find differences in brain activation between the aloud and control conditions at test. However, activation of areas in the aloud–silent and control–silent contrasts that were, critically, also active during encoding may indicate reinstatement of task-related processes. In particular, activation of areas associated with articulation and auditory processing (somatosensory cortex and posterior STG) may reflect recollection of speech production for both aloud and control items. Importantly, such retrieval would only be diagnostic at test for aloud items; recollection of speaking a nonspecific word ("check") was likely insufficient to differentiate specific words from one another (evidenced by the absence of a behavioural production effect for the control condition). Moreover, activation of the fusiform gyrus (which houses the visual word form area) in the aloud–silent contrast may reflect more vivid recollection of reading the word during the study phase.

5. Conclusion

Producing items during study, particularly by reading them aloud, provides a simple and effective means of enhancing memory (MacLeod & Bodner, 2017). Our fMRI study explored the neural basis of the production effect. Our results are compatible with the dominant distinctiveness account, in demonstrating greater activation of primary sensorimotor cortex (associated with articulation) and auditory cortex (associated with perception) for produced than non-produced words during encoding. This account is further supported by our findings that activation in these regions correlated with later recognition only for produced items, and was somewhat more distinctive for aloud compared to silent and control items. However, our data also suggest that participants may be more engaged during aloud than silent trials. For example, they showed heightened activation of task-relevant regions on aloud trials, and greater recruitment of areas implicated in semantic processing. These differences emerged in analyses based only on items that were later correctly recognized, thus they were not artefacts of the aloud condition yielding proportionately better memory performance. Future studies should investigate these patterns of activation in more detail, for example by using network-based connectivity and/or designs more tailored to MVPA.

CRediT authorship contribution statement

Lyam M. Bailey: Conceptualization, Formal analysis, Data curation, Writing - review & editing, Visualization. **Glen E. Bodner:** Conceptualization, Methodology, Writing - review & editing. **Heath E. Matheson:** Conceptualization, Formal analysis. **Brandie M. Stewart:** Investigation, Writing - review & editing. **Kyle Roddick:** Conceptualization, Methodology, Investigation. **Kiera O'Neil:** Investigation. **Maria Simmons:** Investigation. **Angela M. Lambert:** Investigation. **Olave E. Krigolson:** Conceptualization, Methodology. **Aaron J. Newman:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Jonathan M. Fawcett:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Visualization, Supervision, Project administration, Funding acquisition.

Acknowledgements

We are grateful to the following people for their assistance in this study: Lauren Dunphy, Morgan Johnson, Greg MacLean, Matthew Rodgers, and Sarah Sullivan. Funding was provided by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) to GB, OEK, JF and AN, and from the Nova Scotia Health Research Foundation to OEK and AJN.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandc.2021.105757.

References

Andersson, J. L. R., Jenkinson M., & Smith, S. M. (2007a) Non-linear optimisation. FMRIB technical report TR07JA1.

Andersson J. L. R., Jenkinson M., & Smith S. M. (2007b). Non-linear registration, aka Spatial normalisation. FMRIB technical report TR07JA2.

Astafiev, S. V., Shulman, G. L., & Corbetta, M. (2006). Visuospatial reorienting signals in the human temporo-parietal junction are independent of response selection. *European Journal of Neuroscience, 23*(2), 591–596. https://doi.org/10.1111/j.1460-9568.2005.04573.x.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Special Issue: Emerging Data Analysis, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bodner, G. E., Taikh, A., & Fawcett, J. M. (2014). Assessing the costs and benefits of production in recognition. *Psychonomic Bulletin & Review, 21*(1), 149–154.

Bürkner, P.-C. (2017a). Advanced Bayesian multilevel modeling with the R package brms. ArXiv Preprint ArXiv:1705.11123.

Bürkner, P.-C. (2017b). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 1*(1), 2017. Retrieved from https://www.jstatsoft.org/v080/i01.

Champely, S. (2020). pwr: Basic Functions for Power Analysis. R package version 1.3-0.

Ciaramelli, E., Grady, C., Levine, B., Ween, J., & Moscovitch, M. (2010). Top-Down and Bottom-Up Attention to Memory Are Dissociated in Posterior Parietal Cortex: Neuroimaging and Neuropsychological Evidence. *The Journal of Neuroscience, 30*(14), 4943. https://doi.org/10.1523/JNEUROSCI.1209-09.2010.

Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language, 26*(3), 341–361. https://doi.org/10.1016/0749-596X(87)90118-5.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3*(3), 201.

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping, 8*(2–3), 109–114.

Dale, A. M., Greve, D. N., & Burock, M. A. (1999, June). Optimal Stimulus sequences for Event-Related fMRI. Presented at the 5th International Conference on Functional Mapping of the Human Brain. Duesseldorf, Germany.

Danker, J. F., & Anderson, J. R. (2010). The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychological bulletin, 136*(1), 87.

Danker, J. F., Tompary, A., & Davachi, L. (2017). Trial-by-trial hippocampal encoding activation predicts the fidelity of cortical reinstatement during subsequent retrieval. *Cerebral Cortex, 27*(7), 3515–3524.

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review, 8*(1), 155–161.

Dubois, J., & Adolphs, R. (2016). Building a science of individual differences from fMRI. *Trends in cognitive sciences, 20*(6), 425–443.

Fawcett, J. M. (2013). The production effect benefits performance in between-subject designs: A meta-analysis. *Acta Psychologica, 142*(1), 1–5. https://doi.org/10.1016/j.actpsy.2012.10.001.

Fawcett, J. M., Baldwin, M. M., Drakes, D. H. & Willoughby, H. V. (submitted for publication). Production improves recognition and reduces intrusions in between-subject designs: An empirical and meta-analytic investigation.

Fawcett, J. M., Lawrence, M. A., & Taylor, T. L. (2016). The representational consequences of intentional forgetting: Impairments to both the probability and fidelity of long-term memory. *Journal of Experimental Psychology. General, 145*(1), 56–81. https://doi.org/10.1037/xge0000128.

Fawcett, J. M., & Ozubko, J. D. (2016). Familiarity, but not recollection, supports the between-subject production effect in recognition memory. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale, 70*(2), 99–115. https://doi.org/10.1037/cep0000089.

Fawcett, J. M., Quinlan, C. K., & Taylor, T. L. (2012). Interplay of the production and picture superiority effects: A signal detection analysis. *Memory (Hove, England), 20*(7), 655–666. https://doi.org/10.1080/09658211.2012.693510.

Forrin, N. D., & MacLeod, C. M. (2016). Order information is used to guide recall of long lists: Further evidence for the item-order account. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 70*(2), 125.

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition, 40*(7), 1046–1055. https://doi.org/10.3758/s13421-012-0210-8.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gottfried, J. A., Smith, A. P., Rugg, M. D., & Dolan, R. J. (2004). Remembrance of odors past: Human olfactory cortex in cross-modal recognition memory. *Neuron, 42*(4), 687–695.

Hassall, C. D., Quinlan, C. K., Turk, D. J., Taylor, T. L., & Krigolson, O. E. (2016). A preliminary investigation into the neural basis of the production effect. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale, 70*(2), 139–146. https://doi.org/10.1037/cep0000093.

Hickok, G., & Poeppel, D. (2016). Chapter 25—Neural Basis of Speech Perception. In G. Hickok & S. L. Small (Eds.), Neurobiology of Language (pp. 299–310). https://doi.org/10.1016/B978-0-12-407794-2.00025-0.

Hunt, R. R. (2006). The concept of distinctiveness in memory research. *Distinctiveness and Memory*, 3–25.

Jacoby, L. L., Yonelinas, A. P., & Jennings, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. *Scientific Approaches to Consciousness*, 13–47.

Jamieson, R. K., Mewhort, D. J. K., & Hockley, W. E. (2016). A computational account of the production effect: Still playing twenty questions with nature. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 70*(2), 154.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage, 17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132.

Jenkinson, M., & Smith, S. M. (2001). A Global Optimisation Method for Robust Affine Registration of Brain Images. *Medical Image Analysis, 5:2*(143–156).

Johansen-Berg, H., & Matthews, P. (2002). Attention to movement modulates activity in sensori-motor areas, including primary motor cortex. *Experimental Brain Research, 142*(1), 13–24. https://doi.org/10.1007/s00221-001-0905-8.

Kassambara, A. (2021). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. *R package version*, (7).

Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: A meta-analysis of 74 fMRI studies. *NeuroImage, 54*(3), 2446–2461. https://doi.org/10.1016/j.neuroimage.2010.09.045.

Lenth, R. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. (Version 1.3.0). Retrieved from https://CRAN.R-project.org/package=emmeans.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in systems neuroscience, 2*(4).

Lin, O. Y., & MacLeod, C. M. (2012). Aging and the production effect: A test of the distinctiveness account. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 66*(3), 212.

MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica, 98*(2), 291–310. https://doi.org/10.1016/S0001-6918(97)00047-4.

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science, 26*(4), 390–395.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(3), 671–685. https://doi.org/10.1037/a0018785.

Mangels, J. A., Picton, T. W., & Craik, F. I. M. (2001). Attention and successful episodic encoding: An event-related potential study. *Cognitive Brain Research, 11*(1), 77–95. https://doi.org/10.1016/S0926-6410(00)00066-5.

Martin, A., Schurz, M., Kronbichler, M., & Richlan, F. (2015). Reading in the brain of children and adults: A meta-analysis of 40 functional magnetic resonance imaging studies. *Human Brain Mapping, 36*(5), 1963–1981. https://doi.org/10.1002/hbm.22749.

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage, 59*(3), 2636–2643. https://doi.org/10.1016/j.neuroimage.2011.08.076.

Murphy, K. A., Jogia, J., & Talcott, J. B. (2019). On the neural basis of word reading: A meta-analysis of fMRI evidence using activation likelihood estimation. *Journal of Neurolinguistics, 49*, 71–83.

Nee, D. E., Brown, J. W., Askren, M. K., Berman, M. G., Demiralp, E., Krawitz, A., & Jonides, J. (2012). A Meta-analysis of Executive Components of Working Memory. *Cerebral Cortex, 23*(2), 264–282. https://doi.org/10.1093/cercor/bhs007.

Newman, A. J., Supalla, T., Fernandez, N., Newport, E. L., & Bavelier, D. (2015). Neural systems supporting linguistic structure, linguistic experience, and symbolic communication in sign language and gesture. *Proceedings of the National Academy of Sciences, 112*(37), 11684. https://doi.org/10.1073/pnas.1510527112.

Newman, A. J., Supalla, T., Hauser, P. C., Newport, E. L., & Bavelier, D. (2010a). Prosodic and narrative processing in American Sign Language: An fMRI study. *NeuroImage, 52*(2), 669–676. https://doi.org/10.1016/j.neuroimage.2010.03.055.

Newman, A. J., Supalla, T., Hauser, P., Newport, E. L., & Bavelier, D. (2010b). Dissociating neural subsystems for grammar by contrasting word order and inflection. *Proceedings of the National Academy of Sciences of the United States of America, 107*(16), 7539–7544. https://doi.org/10.1073/pnas.1003174107.

Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General, 129*(2), 242.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4.

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Frontiers in Neuroinformatics, 10*, 27.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functionalneuroimaging studies. *Human Brain Mapping, 25*(1), 46–59. https://doi.org/10.1002/hbm.20131.

Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity. *Memory & Cognition, 40*(3), 326–338. https://doi.org/10.3758/s13421-011-0165-1.

Parsons, S. (2020). splithalf: Robust estimates of split half reliability, version 4.0. figshare. *Software*. https://doi.org/10.6084/m9.figshare.11956746.v4.

Peirce, J. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*, 10. https://doi.org/10.3389/neuro.11.010.2008.

Quinlan, C. K., & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory, 21*(8), 904–915.

R Core Team. (2016). R: A language and environment for statistical computing. (Version 3.5.1). Retrieved from https://www.R-project.org/.

Rinne, T., Pekkola, J., Degerman, A., Autti, T., Jääskeläinen, I. P., Sams, M., & Alho, K. (2005). Modulation of auditory cortex activation by sound presentation rate and attention. *Human Brain Mapping, 26*(2), 94–99. https://doi.org/10.1002/hbm.20123.

Ritchey, M., Wing, E. A., LaBar, K. S., & Cabeza, R. (2013). Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cerebral cortex, 23*(12), 2818–2828.

Rodd, J. M., Vitello, S., Woollams, A. M., & Adank, P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: An Activation Likelihood Estimation meta-analysis. *Brain and Language, 141*, 89–102. https://doi.org/10.1016/j.bandl.2014.11.012.

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., … Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage, 60*(1), 830–846. https://doi.org/10.1016/j.neuroimage.2011.11.050.

Rowe, J., Friston, K., Frackowiak, R., & Passingham, R. (2002). Attention to Action: Specific Modulation of Corticocortical Interactions in Humans. *NeuroImage, 17*(2), 988–998. https://doi.org/10.1006/nimg.2002.1156.

Schultz, H., Tibon, R., LaRocque, K. F., Gagnon, S. A., Wagner, A. D., & Staresina, B. P. (2019). Content tuning in the medial temporal lobe cortex: Voxels that perceive, retrieve. *ENeuro, 6*(5).

Sestieri, C., Shulman, G. L., & Corbetta, M. (2017). The contribution of the human posterior parietal cortex to episodic memory. *Nature Reviews Neuroscience, 18*, 183.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping, 17*(3), 143–155.

Staresina, B. P., Henson, R. N. A., Kriegeskorte, N., & Alink, A. (2012). Episodic Reinstatement in the Medial Temporal Lobe. *The Journal of Neuroscience, 32*(50), 18150. https://doi.org/10.1523/JNEUROSCI.4156-12.2012.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage, 44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061.

Szucs, D., & Ioannidis, J. P. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *Neuroimage, 221*, Article 117164.

Taylor, J. S. H., Rastle, K., & Davis, M. H. (2013). Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. *Psychological Bulletin, 139*(4), 766–791. https://doi.org/10.1037/a0030266.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. Oxford, England: Bureau of Publications, Teachers Co.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne, 26*(1), 1–12. https://doi.org/10.1037/h0080017.

Vaidya, C. J., Zhao, M., Desmond, J. E., & Gabrieli, J. D. (2002). Evidence for cortical encoding specificity in episodic memory: Memory-induced re-activation of picture processing areas. *Neuropsychologia, 40*(12), 2136–2143.

Varao Sousa, T. L., Carriere, J. S. A., & Smilek, D. (2013). The way we encounter reading material influences how frequently we mind wander. *Frontiers in Psychology, 4*, 892. https://doi.org/10.3389/fpsyg.2013.00892.

Venezia, J. H., Vaden, K. I., Rong, F., Maddox, D., Saberi, K., & Hickok, G. (2017). Auditory, Visual and Audiovisual Speech Processing Streams in Superior Temporal Sulcus. *Frontiers in Human Neuroscience, 11*, 174. https://doi.org/10.3389/fnhum.2017.00174.

Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., … Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology,

semantics, and sentence processing. *NeuroImage, 30*(4), 1414–1432. https://doi.org/10.1016/j.neuroimage.2005.11.002.

Vigneau, M., Beaucousin, V., Herve, P.-Y., Jobard, G., Petit, L., Crivello, F., … Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage, 54*(1), 577–593. https://doi.org/10.1016/j.neuroimage.2010.07.036.

Vilberg, K. L., & Rugg, M. D. (2008). Memory retrieval and the parietal cortex: A review of evidence from a dual-process perspective. *Part Special Issue: What Is the Parietal Lobe Contribution to Human Memory?, 46*(7), 1787–1799. https://doi.org/10.1016/j.neuropsychologia.2008.01.004.

Wagner, S., Sebastian, A., Lieb, K., Tuscher, O., & Tadic, A. (2014). A coordinate-based ALE functional MRI meta-analysis of brain activation during verbal fluency tasks in healthy control subjects. *BMC Neuroscience, 15*, 19. https://doi.org/10.1186/1471-2202-15-19.

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2016). The drawing effect: Evidence for reliable and robust memory benefits in free recall. *The Quarterly Journal of Experimental Psychology, 69*(9), 1752–1776. https://doi.org/10.1080/17470218.2015.1094494.

Weber, K., Lau, E. F., Stillerman, B., & Kuperberg, G. R. (2016). The yin and the yang of prediction: An fMRI study of semantic predictive processing. *PLoS ONE, 11*(3), Article e0148637.

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences, 97*(20), 11125–11129.

Wing, E. A., Ritchey, M., & Cabeza, R. (2015). Reinstatement of individual past events revealed by the similarity of distributed activation patterns during encoding and retrieval. *Journal of cognitive neuroscience, 27*(4), 679–691.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage, 92*, 381–397. https://doi.org/10.1016/j.neuroimage.2014.01.060.

Woolrich, M. W., Ripley, B. D, Brady, J. M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modelling of FMRI data. *Neuroimage, 14*(6), 1370–1386.

Yonelinas, Andrew P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language, 46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864.

Yonelinas, A. P., & Jacoby, L. L. (1995). The Relation between Remembering and Knowing as Bases for Recognition: Effects of Size Congruency. *Journal of Memory and Language, 34*(5), 622–643. https://doi.org/10.1006/jmla.1995.1028.

Zeithamova, D., de Araujo Sanchez, M. A., & Adke, A. (2017). Trial timing and pattern-information analyses of fMRI data. *Neuroimage, 153*, 221–231.