

# Learning and Consolidation of Novel Spoken Words

Matthew H. Davis<sup>1</sup>, Anna Maria Di Betta<sup>2</sup>, Mark J. E. Macdonald<sup>3</sup>,  
and M. Gareth Gaskell<sup>2</sup>

## Abstract

■ Two experiments explored the neural mechanisms underlying the learning and consolidation of novel spoken words. In Experiment 1, participants learned two sets of novel words on successive days. A subsequent recognition test revealed high levels of familiarity for both sets. However, a lexical decision task showed that only novel words learned on the previous day engaged in lexical competition with similar-sounding existing words. Additionally, only novel words learned on the previous day exhibited faster repetition latencies relative to unfamiliar controls. This overnight consolidation effect was further examined using fMRI to compare neural responses to existing and novel words learned on different days prior to scanning (Experiment 2). This revealed an elevated response for novel com-

pared with existing words in left superior temporal gyrus (STG), inferior frontal and premotor regions, and right cerebellum. Cortical activation was of equivalent magnitude for unfamiliar novel words and items learned on the day of scanning but significantly reduced for novel words learned on the previous day. In contrast, hippocampal responses were elevated for novel words that were entirely unfamiliar, and this elevated response correlated with postscanning behavioral measures of word learning. These findings are consistent with a dual-learning system account in which there is a division of labor between medial-temporal systems that are involved in initial acquisition and neocortical systems in which representations of novel spoken words are subject to overnight consolidation. ■

## INTRODUCTION

New word learning is a critical component of the human language system. By adulthood, language users have approximately 30,000 words in their mental lexicon (Altmann, 1997; Waring & Nation, 1997), with this lexical knowledge deployed automatically and efficiently during listening and speaking (Levelt, Roelofs, & Meyer, 1999; Marslen-Wilson, 1984). Retrieval of stored word knowledge is also reflected in differential neural responses evoked by familiar words and unfamiliar pseudowords (Binder et al., 2000; Ziegler, Besson, Jacobs, Nazir, & Carr, 1997; Bentin, McCarthy, & Wood, 1985). Here, we explore the learning mechanisms by which novel spoken words obtain the equivalent cognitive and neural status as preexisting, familiar words.

One cognitive marker of word knowledge is competition between lexical representations (Gaskell & Marslen-Wilson, 1997; McClelland & Elman, 1986). We not only learn to recognize new words such as *blog* but must also distinguish them from similar sounding competitors like *blag*, *bog*, and *clog*. Recent work has demonstrated a temporal dissociation between these two aspects of word learning. Participants rapidly become familiar with fictitious novel words such as *cathedruke* (measured by a recognition memory test), whereas an effect of new

learning on existing words (slowed identification of a competitor *cathedral*) is only observed after a delay (Bowers, Davis, & Hanley, 2005; Gaskell & Dumay, 2003). Furthermore, engagement in lexical competition is associated with sleep: A 12-hr delay between learning and testing produces lexical competition when participants sleep in the intervening period but not when they remain awake (Dumay & Gaskell, 2007). Such data fit well with research demonstrating sleep-related memory consolidation in other domains (see Walker, 2005); however, evidence of consolidation-induced changes in the neural representation of novel words is lacking.

In the current article, we present convergent behavioral and fMRI evidence advancing a two-stage neural model for learning the spoken form of new words: (1) initial learning is supported by medial-temporal systems that rapidly adapt as novel words become familiar; and (2) long-term cortical representations of new words are altered by slow, off-line consolidation (cf. O'Reilly & Norman, 2002; McClelland, McNaughton, & O'Reilly, 1995). Existing neuropsychological evidence suggests that the acquisition of new words depends on hippocampal structures (Gooding, Mayes, & van Eijk, 2000; Verfaellie, Croce, & Milberg, 1995; Gabrieli, Cohen, & Corkin, 1988), whereas long-term knowledge of spoken words is supported by cortical systems (Tyler, Marslen-Wilson, & Stamatakis, 2005; Bates et al., 2003; Tranel, Adolphs, Damasio, & Damasio, 2001). Neuroimaging investigations demonstrate both hippocampal

<sup>1</sup>Cognition and Brain Sciences Unit, Cambridge, UK, <sup>2</sup>University of York, UK, <sup>3</sup>University of Pennsylvania

(Breitenstein et al., 2005) and neocortical (Majerus et al., 2005; Cornelissen et al., 2004) changes during exposure to new words. However, no previous study has compared immediate and longer-term consequences of learning new spoken words.

A further critical aspect of learning a new spoken word is that knowledge acquired must be employed in a number of different tasks or contexts. For instance, we might have to produce in speech a word that we have only read previously or interpret a recently learned word in an unfamiliar sentential context. To know a word, then, implies both appropriate usage and generalization across multiple different tasks and situations. For experimental investigations of word learning, we can similarly distinguish between learning-induced changes that occur when both the stimulus and the task or context are repeated (which might reflect task-specific repetition priming; cf. Orfanidou, Marslen-Wilson, & Davis, 2006; Dobbins, Schnyer, Verfaellie, & Schacter, 2004) and changes that imply long-term, stable, and flexible representations of novel words. Although task-specific repetition effects may indeed reflect aspects of word learning, a critical test is nonetheless to assess whether newly acquired knowledge of spoken words generalizes to tasks that differ from those that were used during training. Existing neuroimaging investigations of word learning either do not include a test of generalization (Breitenstein et al., 2005; Majerus et al., 2005) or have failed to show wordlike generalization to an untrained task (Mestres-Misse, Rodriguez-Fornells, & Munte, 2007). Here we assess the hypothesis that generalized knowledge of newly learned spoken words involves overnight consolidation by using test tasks that differ from those employed during initial learning.

Both behavioral and fMRI experiments reported here were conducted over 2 days. On each day, participants learned fictitious novel words in a phoneme-monitoring task. Effects of initial learning and overnight consolidation were assessed in behavioral and fMRI test sessions on the second day, comparing the two sets of words trained on Day 1 or Day 2 with untrained control items (see timelines, Figure 1). By using two sets of items learned in separate study periods, we can test for effects of overnight consolidation within a single test session. This is an efficient design given the potential variability of functional imaging results from different sessions in the same subject (McGonigle et al., 2000; Noll et al., 1997). A further advantage of this design is that differences in word knowledge for items learned on the same or previous day cannot reflect procedural learning of test tasks because these are only administered once. Previous demonstrations of off-line consolidation of word knowledge (Dumay & Gaskell, 2007; Bowers et al., 2005; Gaskell & Dumay, 2003) used a single study period and multiple test sessions. These studies therefore confounded overnight changes in word knowledge and practice at the test task. To assess whether procedural learning

could account for previous observations of off-line consolidation of newly learned words, our first experiment replicates and extends behavioral findings on the overnight emergence of lexical competition from newly learned words (Dumay & Gaskell, 2007).

## EXPERIMENT 1—BEHAVIORAL EVIDENCE FOR LEARNING AND CONSOLIDATION OF NOVEL SPOKEN WORDS

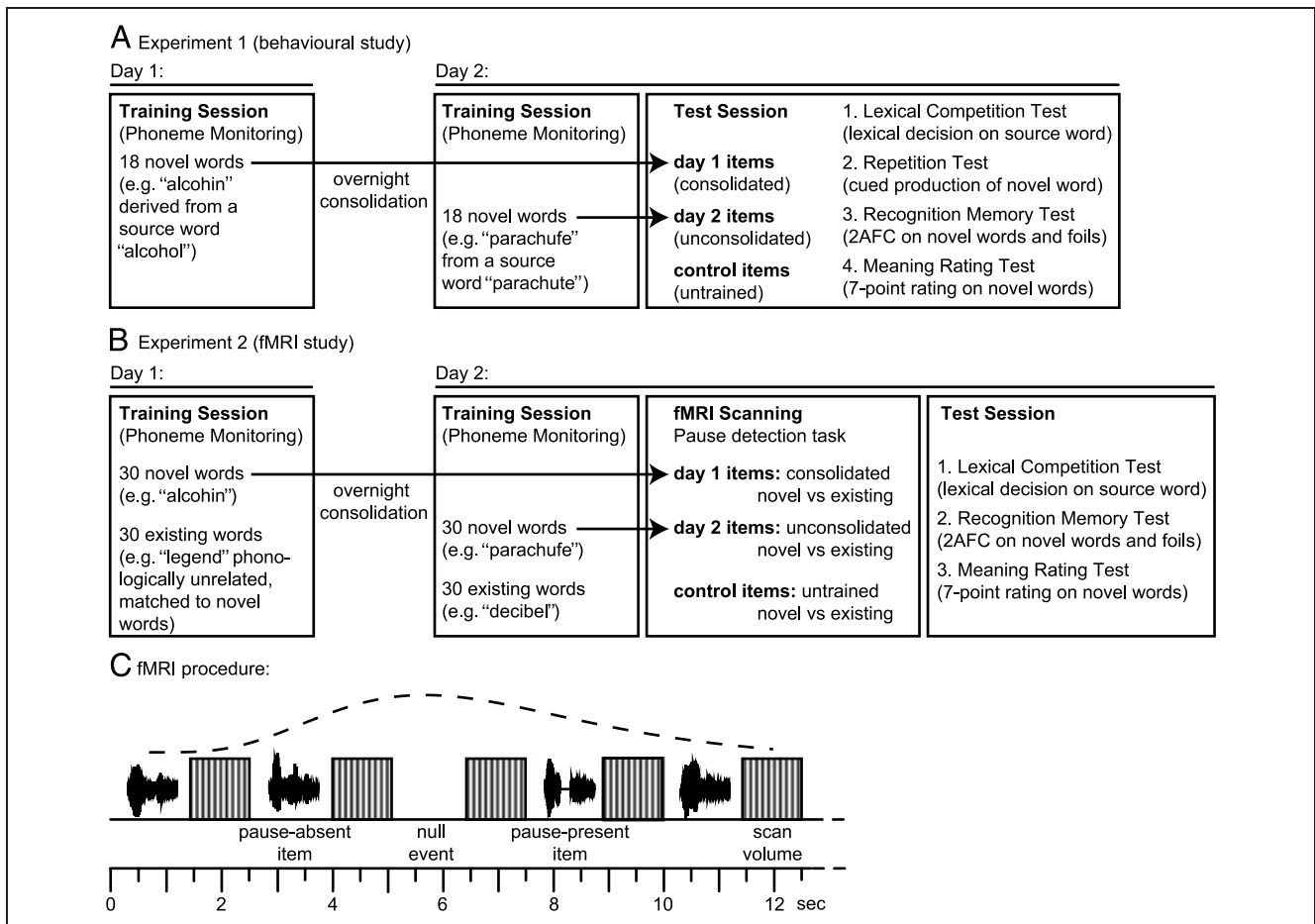
### Materials and Methods

#### Participants

Fifty-seven students from the University of York were tested under the supervision of the University of York Psychology Department Ethics Committee. All were native English speakers with no known hearing or language impairment and received either course credits or payment for their participation. Participants were randomly assigned to three different groups.

#### Materials

We used 54 stimulus triplets taken from Gaskell and Dumay (2003;  $n = 33$ ) and Tamminen and Gaskell (2008;  $n = 21$ ). Each triplet contained a familiar base word (e.g., *alcohol*), a nonword to be learned as a “novel” word (e.g., *alcobin*), and a second nonword (e.g., *alcobid*) that was used as a foil in the recognition memory task. Base words were bisyllabic ( $n = 21$ ) and trisyllabic ( $n = 33$ ) with mean length 8.2 phonemes ( $SD = 1.2$ ) and frequency 5.8 per million (range = 0–19 from CELEX; Baayen, Piepenbrock, & van Rijn, 1993). Their phonemic uniqueness point (cf. Marslen-Wilson, 1984) was 6.0 ( $SD = 1.5$ ). Nonwords diverged from the base word on the final vowel and from each other on the final consonant, ensuring that full lexicalization of the novel word would have the effect of extending the uniqueness point of the associated base word (by adding a new close competitor) and slow down the recognition of the base word (cf. Gaskell & Dumay, 2003). The 54 triplets were divided into three matched lists of 18 items with each group of participants trained on a set of 18 novel words on Day 1 and another set on Day 2. The remaining 18 novel words acted as control items during testing. Thus, we can compare responses to untrained control items with responses to novel words that have been trained and (potentially) consolidated (Day 1 items) or that remain unconsolidated (Day 2 items). The assignment of item lists to training conditions was counterbalanced across three groups of participants. All the spoken words used in this study were recorded on CD-R by a native speaker of British English in a sound-proof booth at a sampling rate of 44.1 kHz. Sound files were digitally transferred to a computer, divided into single sound files using Cool Edit software (Syntrillium Software, Phoenix, AZ), and trimmed to length.



**Figure 1.** (A) Timeline for Experiment 1. All participants were trained on two sets of novel spoken words on successive days. On the second day, training was immediately followed by a behavioral test session comprising four behavioral tasks assessing different forms of knowledge of the newly learned words. (B) Timeline for Experiment 2. After training on two sets of novel and existing spoken words on successive days, participants were fMRI scanned and completed a behavioral test session assessing different knowledge of newly learned words. (C) Timeline of the fast sparse fMRI procedure illustrating the rapid alternation of stimulus presentation and single scan volumes. Dashed line shows an estimate of the predicted BOLD response to a single stimulus (using the canonical hemodynamic response in the SPM software), illustrating how the expected hemodynamic response is sampled over subsequent scans (cf. Orfanidou et al., 2006; Jacquemot et al., 2003).

### Procedure

The experiment was carried out in two sessions on two consecutive days. On Day 1, participants undertook 216 trials of a phoneme-monitoring test with each of 18 novel words presented 12 times in total in a training session lasting approximately 15 min. Prior to each block of 18 trials, a visual display signaled a target phoneme that participants should listen for (/n/, /t/, /d/, /s/, /p/, /m/ used twice each). Participants were required to indicate with a button-press if the target phoneme was present or absent and had 4 sec to respond. On Day 2, approximately 24 hr after the first session, the same phoneme-monitoring paradigm was used with a new list of 18 novel words. After that, participants carried out in fixed order a *lexical competition test*, a *repetition test*, a two-alternative forced-choice *recognition test*, and finally a *word meaning rating test*; each of these tasks is described below. The *lexical competition test* involved

making timed lexical decisions (word/nonword) to the 54 base words (e.g., *alcohol*) testing whether responses were significantly slowed by competition from newly learned novel words (such as *alcoholin*). The 54 base words along with 206 filler items (76 words and 130 nonwords) were presented in random order with a response deadline of 3 sec and intertrial interval of 1 sec. Response latencies were recorded from the onset of the spoken stimulus. The *repetition test* examined whether the speed with which novel words are produced is affected by prior learning. All 54 novel words were presented in random order, including 18 untrained items, 18 trained on the same day, and 18 trained on the previous day. The order of presentation of the stimuli was randomized for each participant, and each novel word was followed after a variable interval of 500, 1000, or 1500 msec by a tone that cued participants to repeat the word. Response latencies were recorded from the beginning of the cue tone with deadlines and intertrial

interval as before. The forced-choice *recognition test* required participants to distinguish the 36 trained novel words (e.g., *alcobin*) from their corresponding foils (e.g., *alcobid*). A two-alternative procedure was employed with participants choosing whether the first or second item presented was a learned novel word. The *word meaning rating task* involved presentation of all 54 novel words in a random order with participants rating the extent to which each word had become associated with meaning or acquired an invented meaning on a 7-point scale. Low ratings (1–2) indicated that no meaning had been attached to the novel word, whereas high ratings (6–7) indicated that a meaning had clearly been attached to the new word. Although the instructions emphasized responding based on meaning, they also acknowledged that differences in familiarity might also arise and so this task might reflect both subjective familiarity and meaningfulness.<sup>1</sup> There was no time limit for responses in the recognition or the meaning rating tests.

All participants were tested individually in separate cubicles, with auditory stimuli delivered via high-quality headphones, and button-press and vocal responses were recorded using DMDX software (Forster & Forster, 2003).

## Results

The results of these behavioral tests were averaged over items for each participant, and effects of word learning and consolidation were assessed using repeated measures ANOVAs with SPSS. For this fully counterbalanced design, we report the results of analysis by participants only (following Raaijmakers, Schrijnemakers, & Gremmen, 1999) and include dummy variables to encode the subject-

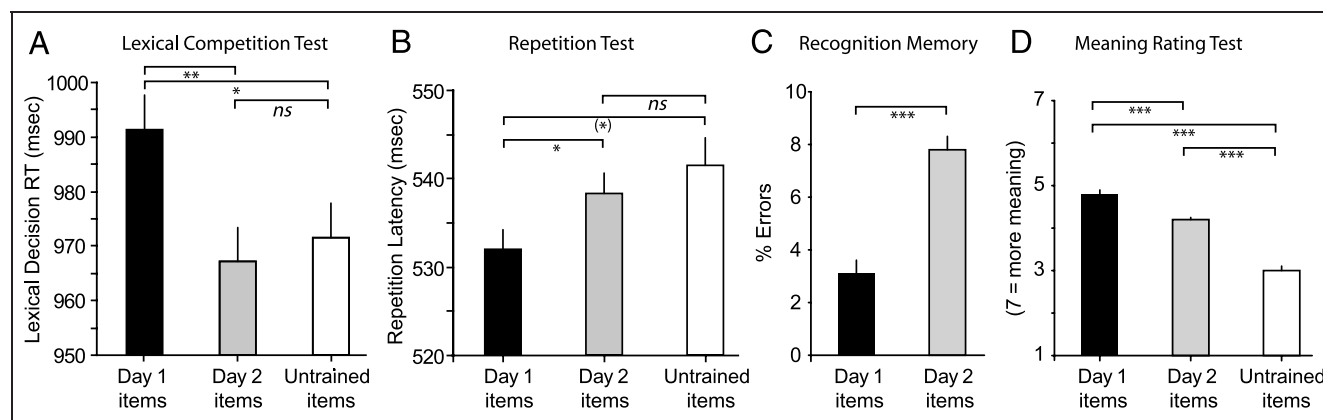
specific assignment of item groups to training conditions (Pollatsek & Well, 1995).

### Lexical Competition Test

Response times were inverse transformed to reduce the influence of response time outliers (Ulrich & Miller, 1994). Mean response times shown in Figure 2A are retransformed for ease of presentation (harmonic means). Analysis of variance revealed significant differences between the three training conditions [ $F(2,108) = 3.635, p < .05$ ]. Planned pairwise comparisons showed that, compared with controls for which no novel neighbor had been presented, there was significant slowing of responses to base words due to competition from consolidated novel items that were learned on Day 1 [ $F(1, 54) = 7.095, p < .01$ ], whereas existing neighbors of unconsolidated items learned on Day 2 showed no such slowing ( $F < 1$ ). Critically, there was a significant difference between responses to neighbors of consolidated (Day 1) and unconsolidated (Day 2) items [ $F(1, 54) = 5.287, p < .05$ ]. Thus, significant lexical competition only arises for existing word neighbors of novel words on the day after initial acquisition. There were no significant differences in error rates between the three training conditions (all  $F < 1$ ).

### Repetition Test

Response latencies when producing the spoken form of novel words in the three training conditions (shown in Figure 2B) were assessed by hand-marking speech onsets with the assistance of CheckVocal software (Protopapas, 2007). Response times showed a marginally significant difference between repetition responses in the three conditions [ $F(2, 108) = 2.571, p < .1$ ] and no significant



**Figure 2.** (A) Lexical decision response times for real-word competitors of novel words trained on Day 1 (hence consolidated), Day 2 (unconsolidated), or untrained controls. Error bars show 1 *SEM* after between-subjects variability is removed, suitable for repeated measures comparisons (cf. Loftus & Masson, 1994). Statistical significance of planned pairwise comparisons (\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ , ( $*$ )  $p < .1$ , *ns* = nonsignificant). (B) Response times for novel words trained on Day 1 (consolidated), Day 2 (unconsolidated), or untrained controls in the auditory repetition test. Error bars and statistical significance as before. (C) Recognition memory performance for novel words trained on Day 1 or Day 2. (D) Rated strength of meaning for novel words trained on Day 1, Day 2, or untrained controls.

difference in error rate ( $F < 1$ ). Similar to the lexical decision results, planned pairwise comparisons revealed a (marginal) difference in response latencies for consolidated items compared with untrained controls [ $F(1, 54) = 3.780, p < .1$ ] and no difference between unconsolidated items and untrained controls ( $F < 1$ ). Critically, though, there was again a significant difference between repetition latencies for consolidated and unconsolidated items [ $F(1, 54) = 4.279, p < .05$ ]. Production speed for novel words is, therefore, not affected by perceptual training on the same day as testing but receives further significant facilitation if a period of overnight consolidation intervenes between an initial perceptual acquisition and a speech production test.

#### *Recognition Memory Test*

Recognition memory performance (Figure 2C) was substantially above chance for both sets of trained items with fewer than 10% errors in this two-alternative forced-choice test. However, this test of explicit memory was also affected by overnight consolidation with better recognition memory for consolidated items trained on Day 1 than unconsolidated items learned on Day 2 [ $F(1, 54) = 20.59, p < .001$ ].

#### *Meaning Rating Task*

Due to a failure to follow instructions to respond to all items, data from four participants were uninterpretable. Repeated measures ANOVAs on data from the remaining 53 participants (Figure 2D) showed a significant effect of training condition on strength of meaning ratings [ $F(2, 100) = 75.79, p < .001$ ], with significant pairwise differences between all three training conditions (all comparisons  $p < .001$ ). This finding further confirms that changes to the mental representation of novel words are produced both by initial familiarization and by subsequent, overnight consolidation. Subjectively, novel words become more meaningful or familiar a day after initial presentation.

In summary, these results confirm that the competitive environment of newly acquired spoken words changes during the first 24 hr after initial training. Only those novel words for which overnight consolidation was possible significantly slowed identification of lexical competitors (cf. Dumay & Gaskell, 2007). Although explicit memory for novel words was also enhanced by overnight consolidation, the emergence of lexical competition did not reflect a failure of memory encoding for nonconsolidated items. Lexical decision responses were similarly slowed for consolidated relative to unconsolidated novel words if analysis was restricted, for each participant, to novel words that they correctly recognized in the memory test [ $F(2, 108) = 3.506, p < .05$ ]. Furthermore, these differences in response times for

consolidated items in the lexical decision tests were observed despite there being only a single testing session. This finding provides valuable confirmation that previous observations of sleep-association lexicalization (Dumay & Gaskell, 2007) are not dependent on repeated presentation of test tasks. The results also provide more direct evidence for changes in the representation of novel words following overnight consolidation. We found significant changes in both the speed of spoken productions of novel words and the strength of meaning ratings for those novel words that are subject to overnight consolidation.

## **EXPERIMENT 2—NEURAL EVIDENCE FOR LEARNING AND CONSOLIDATION OF NOVEL WORDS**

### **Materials and Methods**

#### *Design*

Experiment 2 used fMRI to measure neural responses to novel spoken words at different stages of learning and consolidation using the same 2-day training procedure (see Figure 1B). In this fMRI study, we assessed the neural representation of novel words directly during the test phase (as in the repetition and meaningfulness rating tasks) rather than by indirectly assessing responses to competitors of newly learned words (as in the lexical competition test). A further goal in assessing neural responses to novel words was to determine whether and when novel words evoke an equivalent neural response to preexisting, familiar words. To avoid confounds due to recent training in comparing novel and existing words, we included both novel words (e.g., “alcoholin”) and matched phonologically unrelated existing words (e.g., “legend”) in the training regime. In this way, changes in the neural responses to novel words due to learning and consolidation can be separated from effects of recent training on items with which participants are already familiar. Thus, all three levels of preexposure (consolidated, unconsolidated, and untrained controls) in Experiment 2 included both novel nonwords and phonologically unrelated preexisting words. To increase statistical power in assessing fMRI responses, we presented all stimulus items four times over the course of the experiment (three times intact and once with an inserted pause for a detection response). Comparisons of neural responses for untrained novel words across the three intact presentation provides a neural correlate of initial familiarization with novel words.

#### *Participants*

Sixteen right-handed, native speakers of British English aged between 18 and 40 years participated in an fMRI study approved by the Cambridge NHS Research Ethics

Committee. All participants reported normal hearing and no neurological or language impairment.

### Materials

Because of the need to obtain robust fMRI responses to novel words in a smaller number of participants, each training session included more repetitions of a larger stimulus set in each condition. Experiment 2, therefore, assessed neural responses to 90 “novel” spoken words and 90 “existing” real words, both divided into three matched groups of 30 items that varied in their time and level of preexposure prior to scanning: (1) “untrained” items were first presented to participants during the first fMRI scanning session; hence, the novel words in this condition were entirely unfamiliar at the start of scanning; (2) “unconsolidated” items were presented to participants during a training session on the same day as fMRI scanning such that novel items would be familiar to participants but not subject to overnight consolidation; (3) “consolidated” items were presented to participants during training on the day prior to scanning such that there was an opportunity for overnight, sleep-associated consolidation of this set of novel words. On the basis of the previous behavioral data and the results of Experiment 1, it is only this set of consolidated novel words that we predicted to have a neural status approaching that of existing words.

The novel words were again taken from previous behavioral studies used to demonstrate novel word lexical competition effects. There were 29 novel items (e.g., *alcobin*) derived by altering the final vowel and consonant(s) of existing base word (e.g., *alcohol*; Gaskell & Dumay, 2003). A further 61 items from Dumay and Gaskell (2007) were used in which the novel word (e.g., *fellowks*) was generated by adding one consonant (55 items) or two consonants (6 items) to the existing word (e.g., *fellow*). Items with two additional consonants had offset segments that existed in monomorphemic English words (e.g., the final syllable of “fellowks” rhymes with *hoax* or *coax*).<sup>2</sup> For each novel word, two matched items were selected: a preexisting English word and a nonword. The existing words were matched to the novel words on the number of syllables, phonemes, and consonant–vowel structure to the novel words and were matched for lemma frequency to the source word. These matched words were phonologically unrelated to the base words from which the novel items had been derived (e.g., the novel word *fellowks*, derived from *fellow*, was matched to the existing word *legend*). The matched nonwords were used as foil items in recognition memory tests and were highly similar to the novel words used in training (e.g., for *fellowks* the nonword *fellowkt* was used).

Stimuli were divided into three groups, matched for word frequency, length in phonemes/syllables, and uniqueness/divergence points (cf. Marslen-Wilson, 1984).

Each group of items was assigned to the three training conditions in a counterbalanced fashion for each volunteer. Stimuli were recorded and digitally transferred as before. To create target items for the pause-detection task used during scanning, we inserted a 200-msec silent period into each speech stimulus using Cool Edit software.

### Procedure

The two training sessions were similar to those used in Experiment 1. In each session, participants were exposed to one of the three sets of novel and existing words (for details, see Figure 1B). One training session occurred on the previous day to scanning (consolidated items: mean = 27 hr before scanning; range = 25 to 31 hr) and another session on the same day as scanning (unconsolidated items: mean = 4 hr before scanning; range = 2 to 8 hr). Participants were trained on both novel and existing words described above using the same phoneme-monitoring task as previously with six consonants (/p/, /b/, /k/, /m/, /n/, /s/), which occurred with equal frequency in the three groups of 30 novel and 30 existing words. Participants were presented with a randomly ordered sequence of items (intertrial interval 2 sec) and instructed to press a button when they detected a prespecified phoneme. Each participant monitored for each target phonemes five times, such that each stimulus was presented 30 times for a total of 1800 phoneme-monitoring trials (30 trials for 30 novel words and 30 familiar words) in a training session that lasted approximately 60 min.

During scanning, participants heard all the novel and existing words in a constrained, pseudorandom order generated using Mix software (van Casteren & Davis, 2006). In each of three 13-min scanning runs, all 180 stimulus items (90 novel and 90 existing words) were presented along with 60 trials (30 novel and 30 existing words, drawn equally from the three training conditions), in which a 200-msec silent pause was inserted into each speech stimulus. Participants performed a pause-detection task (cf. Mattys & Clark, 2002), pressing a button with their right index finger when they heard one of the 25% trials with an inserted pause. Our fMRI analysis focused on the majority pause-absent trials in which words were presented intact, and no button press was made. A further 60 null event trials (when no stimulus was presented) were included in each scanning run to provide a resting baseline against which responses to heard stimuli can be assessed (Josephs & Henson, 1999) and an initial search volume for fMRI data analysis. Auditory stimuli were presented over high-quality headphones (Resonance Technology, Commander XG system) using DMDX software (Forster & Forster, 2003). Participants were given an opportunity to practice the pause-detection task before entering the scanning.

After leaving the scanner, participants completed three behavioral tasks from Experiment 1 (lexicalization,

recognition memory, and meaning rating). Due to the lack of a suitable environment for recording naming responses, the repetition task was omitted. The lack of a quiet, sound-proof testing environment also impacted on response times in the lexicalization test, which were more variable than previously. These tests were conducted using a Sony VAIO laptop running DMDX software (Forster & Forster, 2003) interfaced with a two-button response box. All behavioral procedures were identical to Experiment 1 but used the larger set of novel words and competitors from Experiment 2.

### *fMRI Scanning and Data Analysis*

MR Imaging was performed with a 3-T Bruker Medspec MRI scanner using a head coil. Functional images were collected using 21 axial slices angled away from the eyes and covering most of the brain (slice thickness 4 mm, interslice distance 1 mm, matrix size  $64 \times 64$ , field of view  $20 \text{ cm} \times 20 \text{ cm}$ , in plane resolution of approximately  $3 \times 3 \text{ mm}$ ) with an EPI sequence (TR = 2.506 sec). To avoid interfering effects of scanner noise, we used bunched image acquisition in which a single volume (TA = 1.1 sec) is acquired, followed by a silent period (1.406 sec) during which a single stimulus item is presented (for details, see Figure 1C). Due to the slow evolution of the stimulus-evoked BOLD response, the hemodynamic response to each stimulus was sampled with an adequate temporal resolution over five subsequent scans. In each of three experimental sessions, 306 functional EPI images were acquired (~13 min scanning time per session). Six images at the start of each run were discarded before preprocessing to allow the EPI signal to reach equilibrium. High-resolution anatomical images (SPGR) and fieldmaps were also acquired for use in preprocessing and normalization.

Functional imaging data were preprocessed and analyzed using Statistical Parametric Mapping software (SPM2, Wellcome Department of Cognitive Neurology, London, UK): Images were corrected for motion by realigning them to the first image, and a magnetic field map was used to correct for geometric distortions to the EPIS (Cusack, Brett, & Osswald, 2003). The mean of the realigned, undistorted images was coregistered with the structural T1 volume, which was then spatially normal-

ized to the standard MNI template image. The same spatial transformation was then applied to the EPI volumes. Normalized images were smoothed with a 12-mm FWHM Gaussian kernel suitable for random-effects analysis (Xiong et al., 2000).

Data from each participant were entered into a general linear model for an event-related analysis (Josephs & Henson, 1999) with 13 event types in each session (the factorial crossing of pause-present and pause-absent trials, novel and existing words, untrained, unconsolidated, or consolidated at the time of scanning, plus an additional event type for rare pause-detection errors). Events were modeled using the SPM2 canonical hemodynamic response function (HRF) with temporal and dispersion derivatives (Henson, Price, Rugg, Turner, & Friston, 2002). Movement parameters estimated at the realignment stage of preprocessing were added as regressors of no interest. A high-pass filter (128 sec) was applied and AR1 correction for serial autocorrelation was made. Contrasts of parameter estimates from the least-mean square fit of single-subject models were entered into random-effects analyses (one-sample *t* tests) comparing the mean parameter estimate for the canonical response to zero over subjects. Significant results that passed voxel-wise false-discovery rate correction for multiple comparisons are reported (FDR,  $p < .05$ ; Genovese et al., 2002). To ensure consistent presentation despite changes in the FDR-corrected threshold for different contrasts, figures show results thresholded at  $p < .001$  uncorrected, with probability scales marked to indicate the FDR-corrected threshold. Graphs showing signal change in specific brain regions present the mean parameter estimate for the SPM canonical HRF in each single condition with zero reflecting the fit relative to scans following unmodeled null events (cf. Josephs & Henson, 1999).

### **Results**

Behavioral responses collected in the scanner were scored for the accuracy and the speed of pause-detection responses. Responses were fast and accurate throughout (see Table 1), with faster and more accurate pause-detection responses for preexisting, real words [accuracy:  $F(1,16) = 8.418, p < .01$ ; response time:  $F(1,16) =$

**Table 1.** Accuracy ( $d'$ ) and Response Time (msec from Word Onset) of Pause-Detection Responses as a Function of Lexicality and Training in Experiment 2

	Accuracy ( $d'$ )			Response Time (msec)		
	Consolidated	Unconsolidated	Untrained	Consolidated	Unconsolidated	Untrained
Novel words	4.02 (0.13)	4.08 (0.16)	3.66 (0.17)	968 (8.68)	945 (8.10)	952 (10.44)
Existing words	4.27 (0.11)	4.15 (0.14)	3.82 (0.15)	912 (9.09)	919 (8.79)	932 (7.33)

Bracketed values show the *SEM*, after between-subject variation is removed, suitable for repeated measures comparisons (cf. Loftus & Masson, 1994).

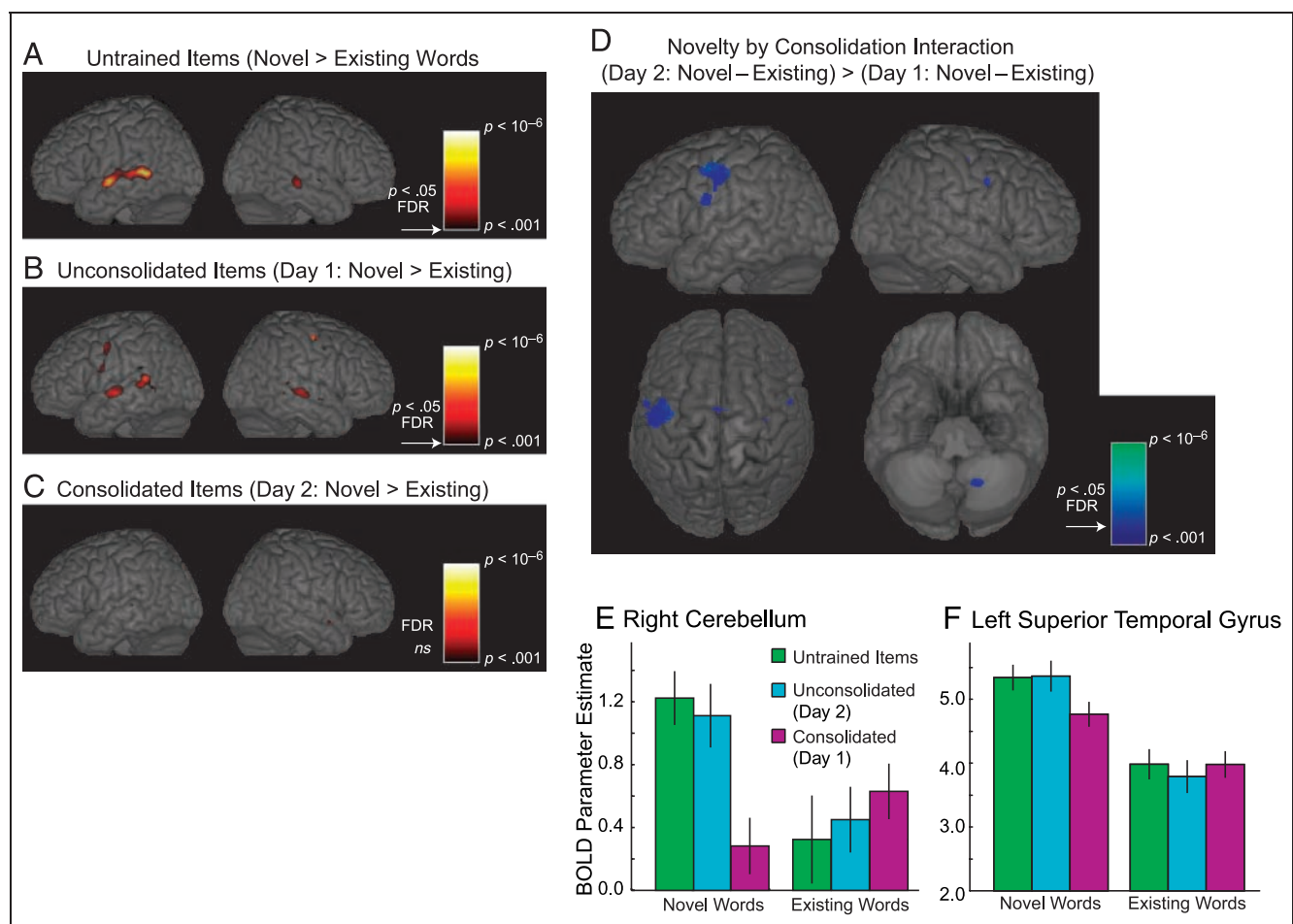
51.766,  $p < .001$ ]. Responses were also significantly more accurate for both consolidated and unconsolidated trained items [ $F(2,32) = 5.851, p < .01$ ], although changes in response times were only marginally significant [ $F(2,32) = 2.817, p = .075$ ]. There was no interaction between effects of lexicality and training condition on accuracy or response time (both  $F < 1$ ).

### Neural Effects of Consolidation

In assessing the fMRI data, we focus on pause-absent trials so that neural responses can be assessed in the absence of overt behavioral responses. We begin by contrasting responses to all pause-absent trials with the resting baseline provided by the null events. This con-

trast highlights a bilateral network of superior and middle temporal regions, motor cortex, SMA, cerebellum, and hippocampus (HC) as well as a more lateralized response in the left inferior frontal gyrus, insula, and premotor cortex. These regions provide an initial search volume for responses to spoken words as a function of preexperimental familiarity (novel/existing words) and training (untrained/unconsolidated/consolidated items).

We then examined brain areas that showed a differential response to novel and existing words (see Figures 3A–3C; Tables 2A–2C) in each training condition individually. In none of the three training conditions was additional activity observed for familiar compared with novel words. Furthermore, no additional response to existing compared with novel words was observed when



**Figure 3.** Brain regions showing response differences between novel and existing spoken words (pause-absent trials) in the fMRI experiment rendered onto the MNI canonical brain. (A) Untrained novel > existing words, thresholded at  $p < .001$  uncorrected, all activations shown exceed FDR corrected  $p < .05$ . No voxels show an elevated response to existing words. (B) Unconsolidated novel > existing (comparison of items trained on Day 2) thresholded as before. (C) Consolidated novel > existing words (comparison of items trained on Day 1), thresholded at  $p < .001$  uncorrected, no voxels approach FDR  $p < .05$ . (D) Brain regions showing an interaction between novelty and consolidation (i.e., an additional response for novel words trained on Day 2, not seen for novel words trained on Day 1), results displayed at  $p < .001$ , all peak voxels in the left hemisphere and right cerebellum exceed FDR corrected  $p < .05$ , as indicated by the arrow on the color scale. (E) Mean BOLD parameter estimate (peak of the fitted canonical HRF function on an arbitrary scale) of the peak voxel in the right cerebellum ( $x = 16, y = +60, z = -36$ ) for novel and existing words in each of the training conditions. Error bars show SEM after between-subject variance has been removed, suitable for repeated measures comparisons between conditions (Loftus & Masson, 1994). (F) Response profile of a cluster of voxels in the left STG (center of mass,  $x = -54, y = -30, z = +6$ ), which show an elevated response to novel words and a significant interaction between novelty and consolidation. Error bars as in panel E.



**Table 2.** MNI Coordinates for Peak Voxels (Maximum of Three Peaks/Cluster, Separated by >8 mm) Showing Increased Activity for Novel versus Existing Words in the Three Training Conditions

<i>Location</i>	<i>Cluster Size (Voxels)</i>	<i>Z</i>	<i>x</i>	<i>y</i>	<i>z</i>
<i>(A) Untrained Words: Novel &gt; Existing (Results Thresholded at p &lt; .001 Uncorrected, All Voxels Exceed FDR p &lt; .05)</i>					
Left posterior MTG	785	4.64	-48	-44	4
Left anterior STG		4.57	-60	-8	-4
Left posterior MTG		3.94	-62	-30	4
Right posterior MTG	107	3.71	60	-28	-6
Left cerebellum (Lobe 6)	61	3.56	-22	-62	-26
Right anterior STG	8	3.15	58	-6	-4
Right inferior frontal gyrus	1	3.13	50	20	-10
<i>(B) Unconsolidated Words: Novel &gt; Existing (Results Thresholded at p &lt; .001 Uncorrected, All Voxels Exceed FDR p &lt; .05)</i>					
Right motor cortex	34	4.21	34	-12	56
Left SMA	224	4.18	-12	0	44
SMA		3.42	0	-4	54
Left posterior STG	250	4.11	-52	-46	14
Left posterior MTG		3.83	-64	-40	8
Left posterior MTG		3.26	-50	-56	0
Left anterior MTG	167	3.93	-62	-12	-2
Right posterior MTG	148	3.72	56	-26	0
Left putamen	48	3.72	-32	-20	-4
Left anterior insula	51	3.60	-20	36	8
Left anterior insula		3.28	-22	28	4
Right cerebellum (Lobe 6)	25	3.59	20	-58	-22
Left dorsal PCG	63	3.48	-48	-4	46
Right dorsal PCG	5	3.38	30	-20	52
Right ventral PCG	8	3.36	36	-20	34
Left ventral PCG	54	3.35	-58	0	24
Left PCG		3.16	-44	-2	26
Right posterior MTG	5	3.30	66	-40	10
Right posterior MTG	6	3.30	50	-42	-2
Right ventral PCG	6	3.26	52	-2	48
Right anterior STG	6	3.21	58	-4	-8
Left SMG	8	3.18	-44	-40	26
Left rolandic operculum	2	3.14	-46	-28	20

**Table 2.** (continued)

<i>Location</i>	<i>Cluster Size (Voxels)</i>	<i>Z</i>	<i>x</i>	<i>y</i>	<i>z</i>
<i>(C) Consolidated Words: Novel &gt; Existing (Results Thresholded at p &lt; .001 Uncorrected, No Voxels Exceed FDR p &lt; .05)</i>					
Right anterior STG	5	3.44	58	6	-14
Left posterior MTG	4	3.20	-58	-32	4

MTG = middle temporal gyrus; PCG = precentral gyrus; SMA = supplementary motor area; SMG = supramarginal gyrus; STG = superior temporal gyrus.

data from all three training conditions were combined to increase statistical power (no voxels reach  $p > .001$  uncorrected or FDR corrected). As discussed later, the absence of an elevated response to familiar words might reflect our use of an acoustic-phonetic task (pause detection) that emphasizes phonological processing. Phonological processes are more difficult for less familiar words. Consequently, we consistently observed an elevated response to novel compared with existing words in this study.

For unconsolidated and untrained items, we observed an elevated response to novel compared with existing words in a bilateral region of the STG, extending into posterior middle temporal gyrus on the left (see Figures 3A and 3B; Tables 2A and 2B). In addition to these lateral temporal responses, there was an elevated response to novel words in bilateral motor cortex, SMA, and cerebellum, although these effects were more robust for items trained on the same day as the scanning (unconsolidated items) than for the items that had not been presented prior to scanning (untrained). Whereas both unconsolidated and untrained items showed a similar lexicality effect (confirmed by statistical comparisons reported later), the same comparison of novel and existing words for items on which participants had been trained on the day prior to scanning did not reach a corrected level of significance (see Figure 3C; Table 2C).

The lack of a difference between the neural responses to existing and novel words trained on the day prior to scanning (hence that had been consolidated) is striking because it suggests that these novel words have acquired more wordlike representations. However, to be confident that this null effect is meaningful, a significant difference between lexicality effects for items trained on the same day and the previous day to scanning is required. That is, we must show a significant interaction reflecting a reduction in the lexicality effect due to consolidation: (unconsolidated novel–unconsolidated existing) > (consolidated novel–consolidated existing). As shown in Figure 3D and Table 3, this contrast picks out a number of regions in which the elevated response to novel words over existing words was significantly

**Table 3.** Peak Voxels (Maximum of Three Peaks/Cluster, Separated by >8 mm) Showing a Significant Interaction between Novelty (Novel versus Existing Words) and Consolidation (Unconsolidated versus Consolidated)

Location	Cluster Size (Voxels)	Z	x	y	z
Left precentral gyrus	319	5.48	-44	-4	52
Left postcentral gyrus		3.71	-52	-18	46
Left precentral gyrus		3.46	-56	-8	42
Left precentral gyrus	62	4.07	-56	0	26
Right precentral gyrus	16	3.71	54	0	38
Right cerebellum (Lobe 8)	25	3.56	16	-60	-36
Supplementary motor area	81	3.52	-4	-6	56
Right precentral gyrus*	3	3.30	36	-14	56
Right postcentral gyrus*	2	3.18	44	-20	38

Results thresholded at  $p < .001$ . All peak voxels reach  $p < .05$  FDR corrected, except those marked with an asterisk (\*).

reduced for consolidated items: in bilateral motor and somatosensory regions on the precentral and postcentral gyri, the left premotor cortex, the SMA, and the right cerebellum. The response profile picked out by this interaction reflects a significant reduction in the magnitude of the novelty effect specific to items that were trained on the day prior to scanning (see, e.g., the peak voxel in the right cerebellum shown in Figure 3E). A second interaction contrast tested for any differences between lexical responses to untrained and unconsolidated items. This comparison (unconsolidated novel–unconsolidated existing) > (untrained novel–untrained existing) only revealed a single, small cluster in the left anterior insula, which failed to reach a corrected level of significance (30 voxels at  $p < .001$  uncorrected, peak voxel,  $x = -20$ ,  $y = +32$ ,  $z = +4$ ,  $Z = 3.43$ ,  $p = .611$  FDR). Thus, in contrast to the significant differences following overnight consolidation, repeated exposure to novel words earlier in the same day has little immediate effect on their neocortical representation.

We also conducted an ROI analysis using MarsBar (Brett, Anton, Valabregue, & Poline, 2002) to assess the response of the superior regions of the left temporal lobe, an area that shows a robust additional response to novel compared with existing words both in the current study and in the previous work (Kotz, Cappa, von Cramon, & Friederici, 2002; Newman & Twieg, 2001). This ROI was defined using the cluster of voxels that showed a main effect of lexicality (novel > existing words, averaged over all three training conditions), generating a large cluster centered on the posterior superior and middle temporal gyri (center of mass,  $x = -54$ ,  $y = -30$ ,  $z = +6$ ). In assessing the response of this STG region, we computed a positive  $t$  contrast, assessing

the same interaction between lexicality and consolidation shown in the whole-brain analysis [i.e., (unconsolidated novel–unconsolidated existing) > (consolidated novel–consolidated existing)]. This showed that the average response of the STG to novel words is significantly reduced for consolidated, relative to unconsolidated items [ $t(15) = 1.78$ ,  $p < .05$ ; shown in Figure 3F].

### Initial Learning of Novel Spoken Words

To assess neural responses to initial presentations of novel spoken words, we now focus on the untrained items. The novel words in this condition were entirely unfamiliar before the first functional imaging scan but were presented four times over the three scanning runs (one pause-present and three pause-absent presentations). By the time of the postscan recognition memory test, participants were able to distinguish these novel words from highly similar foils with above chance accuracy [mean 2AFC recognition = 75%;  $t(15) = 11.21$ ,  $p < .001$ ]. Nonetheless, these items differ from the other novel items, which were presented 30 times prior to and four times during the scanning session. Mean 2AFC recognition performance was 79% for unconsolidated and 94% for consolidated novel items trained on the day before scanning. Recognition memory performance differed between the three training conditions [ $F(2,30) = 37.67$ ,  $p < .001$ ] with planned pairwise comparisons confirming that (as in Experiment 1) recognition memory was significantly enhanced by overnight consolidation [difference between consolidated and unconsolidated items;  $t(15) = 6.70$ ,  $p < .001$ ]. Interestingly, there was only a marginal enhancement of recognition memory by training on the same day as scanning when compared with untrained items that were only presented during scanning [ $t(15) = 1.71$ ,  $p < .1$  one-tailed].

To assess neural correlates of the initial learning of novel words, we compared untrained novel items with the two sets of novel items that had been trained prior to scanning. Of particular interest are neural responses during the first scanning run since only at this time are these untrained items truly novel. Although some voxels showed an additional response to untrained novel words compared with consolidated and unconsolidated novel words in this first scanning run [including a cluster of 17 voxels at  $p < .001$  uncorrected in the left HC and anterior fusiform gyrus, peak voxel:  $x = -34$ ,  $y = -10$ ,  $z = -22$ ;  $Z = 3.85$ ,  $p = .112$  FDR], this comparison failed to reach whole-brain corrected significance. Given previous imaging studies suggesting a role for the HC in learning new words (Breitenstein et al., 2005) and theoretical accounts that emphasize a role for medial-temporal regions in rapid learning (McClelland et al., 1995), we used a hippocampal ROI to further investigate the initial acquisition of novel spoken words. This ROI was generated by assessing neural activity for pause-absent trials compared with rest at  $p < .05$  FDR within

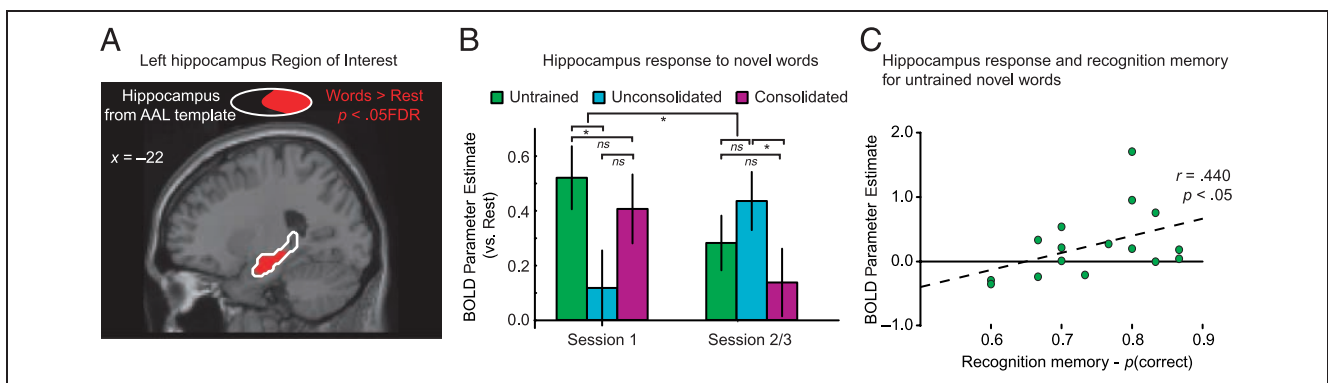
the left HC region of the automated anatomical labeling (AAL) map of the MNI single subject brain (Tzourio-Mazoyer et al., 2002). This procedure provides an ROI of 675 voxels, with center of mass at  $x = -22, y = -18, z = -12$  (shown in Figure 4A), overlapping with an anterior HC activation region that previously showed a reduction in HC activity over five presentations of consistent non-word-picture pairings (Breitenstein et al., 2005; Saykin et al., 1999).

To assess the response of the HC, we used MarsBar (Brett et al., 2002) to extract the mean response of the HC ROI to novel words in each training condition (untrained, unconsolidated, and consolidated) for the initial and subsequent scanning runs (Figure 4B). Statistical analysis of HC responses confirmed significant activity for untrained novel words compared with recently trained novel words in the first scanning run [ $t(15) = 1.938, p < .05$ ]. Paired comparisons suggested that recent training significantly reduces the hippocampal response to novel words; untrained novel items produced an elevated response compared with unconsolidated trained items [ $t(15) = 2.76, p < .05$ ], whereas the equivalent comparison with consolidated trained items was nonsignificant [ $t(15) = 0.662, p > .1$ ].<sup>3</sup> Such a result replicates other fMRI findings (Bosshardt et al., 2005) in demonstrating that hippocampal responses following item familiarization change as a function of time since encoding. That the elevated response to untrained novel words is reduced by training is confirmed by equivalent statistical tests for subsequent scanning sessions. The difference between untrained and unconsolidated novel words was nonsignificant in subsequent scanning sessions [ $t(15) = 1.03, p > .1$ ], and there was a significant reduction in this difference between untrained and trained novel words between the first and the subsequent sessions [ $t(15) = 2.54, p < .05$ ]. Such results suggest that the HC contributes primarily to the initial acquisition of novel spoken words.

Further evidence that the elevated HC response to novel words contributes to initial acquisition of novel words comes from correlational analyses relating HC activity for untrained novel words to individual's recognition memory performance for those same novel words after scanning. The magnitude of the HC response in Session 1 for untrained compared with trained novel words is positively correlated with recognition memory performance [ $r(16) = 0.440, p < .05$ ; Figure 4C]. Furthermore, the degree of reduction in the HC response between first and subsequent scanning runs shows a marginally significant correlation with recognition memory performance [ $r(16) = 0.408, p = .058$ ] such that those subjects with better subsequent recognition memory produced larger reductions in HC activity. These results support previous findings that link changes in hippocampal activity to successful acquisition of novel spoken words (Breitenstein et al., 2005).

### Postscanning Behavioral Results

Lexical decision data were lost for one participant due to a software error. Data from the remaining 15 participants failed to reveal an effect of consolidation on responses to real-word neighbors of newly learned words ( $F < 1$ ). This may reflect a priming effect from recent presentations of novel words or more likely (and less interesting) reduced power due to their being fewer participants and a nonoptimal testing environment for this response time experiment. However, despite these problems with this on-line test, the meaning/familiarity rating task showed the same effects of training and consolidation observed previously [ $F(2,30) = 27.60, p < .001$ ] with significant pairwise differences (all  $p < .01$ ) between ratings for previously untrained items (mean = 3.19 of 7), unconsolidated items (3.63), and consolidated items (4.34). Hence, the same participants that show differential fMRI responses as a function of



**Figure 4.** (A) Single slice of the MNI canonical brain showing the ROI within the left HC, defined as the portion of the AAL left HC map (white outline), which shows an additional response to pause-absent trials compared with rest thresholded at FDR  $p < .05$  (red). (B) Response of this left HC ROI to novel words presented in the first scanning run (i.e., the first presentations for untrained items) and in later scanning runs (mean of second/third run). Error bars as in Figure 3E. Significance of statistical comparisons shown with braces as in Figure 2A. (C) Correlation between responses to untrained novel words during Session 1 in the left HC ROI and subsequent recognition memory performance for single participants. The dotted line shows the best fitting linear regression line ( $y = 2.66x - 1.73$ ).

learning and consolidation of novel words also show changes in behavioral response to newly learned words equivalent to those seen in Experiment 1.

## DISCUSSION

We have demonstrated changes in the cognitive and the neural representation of previously unfamiliar spoken words as a consequence of learning and overnight consolidation. Previous behavioral demonstrations of word consolidation used multiple test sessions (Dumay & Gaskell, 2007; Bowers et al., 2005; Gaskell & Dumay, 2003) hence confounded changes in levels of word knowledge and test practice. In contrast, both our experiments employed a single test session following 2 days of training with different items. The emergence of lexical competition between consolidated novel words (*cathedruke*) and familiar words (*cathedral*) in Experiment 1 must, therefore, reflect changes in the underlying representations of novel words rather than practice at the test task. Although the current research was not intended to test directly whether consolidation is associated with sleep, previous research in which participants either remained awake or slept for equivalent periods of time showed that the mere passage of time is insufficient to generate lexical competition (Dumay & Gaskell, 2007). It seems likely, therefore, that the emergence of lexical competition ordinarily requires overnight, sleep-associated consolidation of newly learned words.<sup>4</sup>

Our study also showed other behavioral and neuroimaging differences between words learned on the same day as testing and those learned the day before. For these findings, a similar comparison between groups that sleep or remain awake following training has not previously been conducted. We, therefore, cannot rule out the possibility that certain of our findings might be due to interference between novel words learned first and second or due to the mere passage of time in the absence of sleep. However, existing data on proactive and retroactive interference would predict relatively little between-list interference given that the two sets of pseudowords taught to participants were entirely dissimilar to each other (cf. Bower, Thompson-Schill, & Tulving, 1994). Recent evidence would further suggest that the degree of between-list interference will be further reduced by postlearning sleep (Drosopoulos, Schultze, Fischer, & Born, 2007). Although future studies that include a between-groups assessment of words learned and tested on the same day or subsequent days would be helpful, we consider it most likely that many of the response differences between novel words learned on the same day and previous day to testing similarly reflect overnight changes in the cognitive and neural representation of newly learned words. In support of this conclusion, we note that comparisons between untrained and unconsolidated conditions were nonsignificant in both the lexical competition and the repeti-

tion test of Experiment 1. In Experiment 2, a relatively long delay of approximately 4 hr intervened between Day 2 training and scanning, yet we observed no effect of this training on cortical responses to novel words. This might suggest that our results do not arise from the same form of nonsleep associated memory stabilization that has been suggested for motor skill learning (Walker, Brakefield, Hobson, & Stickgold, 2003).

One interesting result consistent with consolidation-induced changes comes from the speeded-repetition task. Production of novel words was unaltered by same-day training but showed reliable facilitation following overnight consolidation. Strength of meaning ratings also showed reliable increases for novel words learned on the previous day. These findings have interesting parallels in the learning of other cognitive skills. Participants trained on visual target sequences containing hidden probabilistic rules developed implicit, procedural knowledge (shown by RT reductions; Fischer, Drosopoulos, Tsen, & Born, 2006). However, they only showed explicit sequence knowledge after sleep. We demonstrate the reverse dissociation, with participants showing explicit knowledge of newly learned words (good recognition memory) but failing to show implicit procedural effects (in word recognition and production tasks) until the following day. These results suggest that one facet of sleep's role in memory consolidation is to provide for bidirectional information flow between explicit and implicit learning systems.

Evidence of overnight consolidation of newly learned words is also provided by differential neural responses to novel words familiarized prior to or following sleep. Our fMRI comparison between real and novel words assessed whether familiarization produces pseudoword representations comparable with those of words. Previous fMRI studies comparing real and novel words (without training) showed elevated responses for words in brain regions supporting access to meaning including the supramarginal gyrus (Orfanidou et al., 2006; Xiao et al., 2005; Kotz et al., 2002; Binder et al., 2000), inferior temporal, and fusiform gyri (Orfanidou et al., 2006; Majerus et al., 2002; Binder et al., 2000). In contrast, pseudowords increase phonological processing demands and evoke elevated activity compared with words in phonologically associated brain regions, such as the superior and the middle temporal gyri (Majerus et al., 2005; Xiao et al., 2005; Kotz et al., 2002; Newman & Twieg, 2001).

Consistent with this previous literature and our use of an acoustic or a phonological task (pause detection), we observed a stronger bilateral STG response for pseudowords than matched words (for items that were not consolidated at the time of scanning). An additional response to pseudowords was also observed in the inferior frontal, premotor, and motor cortices and in the cerebellum, an elevated response shown most clearly for learned but not consolidated items. Recent studies have

shown that left frontal and premotor regions are engaged in phonological processing of spoken language if motoric processes are not concurrently engaged for overt decisions (Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Pulvermuller et al., 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004; Watkins, Strafella, & Paus, 2003; Fadiga, Craighero, Buccino, & Rizzolatti, 2002). Functional and anatomical connectivities between posterior temporal and prefrontal/premotor regions (Wilson & Iacoboni, 2006; Buchsbaum, Olsen, Koch, & Berman, 2005; Catani, Jones, & ffytche, 2005) support the recruitment of this distributed network for sensorimotor integration of spoken language (for a discussion, see Davis & Johnsrude, 2007; Scott, 2005; Hickok & Poeppel, 2004). In the present study, we thus see additional recruitment of prefrontal and premotor regions for pseudowords.

Strikingly, these elevated neural responses were not reduced by 30 prior presentations of pseudowords on the same day as scanning. Extensive prior familiarization does not lead to wordlike neural responses in brain regions involved in phonological processing and speech-motor integration. However, equivalent familiarization followed by nocturnal sleep does result in a more wordlike neural response. Thus, behavioral and fMRI data converge in suggesting that overnight consolidation creates more efficient representations of newly learned words in cortical systems involved in speech perception and production.

### **A Two-stage Account of Word Learning and Consolidation**

Our findings suggest a two-stage account of word learning, with rapid initial familiarization followed by slower, overnight consolidation, consistent with dual memory systems theories (McClelland et al., 1995). One motivation for these dual-process theories is the “catastrophic interference” suffered by distributed connectionist networks if new information is learned too quickly (McClosky & Cohen, 1989). A proposed solution assumes functional and anatomical separation between a hippocampal system that specializes in initial storage of new memories and neocortical networks that use slower, interleaved learning to combine new and existing knowledge without interference (O’Reilly & Norman, 2002; French, 1999; McClelland et al., 1995). Evidence consistent with this theory suggests that sleep may play an important role in the consolidation of hippocampal learning (e.g., Skaggs & McNaughton, 1996; Wilson & McNaughton, 1994). For instance, single-cell recordings provide evidence that neural activity in the HC, and visual cortex during awake exploration is “replayed” during slow-wave sleep (Ji & Wilson, 2007). The present study supports an extension of this dual-process theory to the acquisition of novel spoken words.

Our data showing the immediate consequences of novel word acquisition lend further support to this

dual-process account of word learning. The recognition memory data showed that participants are sufficiently familiar with recently presented novel words to distinguish them from similar-sounding foils. The neuroimaging data demonstrated stronger activity in the HC for untrained novel words on their initial presentation, with lesser activation on second and third presentation, and a particularly subdued response after extensive prior training on the same day. Further evidence to link hippocampal responses to initial acquisition comes from correlational analyses, which show that those participants that produce a larger HC response to untrained words and a greater reduction in HC responses over the three scanning sessions have better subsequent recognition memory for novel words. We, therefore, propose a specific role for the HC in the initial acquisition of novel spoken words such that learners can distinguish newly learned words from similar-sounding foils after only three intact presentations. This conclusion is consistent with similar results for the learning of written word to picture pairings (Breitenstein et al., 2005). Hippocampal involvement in word learning is also supported by evidence that medial-temporal lobe lesions substantially impair vocabulary acquisition (Gooding et al., 2000; Verfaellie et al., 1995; Gabrieli et al., 1988).

The overnight changes that we have seen in cortical responses to novel words may appear at odds with the fact that retrograde amnesia can impair recall of memories formed years or even decades prior to hippocampal damage (Squire, 1992). To be clear, we are not proposing that all representations of novel words are transferred from HC to neocortex overnight. However, we do suggest that a substantial (and perhaps the largest) change in hippocampal involvement occurs on the first night after learning. Thus, behavioral and fMRI studies can demonstrate observable effects of just a single night of consolidation. This profile of hippocampal involvement fits with a previous neuroimaging study of declarative memory for photographs over the course of 90 days following learning (Takashima et al., 2006). By far, the largest changes to hippocampal and ventral medial prefrontal activation on subsequent recognition occurred within 24 hr of exposure, with more modest changes occurring in the following 90 days. Furthermore, the time course of hippocampal dependence on new learning may relate to both the task used to assess learning and the degree of prior experience of a particular domain of knowledge. For instance, studies of place–food learning in rats have shown that the duration of hippocampal dependence of newly acquired associations depends on the degree of prior experience and schematization of place–food associations (Tse et al., 2007). For animals with the appropriate schema, newly learned associations became independent of the HC in less than 48 hr—consistent with the time span of sleep-associated consolidation proposed in this and previous studies of word learning (e.g., Dumay & Gaskell, 2007).

By analogy to lexical learning, then, we might expect that participants could more rapidly learn and consolidate new word forms that fit the phonotactic structure of their native language (as in the present study because all our novel words contained legal phonological sequences) compared with “foreign” words that include nonnative phonological sequences (cf. Warker & Dell, 2006).

Overnight transfer of representations from HC to neocortex may involve some process of elimination of existing memories to make room for new memories the next day (cf. Rosenzweig, Barnes, & McNaughton, 2002). Consistent with this idea, there was some suggestion that the subdued hippocampal response to novel words that had been presented repeatedly during the day then recovered on the first presentation following a night’s sleep (see Figure 4B). However, the critical comparison of responses to unconsolidated and consolidated novel items failed to reach statistical significance. Further investigations of responses to recently and not-so-recently learned items would be valuable to address this question.

Although our data are in line with a dual-systems account of novel word learning, uncertainty remains as to the degree of interplay between neocortical and hippocampal systems in the consolidation process for novel words. Despite some striking examples of post-morbid vocabulary acquisition (e.g., Van der Linden et al., 2001), the neuropsychological literature on retrograde amnesia suggests that lesions to hippocampal and parahippocampal structures severely impair vocabulary learning (e.g., Verfaellie, Koseff, & Alexander, 2000). However, most case studies have focused on retrieval or recognition of meanings or familiarity of forms. Given that spoken word recognition is an overlearned, automatized skill, it is reasonable to think of the engagement of a novel word in lexical competition as procedural knowledge. Hence, the aspects of novel word learning that lead to a delay in the recognition of existing words, as in Experiment 1, may still be evident in cases where normal hippocampal functioning is lost. Support for this point of view comes from research showing amnesics’ spared learning of “common ground” in communication, which was viewed as a procedural tuning of existing systems (Duff, Hengst, Tranel, & Cohen, 2006). Nonetheless, we would predict that in the absence of support from medial-temporal lobe learning systems, it would take many more exposures of a novel word for an amnesic to show lexical knowledge.

One potential challenge to this dual-process, hippocampal learning followed by cortical consolidation account of word learning comes from previous functional imaging studies reporting more immediate changes in cortical responses during repeated presentations of non-words (Mestres-Misse et al., 2007; Breitenstein et al., 2005; Majerus et al., 2005; Cornelissen et al., 2004). However, response reductions can also result from task-

specific neural repetition priming (a form of procedural learning documented for pictures and for spoken words and pseudowords; Orfanidou et al., 2006; Dobbins et al., 2004), which could, in principle, be unrelated to word learning. For instance, response reductions for newly learned words in the fusiform gyrus (Breitenstein et al., 2005) are inconsistent with more wordlike representations because fMRI and PET studies show that this region most often shows a heightened response to real words (Orfanidou et al., 2006; Majerus et al., 2002, although other factors such as the semantic category of words and the task used can also alter responses; for a review, see Binder et al., 2000). Therefore, apparent learning-related reductions in cases where both stimulus and task are repeated might equally be ascribed to the effect of task-specific repetition priming on neural responses. This ambiguity can be resolved by using different tasks during training and testing (cf. Dobbins et al., 2004). The results of our imaging study satisfy two stringent criteria for demonstrating neural correlates of word learning: (1) generalization to a novel test task should occur and (2) responses to novel words should resemble responses to existing, familiar words. One previous study that included a test of these two criteria immediately after learning failed to show ERP responses to novel words that resemble real words in a generalization test (Mestres-Misse et al., 2007). Although the consequences of swift learning and repetition priming are difficult to tease apart, our work would suggest that one means of satisfying both the above criteria involves learning followed by overnight consolidation.

### **Effects of Consolidation on Newly learned Words**

What benefit does consolidation confer on neural representations of newly learned words? The emergence of lexical competition in this and previous work (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003) shows that consolidation strengthens neural representation such that newly learned words can be recognized more rapidly, hence compete with existing words. We propose that the emergence of lexical competition effects from novel words (Experiment 1) reflects an increase in the speed with which novel words can be identified because it is only after consolidation that novel words are recognized quickly enough to compete for identification with existing words. The results of our fMRI study (Experiment 2) and a previous study reporting a neural correlate of lexical competition for existing words in superior temporal regions (Okada & Hickok, 2006) would, thus, be consistent with the hypothesis that phonological representations in superior temporal regions are modulated by consolidation. In our study, consolidation was such that superior temporal responses to newly learned words more closely resembled responses to existing, familiar words. This might reflect sharper, better-tuned neural representations following consolidation (cf. Grill-

Spector, Henson, & Martin, 2006). However, although mean activity levels are reduced in cortical regions involved in phonological processing, the emergence of behavioral competition following consolidation might suggest that representations of newly learned words have more not less overlap with representations of existing words. Future work to test for direct correlations between consolidation induced neural changes, and the behavioral emergence of lexical competition would, therefore, be valuable. Of particular interest would be studies that assess whether the emergence of lexical competition for spoken words alter inferior frontal responses (as observed during written word recognition; Fiebach, Ricker, Friederici, & Jacobs, 2007) or whether competition effects are confined to temporal lobe regions (cf. Okada & Hickok, 2006).

Further evidence for the benefits conferred by overnight consolidation comes from the speeded repetition task: Consolidation reduces the time taken for listeners to repeat a heard nonword. This increase in the efficiency of perceptuomotor transformations of speech is likely associated with the consolidation-induced reductions in neural activity in cerebellar, precentral, and supplementary motor regions observed in Experiment 2. This conclusion is supported by existing evidence concerning a role for these regions in verbal rehearsal (Chen & Desmond, 2005; Smith, Jonides, Marshuetz, & Koeppel, 1998) and speech production (Alario, Chainay, Lehericy, & Cohen, 2006; Bohland & Guenther, 2006). However, at present we have only indirect evidence for links between these response time effects in behavior and functional imaging results.

Effects of consolidation on cortical processes of perceptuomotor transformation highlight the fact that word knowledge involves multiple sensory modalities and motor systems (e.g., we use different perceptual and motor systems in recognizing and producing spoken and written words). Because both our experiments used a purely auditory training procedure, links between perceptual and motor systems are required for newly learned words to be efficiently produced. Our behavioral and neuroimaging findings suggest that this transfer of knowledge requires overnight consolidation. We propose that on-line learning processes without consolidation may be limited to altering neural responses in sensory systems and tasks that were directly engaged during training (as in repetition priming studies). However, to transfer newly acquired perceptual knowledge to other sensory modalities and tasks (for instance, efficiently producing words that have only been perceived previously) requires overnight consolidation. Whether this cross-modal transfer process also reflects a division of labor between hippocampal and neocortical learning is unclear. Neuropsychological evidence for deficits in supraspan phonological STM following medial-temporal lesions is at least consistent with the proposal that binding perceptual and motor representations of spo-

ken words can involve medial-temporal regions (Knott & Marslen-Wilson, 2001).

Perhaps the clearest example of multimodal word knowledge is in the domain of semantics (Hauk, Johnsrude, & Pulvermüller, 2004; Martin & Chao, 2001; Barsalou, 1999). Learning the meaning of a novel word involves associating representations of word form with various perceptual and motor representations that encode the likely referent of that word. Because the present studies only assessed word-form learning, we can only speculate as to whether a similar two-stage neural process is critical for learning form-meaning associations. Existing behavioral evidence suggests that the automatic activation of associated meanings for newly learned words (semantic priming) requires off-line consolidation (Dumay, Gaskell, & Feng, 2004). However, neuroimaging data have so far focused on initial acquisition of word meaning (Mestres-Misse et al., 2007; Breitenstein et al., 2005; Cornelissen et al., 2004). Further investigations are, therefore, required to assess the role of overnight consolidation in learning form-meaning associations and to test for neural consequences of consolidation on the activation of semantic representations for newly learned spoken words.

## Acknowledgments

This work was supported by the UK Medical Research Council (U.1055.04.013.00001.01, G0000071) and the Economic and Social Research Council (RES-063-27-0061). Mark MacDonald was supported by an Undergraduate Research Bursary from the Nuffield Foundation. We would like to thank the staff at the Wolfson Brain Imaging Center, University of Cambridge, for their help with data acquisition and our volunteers for their participation.

Reprint requests should be sent to Matthew H. Davis, MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 7EF, UK, or via e-mail: matt.davis@mrc-cbu.cam.ac.uk, or to M. Gareth Gaskell, Department of Psychology, University of York, York, YO10 5DD, UK, or via e-mail: g.gaskell@psych.york.ac.uk.

## Notes

1. Recent work suggests behavioral differences between words taught with and without associated meanings (Leach & Samuel, 2007). Our intention was, therefore, that this rating task should index the degree to which participants "invented" a meaning for novel spoken words that were presented without an associated meaning allowing comparison with other studies in which word meanings are taught to participants.
2. An anonymous reviewer suggested that the inclusion of these two segment offsets made certain of our items phonologically unusual. We quantified this by computing biphone probabilities for the entire set of 90 items, which showed a small yet reliable difference in the predicted direction: mean (*SD*) biphone frequency, words = 10,487 (5336), pseudowords = 8424 (5095),  $t(89) = 3.09$ ,  $p < .01$ . Although this small difference in phonological typicality might contribute to the overall response differences between words and pseudowords, it cannot explain changes in word/pseudoword differences due

to learning and consolidation because items were assigned to training conditions in a counterbalanced fashion.

3. This pattern might suggest that the hippocampal response recovers following sleep, although direct comparison of responses to unconsolidated and consolidated novel items does not reach statistical significance [ $t(15) = 1.65, p = .12$ ] in the first scanning session. This difference was significant in subsequent scanning sessions, although responses to unconsolidated novel items were then greater than for consolidated items [ $t(15) = 2.20, p < .05$ ].

4. Our study does not demonstrate that overnight consolidation is necessary for enacting the changes in representation that we discuss because we cannot rule out the possibility that some unspecified training regime might produce such changes in the absence of overnight consolidation. However, we have demonstrated that in the context of a reasonably representative exposure session, overnight consolidation facilitates the engagement of novel word representation into various aspects of normal word processing including lexical competition. Indeed, behavioral experiments have assessed a range of training methods including supplying a meaning and a sentential context and have found no evidence of engagement in lexical competition prior to sleep (e.g., Dumay et al., 2004).

## REFERENCES

- Alario, F. X., Chainay, H., Lehericy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research, 1076*, 129–143.
- Altmann, G. T. M. (1997). *The ascent of Babel*. Oxford, UK: Oxford University Press.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–609.
- Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., et al. (2003). Voxel-based lesion-symptom mapping. *Nature Neuroscience, 6*, 448–450.
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60*, 343–355.
- Binder, J. R., Frost, J. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., & Posing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex, 10*, 512–528.
- Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *Neuroimage, 32*, 821–841.
- Bosshardt, S., Schmidt, C. F., Jaermann, T., Degonda, N., Boesiger, P., Nitsch, R. M., et al. (2005). Effects of memory consolidation on human hippocampal activity during retrieval. *Cortex, 41*, 486–498.
- Bower, G. H., Thompson-Schill, S. T., & Tulving, E. (1994). Reducing retroactive interference: An interference analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 51–66.
- Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition, 97*, B45–B54.
- Breitenstein, C., Jansen, A., Deppe, M., Foerster, A. F., Sommer, J., Wolbers, T., et al. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *Neuroimage, 25*, 958–968.
- Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox. *Neuroimage, 16*, 497.
- Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron, 48*, 687–697.
- Catani, M., Jones, D. K., & ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology, 57*, 8–16.
- Chen, S. H., & Desmond, J. E. (2005). Cerebrocerebellar networks during articulatory rehearsal and verbal working memory tasks. *Neuroimage, 24*, 332–338.
- Cornelissen, K., Laine, M., Renvall, K., Saarinen, T., Martin, N., & Salmelin, R. (2004). Learning new names for new objects: Cortical effects as measured by magnetoencephalography. *Brain and Language, 89*, 617–622.
- Cusack, R., Brett, M., & Osswald, K. (2003). An evaluation of the use of magnetic field maps to undistort echo-planar images. *Neuroimage, 18*, 127–142.
- Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research, 229*, 132–147.
- Dobbins, I. G., Schnyer, D. M., Verfaellie, M., & Schacter, D. L. (2004). Cortical activity reductions during repetition priming can result from response learning. *Nature, 428*, 316–319.
- Drosopoulos, S., Schultze, C., Fischer, S., & Born, J. (2007). Sleep's function in the spontaneous recovery and consolidation of memories. *Journal of Experimental Psychology: General, 136*, 169–183.
- Duff, M. C., Hengst, J., Tranel, D., & Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. *Nature Neuroscience, 9*, 140–146.
- Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*, 35–39.
- Dumay, N., Gaskell, M. G., & Feng, X. (2004). A day in the life of a spoken word. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 339–344). Mahwah, NJ: Erlbaum.
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience, 15*, 399–402.
- Fiebach, C. J., Ricker, B., Friederici, A. D., & Jacobs, A. M. (2007). Inhibition and facilitation in visual word recognition: Prefrontal contribution to the orthographic neighborhood size effect. *Neuroimage, 36*, 901–911.
- Fischer, S., Drosopoulos, S., Tsen, J., & Born, J. (2006). Implicit learning-explicit knowing: A role for sleep in memory system interaction. *Journal of Cognitive Neuroscience, 18*, 311–319.
- Forster, K. L., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, 35*, 116–124.
- French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences, 3*, 128–135.
- Gabrieli, J. D., Cohen, N. J., & Corkin, S. (1988). The impaired learning of semantic knowledge following bilateral medial temporal-lobe resection. *Brain and Cognition, 7*, 157–177.
- Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89*, 105–132.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12*, 613–656.
- Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage, 15*, 870–878.



- Gooding, P. A., Mayes, A. R., & van Eijk, R. (2000). A meta-analysis of indirect memory tests for novel material in organic amnesics. *Neuropsychologia*, *38*, 666–676.
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Science*, *10*, 14–23.
- Hauk, O., Johnsrude, I., & Pulvermuller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301–307.
- Henson, R. N., Price, C. J., Rugg, M. D., Turner, R., & Friston, K. J. (2002). Detecting latency differences in event-related BOLD responses: Application to words versus nonwords and initial versus repeated face presentations. *Neuroimage*, *15*, 83–97.
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*, 67–99.
- Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. *Journal of Neuroscience*, *23*, 9541–9546.
- Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience*, *10*, 100–107.
- Josephs, O., & Henson, R. N. (1999). Event-related functional magnetic resonance imaging: Modelling, inference and optimization. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *354*, 1215–1228.
- Knott, R., & Marslen-Wilson, W. (2001). Does the medial temporal lobe bind phonological memories? *Journal of Cognitive Neuroscience*, *13*, 593–609.
- Kotz, S. A., Cappa, S. F., von Cramon, D. Y., & Friederici, A. D. (2002). Modulation of the lexical-semantic network by auditory semantic priming: An event-related functional MRI study. *Neuroimage*, *17*, 1761–1772.
- Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*, 306–353.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence-intervals in within-subject designs. *Psychonomic Bulletin and Review*, *1*, 476–490.
- Majerus, S., Collette, F., Linden, M. V. D., Peigneux, P., Laureys, S., Delfiore, G., et al. (2002). A PET investigation of lexicality and phonotactic frequency in oral language processing. *Cognitive Neuropsychology*, *19*, 343–361.
- Majerus, S., Van der Linden, M., Collette, F., Laureys, S., Poncet, M., Degueldre, C., et al. (2005). Modulation of brain activity during phonological familiarization. *Brain and Language*, *92*, 320–331.
- Marslen-Wilson, W. (1984). Function and processing in spoken word recognition: A tutorial review. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processing*. Hillsdale NJ: Erlbaum.
- Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology*, *11*, 194–201.
- Mattys, S. L., & Clark, J. H. (2002). Lexical activity in speech processing: Evidence from pause detection. *Journal of Memory and Language*, *47*, 343–359.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–457.
- McClosky, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 109–165). New York: Academic Press.
- McGonigle, D. J., Howseman, A. M., Athwal, B. S., Friston, K. J., Frackowiak, R. S., & Holmes, A. P. (2000). Variability in fMRI: An examination of intersession differences. *Neuroimage*, *11*, 708–734.
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, *17*, 1692–1696.
- Mestres-Misse, A., Rodriguez-Fornells, A., & Munte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex*, *17*, 1858–1866.
- Newman, S. D., & Twieg, D. (2001). Differences in auditory processing of words and pseudowords: An fMRI study. *Human Brain Mapping*, *14*, 39–47.
- Noll, D. C., Genovese, C. R., Nystrom, L. E., Vazquez, A. L., Forman, S. D., Eddy, W. F., et al. (1997). Estimating test-retest reliability in functional MR imaging. II: Application to motor and cognitive activation studies. *Magnetic Resonance in Medicine*, *38*, 508–517.
- Okada, K., & Hickok, G. (2006). Identification of lexical-phonological networks in the superior temporal sulcus using functional magnetic resonance imaging. *NeuroReport*, *17*, 1293–1296.
- O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in Cognitive Sciences*, *6*, 505–510.
- Orfanidou, E., Marslen-Wilson, W. D., & Davis, M. H. (2006). Neural response suppression predicts repetition priming of spoken words and pseudowords. *Journal of Cognitive Neuroscience*, *18*, 1237–1252.
- Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research—A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 785–794.
- Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*, 859–862.
- Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 7865–7870.
- Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with “The language-as-fixed-effect fallacy”: Common misconceptions and alternative solutions. *Journal of Memory and Language*, *41*, 416–426.
- Rosenzweig, E. S., Barnes, C. A., & McNaughton, B. L. (2002). Making room for new memories. *Nature Neuroscience*, *5*, 6–8.
- Saykin, A. J., Johnson, S. C., Flashman, L. A., McAllister, T. W., Sparling, M., Darcey, T. M., et al. (1999). Functional differentiation of medial temporal and frontal regions involved in processing novel and familiar words: An fMRI study. *Brain*, *122*, 1963–1971.
- Scott, S. K. (2005). Auditory processing: Speech, space and auditory objects. *Current Opinion in Neurobiology*, *15*, 197–201.
- Skaggs, W. E., & McNaughton, B. L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science*, *271*, 1870–1873.
- Smith, E. E., Jonides, J., Marshuetz, C., & Koeppe, R. A. (1998). Components of verbal working memory: Evidence from

- neuroimaging. *Proceedings of the National Academy of Sciences, U.S.A.*, *95*, 876–882.
- Squire, L. R. (1992). Memory and the hippocampus—A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*, 195–231.
- Takashima, A., Petersson, K. M., Rutter, F., Tendolkar, I., Jensen, O., Zwarts, M. J., et al. (2006). Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 756–761.
- Tamminen, J., & Gaskell, M. G. (2008). Newly learned spoken words show long-term lexical competition effects. *Quarterly Journal of Experimental Psychology*, *61*, 361–371.
- Tranel, D., Adolphs, R., Damasio, H., & Damasio, A. R. (2001). A neural basis for the retrieval of words for actions. *Cognitive Neuropsychology*, *18*, 655–670.
- Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., et al. (2007). Schemas and memory consolidation. *Science*, *316*, 76–82.
- Tyler, L. K., Marslen-Wilson, W. D., & Stamatakis, E. A. (2005). Differentiating lexical form, meaning, and structure in the neural language system. *Proceedings of the National Academy of Sciences, U.S.A.*, *102*, 8375–8380.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*, 273–289.
- Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General*, *123*, 34–80.
- van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*, 584–589.
- Van der Linden, M., Cornil, V., Meulemans, T., Ivanou, A., Salmon, E., & Coyette, F. (2001). Acquisition of a novel vocabulary in an amnesic patient. *Neurocase*, *7*, 283–293.
- Verfaellie, M., Croce, P., & Milberg, W. P. (1995). The role of episodic memory in semantic learning: An examination of vocabulary acquisition in a patient with amnesia due to encephalitis. *Neurocase*, *1*, 291–304.
- Verfaellie, M., Koseff, P., & Alexander, M. P. (2000). Acquisition of novel semantic information in lesion location. *Neuropsychologia*, *38*, 484–492.
- Walker, M. P. (2005). A refined model of sleep and the time course of memory formation. *Behavioral and Brain Sciences*, *28*, 51–104.
- Walker, M. P., Brakefield, T., Hobson, J. A., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature*, *425*, 616–620.
- Waring, R., & Nation, P. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6–19). Cambridge, UK: Cambridge University Press.
- Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 387–398.
- Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*, 989–994.
- Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, *265*, 676–679.
- Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *Neuroimage*, *33*, 316–325.
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*, 701–702.
- Xiao, Z., Zhang, J. X., Wang, X., Wu, R., Hu, X., Weng, X., et al. (2005). Differential activity in left inferior frontal gyrus for pseudowords and real words: An event-related fMRI study on auditory lexical decision. *Human Brain Mapping*, *25*, 212–221.
- Xiong, J., Rao, S., Jerabek, P., Zamarripa, F., Woldorff, M., Lancaster, J., et al. (2000). Intersubject variability in cortical activations during a complex language task. *Neuroimage*, *12*, 326–339.
- Ziegler, J., Besson, M., Jacobs, A. M., Nazir, T. A., & Carr, T. H. (1997). Word, pseudoword, and nonword processing: A multitask comparison using event-related brain potentials. *Journal of Cognitive Neuroscience*, *9*, 758–775.