Neuropsychologia 111 (2018) 133–144

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Neural representations for newly learned words are modulated by overnight consolidation, reading skill, and age

Nicole Landi[a,*], Jeffrey G. Malins[b], Stephen J. Frost[c], James S. Magnuson[a], Peter Molfese[d], Kayleigh Ryherd[a], Jay G. Rueckl[a], William E. Mencl[c], Kenneth R. Pugh[a]

[a] *University of Connecticut & Haskins Laboratories, United States*
[b] *Yale University & Haskins Laboratories, United States*
[c] *Haskins Laboratories, United States*
[d] *National Institutes of Health, United States*

ARTICLE INFO

Keywords:
Individual differences
Consolidation
Word learning
Language impairment
Reading disability
Overnight
Sleep

ABSTRACT

Word learning depends not only on efficient online binding of phonological, orthographic and lexical information, but also on consolidation of new word representations into permanent lexical memory. Work on word learning under a variety of contexts indicates that reading and language skill impact facility of word learning in both print and speech. In addition, recent research finds that individuals with language impairments show deficits in both initial word form learning and in maintaining newly learned representations over time, implicating mechanisms associated with maintenance that may be driven by deficits in overnight consolidation. Although several recent studies have explored the neural bases of overnight consolidation of newly learned words, no extant work has examined individual differences in overnight consolidation at the neural level. The current study addresses this gap in the literature by investigating how individual differences in reading and language skills modulate patterns of neural activation associated with newly learned words following a period of overnight consolidation. Specifically, a community sample of adolescents and young adults with significant variability in reading and oral language (vocabulary) ability were trained on two spoken artificial lexicons, one in the evening on the day before fMRI scanning and one in the morning just prior to scanning. Comparisons of activation between words that were trained and consolidated vs. those that were trained but not consolidated revealed increased cortical activation in a number of language associated and memory associated regions. In addition, individual differences in age, reading skill and vocabulary modulated learning rate in our artificial lexicon learning task and the size of the cortical consolidation effect in the precuneus/posterior cingulate, such that older readers and more skilled readers had larger cortical consolidation effects in this learning-critical region. These findings suggest that age (even into late adolescence) and reading and language skills are important individual differences that affect overnight consolidation of newly learned words. These findings have significant implications for understanding reading and language disorders and should inform pedagogical models.

1. Introduction

Word learning depends not only on efficient online binding among phonological, semantic, and orthographic features, but also on consolidation of new word representations into permanent lexical memory (Ashworth et al., 2014). Indeed, across a number of studies researchers have shown that word learning consists of an initial episodic encoding for an isolated encounter, followed by a longer-term, sleep–mediated, consolidation process by which a word becomes lexicalized (Brown et al., 2012; Davis et al., 2009; Davis and Gaskell, 2009a, 2009b; Henderson et al., 2014). Additionally, studies of word learning under a variety of contexts indicate that an individual's oral and printed language skill (e.g., vocabulary knowledge, sight word reading efficiency) can impact the facility of novel word learning in both the printed and spoken domains (cf., Bishop and Hsu, 2015; Litt and Nation, 2014; Warmington and Hulme, 2012). Critically, recent work suggests that these factors (offline sleep-mediated consolidation) and language skill may not be independent. For example, we have found that adults with language impairment have diminished ability to consolidate newly learned speech sounds, even under conditions where initial encoding

---

* Corresponding author.
E-mail address: Nicole.landi@uconn.edu (N. Landi).

https://doi.org/10.1016/j.neuropsychologia.2018.01.011
Received 7 August 2017; Received in revised form 11 January 2018; Accepted 12 January 2018

was unimpaired, implicating specific impairments in offline consolidation of newly learned linguistic information (Earle et al., 2015, 2018). With respect to word learning specifically, recent research finds that individuals with specific language impairment (SLI) show impairments in both initial word form learning and in maintaining newly learned representations over time, implicating mechanisms associated with maintenance that may be driven by deficits in sleep-associated consolidation (McGregor et al., 2013; see also James et al., 2017 for theoretical discussion of how sleep and oral language skill may interact).

Neuroimaging studies have the potential to inform these relations given that both the neural systems for word learning (in both typically developing and language impaired populations) and the neural systems associated with overnight consolidation have been well studied independently. The current study seeks to connect these lines of research by specifically investigating how individual differences in language skill modulate patterns of neural activation associated with newly learned words following a period of offline, sleep-mediated consolidation.

1.1. *Individual differences in the neurobiological mechanisms for word learning*

Neuroimaging studies have been able to provide information about the neural networks that support printed and spoken word learning under varied learning conditions, as well as the plasticity in cortical and subcortical structures that makes this learning possible. For example, Breitenstein and colleagues (2005) examined the implicit acquisition of novel word meanings using fMRI. In their task, adult participants “learned” auditory pseudoword-picture pairings, which co-occurred over ten presentations in the scanner. This was contrasted with a no-learning condition, in which a given picture-pseudoword pair was presented only once. Behavioral performance in the learning condition improved across repetitions, and the imaging data revealed a significant linear decrease in activation across repetitions in the left hippocampus, a critical region for memory formation (cf., Preston and Wagner, 2007) and the left fusiform gyrus, a key region for printed and spoken word representations (e.g., Dehaene et al., 2010; Ludersdorfer et al., 2016; Yoncheva et al., 2010). Further analyses revealed links between activation in the left hippocampus and cortical regions over the course of learning such that individual differences in the magnitude of the decrease in activation levels in the left hippocampus across repetitions were significantly correlated with decreases in activation in left fusiform gyrus and with increases in activation in left inferior parietal lobule (IPL). Moreover, the magnitude of signal decrease across blocks in the left hippocampus was negatively correlated with vocabulary knowledge. This association between vocabulary skill and continued engagement of the hippocampus over the course of implicit learning suggests a relationship between the quality of existing vocabulary knowledge and domain-general mechanisms for learning, providing additional motivation for the exploration of the relationship between language skill and memory consolidation.

A related neuroimaging study of *printed* word learning by Pugh and colleagues (2008) examined in–scanner repetition learning of visual word forms in relation to reading skill. During scanning, good readers and reading disabled (RD) participants were presented with printed real words that were displayed once or repeated six times across a run. Both groups became both faster and more accurate at identifying the words across repetitions; however, whereas good readers showed expected patterns of repetition suppression in a number of language and reading relevant left hemisphere regions (consistent with what Breitenstein et al., 2005 found for spoken words), struggling readers instead showed increased activity in these same regions, including the left fusiform/occipito-temporal gyrus and left superior temporal gyrus, suggesting repetition-associated engagement. Although this study did not examine activation or behavioral measures of reading on subsequent days after the scanning session, other behavioral research suggests that these short-term online repetition gains observed during reading instruction may not translate into longer-term stability for individuals with language or reading disorders (cf. McGregor et al., 2013), possibly due to a failure of the mechanisms involved in overnight, sleep associated consolidation. Findings from this study motivate further exploration of how individual differences in reading impact the neural circuitry that supports word learning over a longer duration of time that includes a period of offline sleep.

A third relevant imaging study of *printed* word learning by Sandak et al. (2004) explored three types of repetition learning of novel printed words prior to fMRI scanning. During exposure, some words were learned via orthographic training (participants made judgments about consonant-vowel patterns), others were learned via phonological training (participants made judgments about pronunciations), and still others were learned via semantic training (through novel picture-pseudoword pairings). Behaviorally, participants showed evidence of better learning (faster naming) for words learned under semantic and phonological training conditions relative to orthographic training. At the neural level, patterns of activation for trained pseudowords were more like those elicited by real words than those elicited by completely novel items across a number of language critical left hemisphere regions (e.g., middle and superior temporal gyrus [MTG, STG], inferior frontal gyrus [IFG], and the supramarginal gyrus [SMG]), suggesting that trained items had become at least partially lexicalized. With respect to training type, phonological training was associated with relatively reduced activation in a number of left hemisphere regions that are important for phonological processing (e.g., SMG and inferior frontal gyrus [IFG]), suggesting greater efficiency of phonological processing for these items. In contrast, semantic training resulted in a pattern of relatively enhanced activation in bilateral MTG and STG. These findings suggest that although different training foci (here, semantic or phonological) can result in similar patterns of word learning at the behavioral level, the neural circuits that support these training mechanisms are, in fact, quite distinct.

These studies raise questions and motivate the current exploration of the neural bases of learning and consolidation of novel form-meaning pairs as a function of reading skill. The current study combines the auditory paired associated learning approach taken by Breitenstein et al. (2005), the out-of-scanner training used by Sandak et al. (2004), and the focus on individual differences in Pugh et al. (2008; reading) and Breitenstein et al. (2005; vocabulary) with the addition of an overnight design to explore the effects of overnight consolidation on the neural systems for representing newly learned spoken form-meaning pairs.

1.2. *The role of sleep-associated consolidation in word learning*

The idea that a period of sleep or rest facilitates retention of newly learned information is not a new one. In the early 20th century, Müller and Pilzecker (1900) noted that the recall of a list of nonsense syllables was more accurate when tested after a period of time rather than immediately after list learning. Based on research beginning in the late 1980s, a two-stage complementary learning systems (CLS) model was proposed for acquisition and neural instantiation of new semantic memories, including words (see Davis and Gaskell, 2009a, 2009b; McClelland et al., 1995; Norman and O’Reilly, 2003). The CLS model postulates separate systems for encoding during the awake state supported by the hippocampus, and for longer-term storage supported by the cortex. Patient data in which one system remains intact after damage to the other provides some support for the independence of these two systems (Squire, 1992), and connectionist models have provided a hippocampal learning and cortical consolidation framework to support these findings (McClelland et al., 1995). The role of sleep specifically was initially explored in animal models, with studies finding that firing in hippocampal place cells in particular areas of awake rats was correlated with activity in those same areas during subsequent slow wave sleep (SWS), suggesting a replaying of events to strengthen memories (Ji and Wilson, 2007; Nádasdy et al., 1999; Pavlides and Winson, 1989; Skaggs and McNaughton, 1996; Wilson and McNaughton, 1994).

With respect to word learning specifically, behavioral studies have shown enhanced lexical competition for newly learned words following a period of offline sleep in both children and adults. One paradigm used to test this lexical competition effect involves providing participants with repeated exposures to novel words that differ from existing base words (real words) in their final syllable (e.g., 'cathedruke' created from the base word 'cathedral'). Critically, base words are chosen to be words with very few existing lexical competitors. After exposure to the new words, participants show slowed identification of the base words, suggesting competition from the newly learned words (Bowers et al., 2005; Gaskell and Dumay, 2003; Lindsay and Gaskell, 2013). Additionally, post training delays that include a period of offline sleep produce larger competition effects; this enhanced lexical competition after sleep is interpreted as an index of consolidation (Brown et al., 2012; Davis et al., 2009; Dumay and Gaskell, 2007; Henderson et al., 2014; Tamminen et al., 2010).[1]

Following from these behavioral findings, Davis et al. (2009) used fMRI to examine the neural correlates of overnight sleep associated consolidation of newly learned words. In this study, functional activation to spoken pseudowords (e.g., cathedruke) that had been implicitly learned prior to a period of sleep was compared to activation associated with pseudowords that had been learned just prior to scanning, as well as to both real words and untrained pseudowords. One key finding was an elevated response to novel words (both same-day trained and untrained pseudowords) relative to existing words in left STG and MTG, bilateral motor cortex, supplementary motor area (SMA) and cerebellum, with more robust effects for the trained pseudowords. However, when existing words were compared to trained and sleep-consolidated pseudowords, differences failed to reach threshold. The authors interpreted this finding to indicate that trained-consolidated pseudowords were represented more similarly to existing words than to other pseudowords. Targeted analyses of the hippocampus (using the same voxels that were shown to be modulated by initial learning in Breitenstein et al., 2005) were also conducted to examine both initial learning and consolidation. Analyses of initial learning revealed that same-day trained pseudowords exhibited less activation of the hippocampus than completely unfamiliar, untrained pseudowords. These findings suggest that the hippocampus primarily contributes to encoding of novel words, supporting the complementary learning systems account of the role for the hippocampus in initial learning but not later representation (Bosshardt et al., 2005; Davis et al., 2009).

A second neuroimaging study of these overnight sleep associated effects by Takashima and colleagues (2014) examined two learning conditions: form-only, in which participants were exposed to novel spoken word forms (as in Davis et al., 2009), and picture-associated, in which participants learned novel word forms and an associated picture (similar to Breitenstein et al., 2005). Participants were scanned immediately after encoding (Day 1) while performing a recognition task (old-new judgment) for the newly learned items and control stimuli (untrained pseudowords and real words) and also 24 h later (Day 2) on the same task. Behaviorally, lexical competition effects were observed on Day 2 for form-only trained items but not for picture-associated items, suggesting faster lexicalization of form-only trained items. However, in a free recall task, performance for picture-associated forms was better initially and improved to a greater extent from Day 1 to Day 2. These findings suggest a potential dissociation in the nature of the representations for form-only trained versus picture-associated items.

Day 1 - Day 2 comparisons of neural activation collapsing across form-only and picture-associated conditions revealed greater activity for Day 1 in the striatum (caudate/putamen) and greater activity on Day 2 in cortical areas of the left hemisphere language circuit (MTG, IFG, middle frontal gyrus [MFG], and insula). These findings are consistent with a transfer from subcortical to cortical representation, or cortical consolidation, as a function of sleep. Direct comparisons of picture-associated and form-only condition-wise activations on Day 2 revealed greater activation for picture-associated items in the left anterior cingulate and left MTG and fusiform, suggesting that providing meaning produces greater engagement of semantic processing and memory regions. Taken together, these results suggest that during lexicalization of a novel word, the relative involvement of neural networks for memory and language differs according to the richness of the information about that word available at encoding.

In sum, these studies of word learning have established overnight consolidation as a contributing factor to word learning. Importantly, attaching a meaning to novel word forms appears to impact the nature of the consolidation effect, which has implications for understanding the role of consolidation in vocabulary learning. However, the majority of the literature on word learning and consolidation has focused on form-only learning. Moreover, there have been no systematic investigations of how the neural bases for overnight consolidation of newly learned word representations might relate to individual differences in reading or language ability. Given that recent behavioral studies implicate deficits in overnight consolidation for reading- and language-impaired individuals (discussed below), an investigation of differences in neural mechanisms for overnight consolidation of newly-learned words as a function of reading and language skill is warranted.

Although the impact of reading skill on consolidation effects in word learning has not been studied, recent findings suggest consolidation effects are reduced in individuals with autism spectrum disorders (ASD) and SLI. Specifically, Henderson et al. (2014) used an overnight design to examine auditory word form learning (similar to Davis et al., 2009, described above) in children with autism and typically developing (TD) controls. Findings revealed improved word recognition for all participants after 14 h. However, only the TD children showed lexical competition effects after 24 h, suggesting atypical consolidation of auditory word form knowledge for the participants with ASD (Henderson et al., 2014). Further, Earle et al., (2015, 2018) examined non-native speech sound category learning using behavioral and electrophysiological indices of pre- and post-sleep discrimination in individuals with a history of SLI as well as control adults. These authors found overnight improvements in discrimination of newly learned categories for adults with no history of learning disability, but markedly smaller discrimination improvements for adults with a history of SLI (despite comparable initial learning). Moreover, recent work finds that vocabulary knowledge accounts for significant variance in sleep-mediated consolidation effects in word learning (measured as increased lexical competition effects following sleep; Henderson et al., 2015; James et al., 2017), providing further support for the relation between consolidation and facility with oral language. These reports suggest atypical overnight consolidation in individuals who have language-associated disabilities, and improved overnight consolidation for children with greater vocabulary knowledge. However, these studies have not explored any change in the neurobiological circuitry associated with sleep and, as such, they cannot implicate cortical consolidation specifically.

In the current paper, we use fMRI to explore the neural systems associated with overnight consolidation of a recently learned spoken artificial lexicon. Specifically, we compare patterns of functional neural activation to pseudowords learned just before scanning to those learned on the previous day, and to untrained novel pseudowords. Further, we ask how these neural systems associated with overnight consolidation

---

[1] Note that several accounts have challenged the need for a period of offline sleep or rest for lexicalization of newly learned word forms (cf. Kapnoula et al., 2015), as well as the necessity of hippocampal involvement in learning new word meanings under certain contexts (cf. Coutanche and Thompson-Schill, 2015).

are modulated by individual differences in reading and language skills, assessed here with measures of single word and pseudoword reading and oral language (vocabulary). This question is important in that most extant research and theory in RD and SLI focuses on processing deficits (e.g., in language and phonological processes) but we have relatively little understanding of learning and consolidation deficits and how they may add to or interact with processing problems. If there is a relationship between language and reading skill and atypical sleep-mediated cortical consolidation, we would expect to see differences in neural activation as a function of overnight consolidation that are modulated by reading or language skill.

With respect to specific predictions, based on findings from Takashima and colleagues (2014) we expect that recognition of newly learned pseudowords that are presented prior to overnight consolidation will be associated with greater cortical engagement, relative to trained but unconsolidated pseudowords. Further, if findings from Davis and colleagues (2009) hold here, we would predict greater activation for untrained pseudowords, relative to trained pseudowords in the hippocampus. With respect to individual differences, although this work is exploratory, we expect greater consolidation effects (broadly speaking) for individuals with better reading and oral language skills.

2. Materials and methods

2.1. Participants

A community-based sample of 59 young adults and late adolescents (ages 16–25) participated. Data from 19 participants was excluded from all analyses for the following reasons: 4 participants did not reach our learning criteria of 90% accuracy in the training session; 5 participants failed to pass fMRI data quality checks (see below for details on motion detection procedures); one participant did not meet the study inclusion criterion of performance IQ > 80 on the Wechsler Abbreviated Scale of Intelligence II (WASI-II; Wechsler and Hsiao-pin, 2011); missing performance IQ scores[2] resulted in the exclusion of data from 9 participants. Of the remaining 40 participants, 19 were female, and mean age was 21.0 years (SD = 2.6). Due to the failure to complete the assessment of interest, data from 3 participants was not included in the individual difference analyses of vocabulary and data from 6 participants was not included in the individual difference analyses of reading (see Table 1 for N of each subsample). Most participants were right handed (80%). No participants reported a history of major psychological disorders or traumatic brain injury. Three participants in our sample indicated some history of reading impairment and 5 participants indicated a history of ADHD.

2.2. Learning data acquisition and analysis

2.2.1. Behavioral assessments of skill

Participants completed a behavioral assessment battery consisting of tests designed to measure word-level reading skill and oral language (vocabulary). This battery included the following tests, which were chosen for compatibility with existing work on individual differences in reading and oral language skill in relation to learning and consolidation: (1) The Woodcock-Johnson Tests of Achievement (WJ-III; Woodcock et al., 2001) Letter-Word Identification (LWID), which measures accuracy in reading letters and words aloud; (2) The WJ-III Word Attack (WA), which measures accuracy in reading pseudowords; (3) The Peabody Picture Vocabulary Test (PPVT-4; Dunn and Dunn, 2007), which measures receptive vocabulary. Percent correct for assessments 1 and 2 were averaged to create a composite decoding/word reading measure (see Table 1 for summary statistics).

Table 1
Descriptive information concerning the group of participants ($N = 40$; 19 female) who performed the behavioral learning and fMRI experiments.

Measure	Mean	Range	Number of Participants[a]
Age	21.2	16–25	40
WJ-III Letter-Word Identification standard score	105.4	85–122	34
WJ-III Letter-Word Identification raw score	70.8	60–76	34
WJ-III Word Attack standard score	103.5	85–118	34
WJ-III Word Attack raw score	29.2	22–32	34
WJ-III LWID/WA Composite	92.2	74.5–100	34
PPVT – 4 standard score	109.9	78–141	37
PPVT – 4 raw score	206.3	174–224	37

Note. WJ-III = Woodcock-Johnson III Tests of Achievement; PPVT-4 = Peabody Picture Vocabulary Test 4.

[a] The number of participants out of the 40 total for whom we obtained a score for each test.

2.3. Training protocol

2.3.1. Stimuli

Participants learned associations between 48 novel spoken monosyllabic CVC pseudowords and pictures corresponding to unusual fish or minerals, in a spoken artificial lexicon learning task modeled after (and extending) that used by Magnuson et al. (2003). Seventy-two single-syllable pseudoword stimuli were created to form an artificial lexicon that allowed for division into three sets (24 items each, Table 2) that were matched on number of phonemes, number of syllables and neighborhood density within this encapsulated set (see Table 3). Two sets were learned at different times prior to the MRI session (the evening before next-day scanning [Day 1], and the next morning ~ 2 h before scanning [Day 2]), and a third was untrained (and not associated with any picture) but presented during scanning for a contrast to the two trained conditions. For the two training sets, the mappings between pictures and novel pseudowords were counter-balanced so that half the participants learned associations between 24 novel pseudowords and fish in the first set whereas the other half of participants learned associations between 24 novel pseudowords and rare minerals in the first set. In addition to pseudowords from this artificial lexicon we also included a set of high frequency monosyllabic CVC real words during the scanning session that were matched to the pseudowords on number of phonemes and syllables. These served two primary functions. First, a subset of these were color names (e.g., blue) that served as targets for participants to identify in-scanner, which would ensure that participants were attending to the words, and secondarily, activation to these real words could be compared to activation elicited by the pseudowords from the artificial lexicon to assess any lexicality effects.

This study was part of a larger project investigating the relations between learning in a variety of artificial lexicon and statistical learning paradigms and individual differences in reading and language skill.

2.3.2. Procedure

Behavioral data were collected using during two sessions spaced one day apart. In Session 1 (4 P.M., Day 1), training consisted of presentation of pictures of two fish (or minerals) and a spoken prompt asking which picture corresponded to a spoken pseudoword (e.g., "Which one is the /dɑʃ/"; see Fig. 1.). Participants clicked on either of the two pictures to indicate which they believed was paired with the spoken pseudoword (i.e., the target). They heard positive feedback for correct identifications (e.g., "That's right, that's the /dɑʃ/") and heard "Try again" for incorrect identifications. Participants completed between 2 and 7 blocks of trials, each 72 trials in length. Across the 72 trials, each of the 24 stimulus items was presented six times (three times as the target picture and three times as the alternate picture). Training

---

[2] Data are unavailable because these participants were unable to complete the entire behavioral assessment and testing session.

Table 2
The full stimulus list used in the fMRI experiment.

Consolidated Trained Items	Unconsolidated Trained Items	Novel Pseudowords	Existing Words	Color Words
bith /bɪθ/	dup /dˆp/	baitch /baɪtʃ/	car /kær/	aqua /ɑkwə/
thaj /ðɑdʒ/	dahd /dɑd/	bayj /bedʒ/	cat /kæt/	beige /beiʒ/
theb /ðeb/	doth /dɑθ/	jaub /dʒaʊb/	dime /daim/	black /blæk/
dahsh /dɑʃ/	fup /fˆp/	jauth /dʒaʊð/	door /dɔɚ/	blue /blu/
daz /dɑz/	foid /fɔɪd/	jauf /dʒaʊf/	gas /gæs/	brown /braʊn/
dayb /deb/	foith /fɔɪð/	jaus /dʒaʊs/	girl /gɝl/	cyan /saɪæn/
jayb /dʒeb/	gup /gˆp/	jaut /dʒaʊt/	goose /gus/	gold /goʊld/
joog /dʒug/	gehm /gɛm/	jauthe /dʒaʊθ/	hat /hæt/	green /grin/
geed /gid/	gois /gɔɪs/	faitch /faɪtʃ/	hill /hɪl/	grey /grei/
geeth /gið/	mup /mˆp/	fen /fɛn/	hoop /hup/	pink /pɪŋk/
geej /gidʒ/	mois /mɔɪs/	fethe /fɛθ/	house /haʊs/	red /rɛd/
geem /gim/	nup /nˆp/	laitch /laɪtʃ/	jeep /dʒip/	tan /tæn/
geep /gip/	nem /nɛm/	leb /lɛb/	king /kɪŋ/	teal /til/
geesh /giʃ/	paf /pɑf/	ledj /ledʒ/	kite /kait/	white /wait/
mayb /meb/	sooth /sʊð/	nouf /nʊʃ/	lawn /lɑn/	
nin /nɪn/	souk /sʊk/	shaitch /ʃaɪtʃ/	mom /mɑm/	
niz /nɪz/	soum /sʊm/	shet /ʃet/	moth /mɑθ/	
sahj /sɑdʒ/	soun /sʊn/	sheth /ʃeθ/	mouse /maʊs/	
sayb /seb/	soutch /sʊtʃ/	saitch /saɪtʃ/	pill /pɪl/	
soog /sug/	souve /sʊv/	seb /sɛb/	roof /ruf/	
thayb /θeb/	thaf /θɑf/	taitch /taɪtʃ/	rug /rˆg/	
thoon /θun/	chup /tʃˆp/	choush /tʃʊʃ/	song /sɔŋ/	
thoosh /θuʃ/	cheth /tʃɛð/	toun /tʊn/	thief /θif/	
zith /zɪθ/	chethe /tʃɛθ/	tout /tʊt/	van /væn/	

Note: Transcriptions are in standard IPA; *ʊ is represented by “ou” as in could.

Table 3
Stimulus characteristics.

	Consolidated Trained Items	Unconsolidated Trained Items	Novel Pseudowords	Existing Words	Color Words
Number of Syllables	1	1	1	1	1.14
Number of Phonemes	3	3	3	3	3.5
Phonological Neighborhood Density	14.63	15.83	16.6	29.16	21.79
Relative Phonological Neighborhood Density	3.45	3.541	3.04	.54	.21

Phonological neighborhood density was calculated using IPhOD's unstressed neighborhood densities, which uses the SUBTLEXus database.
Relative phonological neighborhood density was calculated by counting the number of words in the experimental set that were neighbors (i.e., differ by a single phoneme) of a given word.

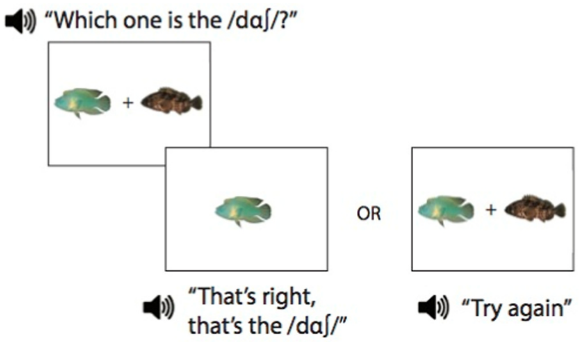

Fig. 1. Sample trial sequence for the behavioral training.

terminated either the first time 90% accuracy on a block was attained (after the two-block minimum) or after a maximum of 7 blocks of testing.

In Session 2 (10 a.m., Day 2), participants returned the next morning for a refresher consisting of two blocks of the same 24 items from Session 1 (hereafter ‘trained consolidated’ items). Immediately following the refresher, participants performed another learning task, identical in structure to Session 1; the only difference was that this task used a new set of 24 novel pseudowords and pictures (hereafter ‘trained unconsolidated items’). The two-block refresher on Day 2 was included to offset the well-known recency effect on memory performance (Crowder, 1976) and initially to provide a behavioral measure of retention of Day 1 learning.[3] Again, participants completed between 2 and 7 blocks of trials, each 72 trials in length, terminating once 90% accuracy was attained. For analysis, reaction times were locked to the onset of the novel pseudoword, and were only analyzed for correct trials. RTs less than 250 ms or greater than 3074 ms were removed (6.2% of trials); RTs below the lower limit were considered invalid anticipations, whereas the upper limit was defined based on 1.5 times the interquartile range above the third quartile of reaction times for correct trials.

2.4. fMRI experiment

2.4.1. In-scanner protocol

On the afternoon of Day 2 at 1 p.m., after the behavioral training sessions, and 20 h after Day 1 learning, participants performed an fMRI task. In an event-related design (Friston et al., 1999; Rosen et al., 1998), participants heard a pseudo-randomly presented set of: (A) the trained consolidated items from Sessions 1 and 2; (B) the trained unconsolidated items from Session 2; (C) a set of completely novel pseudowords; (D) a set of commonly known monosyllabic existing words (e.g., mom) and (E) color words (e.g., white) that were used as catch trials (see Table 2 for the full stimulus list and Table 3 for stimulus

---

[3] One outcome of the inclusion of an automatic end to training after participants reached 90% accuracy on Day 1 was substantial variability in the number of blocks completed and thus an inability to directly compare Day 1 and Day 2 performance for retention.

characteristics). We note that words (D) above were included to examine lexicality effects, not for driving expectations about consolidation-associated lexicalization, per se. Specifically, linking our design and stimulus construction to previous artificial lexicon research resulted in several differences from many of the pertinent studies looking at lexicalization.[4] Moreover, as shown in Table 3, neighborhood metrics were tightly matched among the pseudoword conditions, but differed from those of the words in the study; thus, whereas our design provides optimal matching within the pseudoword set, the ability to look at lexicalization is limited.

The mean duration across conditions for these spoken items was 464 ms (SD = 143 ms). During scanning, the participants' only task was to press one button with their right hand when they heard a color word, and another button when they heard any other word or pseudoword. This task was designed to ensure participants were awake and paying attention.

Data were collected using a Siemens TIM-Trio 3 T magnetic resonance imaging system (Siemens AG, Erlangen, Germany) using a standard 12-channel head coil. Participants were given explicit instructions concerning the task and were queried between runs if there was any uncertainty about the task. Prior to functional imaging, sagittal localizers were prescribed (matrix size = 240 × 256; voxel size = 1 × 1 × 4 mm; FoV = 240/256 mm; TR = 20 ms; TE = 6.83 ms; flip angle = 25°). Next, anatomical scans were acquired for each participant in an axial-oblique orientation parallel to the intercommissural line (MPRAGE; matrix size = 256 × 256; voxel size = 1 × 1 × 1 mm; FoV = 256 mm; TR = 2530 ms; TE = 3.66 ms; flip angle = 7°). Following this, T2*-weighted images were then collected in the same orientation as the anatomical volumes (32 slices; 4 mm slice thickness; no gap) using single-shot echo planar imaging (matrix size = 64 × 64; voxel size = 3.4375 × 3.4375 × 4 mm; FoV = 220 mm; TR = 2000 ms; TE = 30 ms; flip angle = 80°). To allow for stabilization of the magnetic field, the first six volumes within each run were discarded.

Trials were presented at pseudo-random intervals, with inter-trial onset times jittered between 4 and 7 s. Occasional (~10%) *null trials* – i.e., breaks in the stimulus sequence for additional periods of 6–8 s – were included to increase sensitivity (Friston et al., 1999). Each functional run consisted of 8 items per condition (40 trials total). The items of interest, the pseudowords (as well as non-catch trial existing words) were presented twice: once in the first three runs, then again during the last three runs. Note that pseudowords occurred slightly more often during scanning (60% vs. 40%). Each of the six functional runs lasted 4 min 36 s. A sample trial sequence is illustrated in Fig. 2.

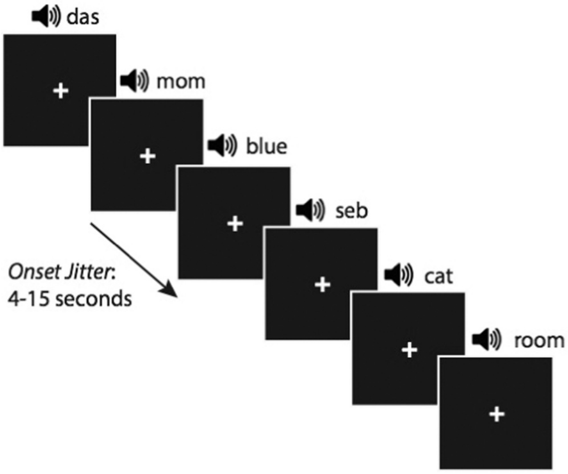

Fig. 2. Sample trial sequence for the fMRI experiment.

2.4.2. fMRI data analysis

Data were analyzed using AFNI (Cox, 1996). Functional images were pre-processed by first correcting for slice acquisition time (*3dTshift*). Following this, functional images were aligned with anatomical images, were corrected for motion using a six-parameter rigid-body transform (*3dvolreg*), and were normalized to the Colin27 brain in Talairach space using a non-linear transform (*3dQwarp*). These three steps were combined into a single transform that also forced a 3 mm isotropic voxel size on the data. Last, all images were smoothed using an 8 mm Gaussian kernel (*3dmerge*).

At the single participant level, data were submitted to a multiple regression analysis with nuisance regressors representing the six movement parameters. This regression used a generalized least squares time series fit, with a restricted maximum likelihood estimation of the temporal auto-correlation structure (*3dREMLfit*). Next, single participant maps were then subjected to a repeated measures analysis of variance (*3dANOVA2*) that tested for a main effect of type (trained consolidated, trained unconsolidated, novel pseudowords, existing words). Resulting groupwise statistical maps were thresholded at a voxelwise threshold of $p = .001$. To control for family-wise error rates, Monte Carlo simulations were performed (*3dClustSim*; 10,000 iterations) using all brain voxels within the TT_N27 template brain, and using the spherical autocorrelation function parameters concerning the error time series (performed in response to the latest recommendations regarding cluster correction in fMRI research; Eklund et al., 2016). The minimum cluster threshold for a corrected alpha level of $p = .05$ was 39 voxels (3 mm isotropic).

3. Results

Mean accuracy and reaction time for the behavioral learning task are shown in Table 4. During Session 1 training, the majority of individuals reached the 90% criterion by the second or third block. During the Session 2 refresher, most individuals reached the 90% criterion during the first block. This high performance for the Session 2 refresher reflects the design of the training schedule: by enforcing a criterion of 90% accuracy in Session 1, we promoted a high degree of learning of the items, which allowed us to maximize power for the contrast between trained consolidated and trained unconsolidated items in the fMRI task. For the new learning blocks in Session 2, participants also learned the novel pairings quickly, and accuracy for the initial block of Session 2 learning was higher than it was for the initial block of Session 1 learning [$t(39) = 4.70$, $p < .001$], suggesting some task transfer effects. These transfer effects are further supported by the observation that participants required fewer blocks to reach the criterion of 90% accuracy in Session 2 learning than they did in Session 1 learning [$t(37) = -4.43$, $p < .001$].

3.1. Individual differences in behavioral training

Although all participants included in our analyses learned our pseudoword picture pairings to criterion by the end of 7 blocks, there was variability in how quickly participants learned the pairings, with some participants reaching our 90% criterion after just two blocks and some taking four or more blocks to reach criterion (Table 4). Using simple correlation analyses, we found that the number of exposures needed to reach criterion was significantly correlated with our individual difference measures of reading and language (see Table 5), such that individuals with better reading skill and vocabulary knowledge required fewer exposures to reach criterion during training. This finding is not unexpected, given previous reports of associations between paired-associate learning and reading and language skill

---

[4] In many studies assessing lexicalization the pseudoword stimuli have a few or no neighbors, (e.g., the only phonological neighbor of 'cathedruke' is 'cathedral'), whereas our pseudoword stimuli typically had many neighbors.

Table 4
Performance in the behavioral training experiment.

Session	Block	Percent Correct	Reaction Time	Number of Participants
Initial Learning (Day 1)	1	64.9 (1.3)	1515 (48)	40
	2	84.0 (1.4)	1334 (41)	40
	3	88.6 (1.3)	1300 (48)	27
	4	88.2 (1.5)	1203 (74)	14
	5	89.0 (2.0)	1302 (88)	9
	6	92.2 (2.6)	1285 (129)	5
	7	90.3	1696	1
Refresher (Day 2)	1	93.1 (.7)	1189 (38)	40
	2	96.9 (.6)	1120 (37)	39
New Learning (Day 2)	1	72.0 (1.4)	1484 (43)	40
	2	89.9 (1.7)	1289 (35)	39
	3	88.3 (4.4)	1396 (93)	12
	4	80.1 (5.3)	1507 (204)	3
	5	81.9	1268	1
	6	93.1	1447	1

Note. Values in parentheses represent standard errors. Reaction times are in milliseconds.

Table 5
Inter-correlations among skill measures, the number of learning blocks required to reach 90% accuracy in Day 1 learning, and age.

Measure	1	2	3
1. WJ-III LWID and WA average percent correct	–		
2. PPVT – 4 raw score	.76***	–	
3. Blocks to Criterion in Day 1 learning	−.70***	−.45**	–
4. Age	.46**	.34*	−.47**

* $p < .05$.
** $p < .01$.
*** $p < .001$.

(Collisson et al., 2015; Li et al., 2009; Litt et al., 2013; Warmington and Hulme, 2012).

3.2. *fMRI task*

3.2.1. *In-scanner behavioral performance*

Performance for the color judgment is shown in Table 6. RTs were only analyzed for correct trials, and were removed from analysis if less than 250 ms or greater than 1775 s (5.3% of trials); RTs below the lower limit were considered invalid anticipations, whereas the upper limit was defined based on 1.5 times the interquartile range above the third quartile of reaction times for correct trials. Importantly, all participants were generally accurate at the task (mean accuracy ranged from 71.3% to 99.6%). In addition, as is evident in Table 6, both accuracy and mean RT were lower for the color word condition compared to the other four conditions. However, because we only focused on the other four conditions (i.e., trained consolidated, trained unconsolidated, novel pseudowords, existing words) for the fMRI analyses, we performed repeated measures ANOVAs to test for a main effect of stimulus type on accuracy and RT across these four conditions. This revealed a main effect of stimulus type on both accuracy [$F(3,117) = 7.77$, $p < .001$, $\eta^2 = .03$] and RT [$F(3,117) = 9.67$, $p < .001$, $\eta^2 = .02$]. Tukey corrected post-hoc *t*-tests revealed that each of the trained conditions (consolidated and unconsolidated) were higher in accuracy compared to novel pseudowords, and trained unconsolidated items were also higher in accuracy compared to existing words. RTs were shorter for trained consolidated items and existing words compared to unconsolidated trained items and novel pseudowords.

Table 6
Behavioral performance for the fMRI task.

Stimulus Type	Percent Correct	Reaction Time
Trained consolidated	95.8 (1.2)	1036 (21)
Trained unconsolidated	96.6 (1.0)	1063 (22)
Novel pseudowords	94.1 (.8)	1080 (23)
Existing (non-color) words	94.3 (1.0)	1036 (21)
Color words	79.7 (2.2)	959 (16)

Note. Values in parentheses represent standard errors. Reaction times are in milliseconds.

3.2.2. *fMRI activation*

We first performed a whole brain analysis to test for a main effect of stimulus type (trained-consolidated, trained-unconsolidated, real words, novel pseudowords) on fMRI activation. As shown in Fig. 3 and listed in Table 7, seven clusters showed a main effect of stimulus type: bilateral STG/MTG; a cluster in the left IPL, including mostly the angular gyrus (AG) as well as a small portion of SMG; a posterior dorsomedial cluster, which included the posterior cingulate and precuneus (PCC/PCun); a cluster in left IFG/MFG; a cluster in anterior cingulate cortex (ACC); and a cluster in the left hippocampus. Following initial whole brain analysis, Tukey corrected post-hoc *t*-tests between stimulus conditions were performed to identify *Consolidation effects*, our primary effects of interest, as well as *Training effects* and *Lexicality effects. Consolidation effects* are reflected in the contrast between trained consolidated items and trained but unconsolidated items. *Training effects* are reflected in the contrast between trained pseudowords (independent of consolidation effects) and untrained pseudowords. Finally, *Lexicality effects* are reflected in the contrast between untrained words and pseudowords.

With respect to *Consolidation effects*, we found greater activation for trained consolidated items relative to trained unconsolidated items in left STG/MTG, left IPL, and in PCC/PCun. With respect to *Training effects*, a number of regions that were not differentially sensitive to consolidated vs. unconsolidated pseudowords showed sensitivity to training. The left hippocampus in particular, showed reduced activation for novel pseudowords relative to both trained consolidated and trained unconsolidated items. Further, the right STG and left MFG showed greater activation for trained consolidated words relative to novel pseudowords. We did not observe any areas that showed greater activation for novel pseudowords relative to either trained consolidated or trained unconsolidated items. With respect to *Lexicality effects*, on average, pseudowords elicited more activation than existing words, but the direct contrast of existing words and untrained pseudowords (simple lexicality effect) revealed greater activation for pseudowords in bilateral STG/MTG, the left hippocampus and in the ACC. Fig. 3 shows all pairwise contrasts for *Consolidation*, *Training* and *Lexicality* effects. Note that we did not observe any effects, as some others have (e.g., Davis et al., 2009), that would be considered lexicalization for the trained items (regions where activity for words was more similar to activity for trained consolidated items than for trained unconsolidated items), which is discussed below.

3.2.3. *Individual differences in activation as a function of consolidation*

Because we were interested in assessing individual differences in neural consolidation, we performed multiple regression analyses on the three clusters that showed significant *Consolidation effects* at the group level (left STG/MTG, left IPL, and PCC/PCun). We performed separate regression analyses for each region; for each analysis, we included the following predictors: reading skill, vocabulary, participant age, and two-way interactions between skill measures and participant age. Given previous findings in the literature demonstrating a link between oral language (vocabulary) skills and consolidation effects and our a priori interest in links between consolidation and reading skill, we conducted separate models for each cluster and skill measure. More specifically, regression models for each cluster included one of the following independent variables: average percent correct for WJ-III LWID and WA, or PPVT-4 raw scores. Given our interest in the relationship between overall reading and language skill and brain activation, rather than age-adjusted reading skill, we used raw scores from each of our assessments

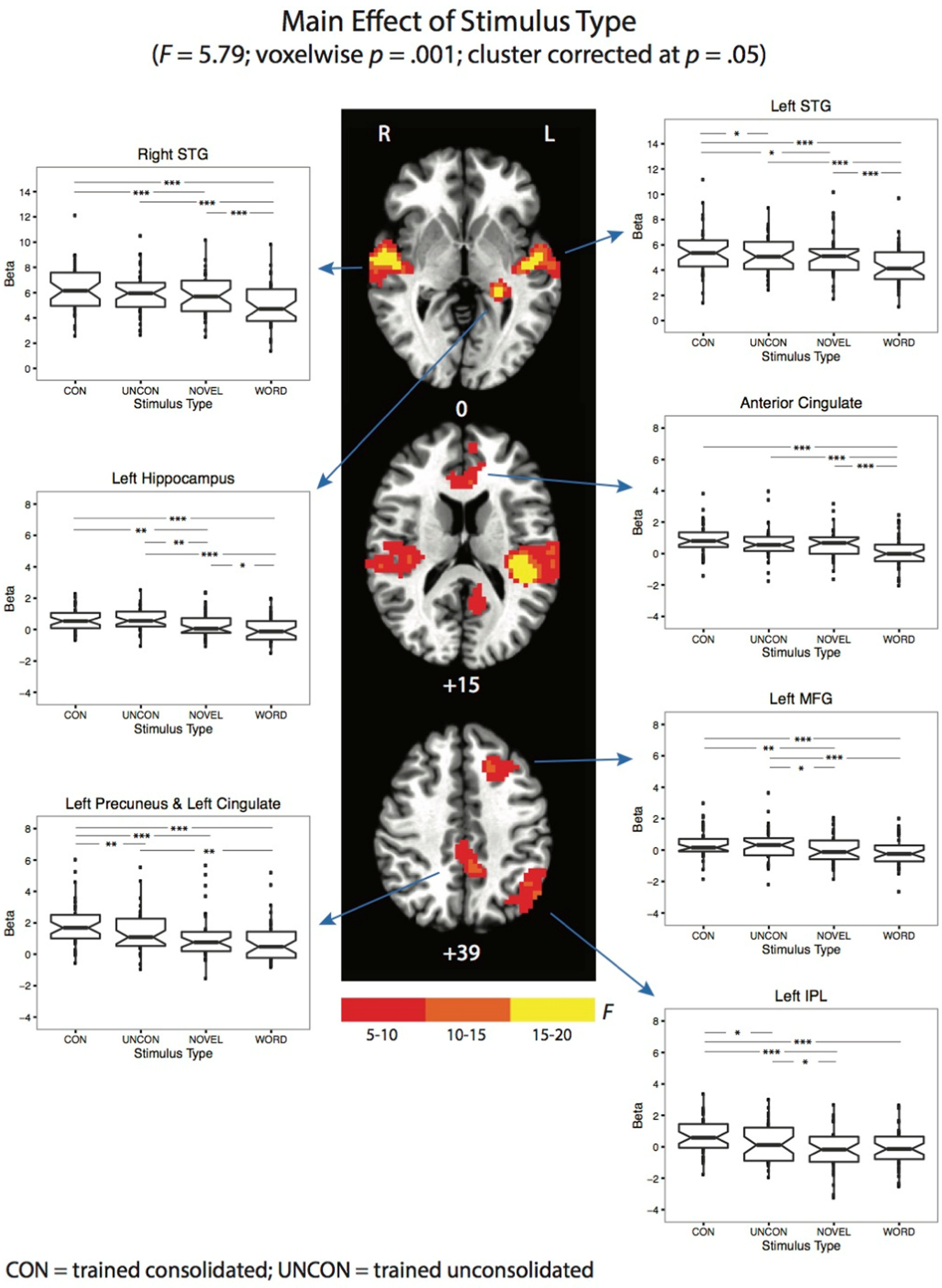

Fig. 3. Clusters showing a main effect of stimulus type in the fMRI experiment. *$p < .05$; ** $p < .01$; *** $p < .001$ (Tukey corrected).

and included main effect and interaction terms for age in each of the regression models; this approach also allowed us to examine any independent effects of age. In all models, the dependent measure was the difference in beta weights between the consolidated and unconsolidated conditions, which were calculated at the individual subject level. The results of these regressions are presented in Table 8.

Table 7
Regions that showed a main effect of stimulus type (voxelwise $p = .001$; cluster corrected at $p < .05$).

Region		Talairach Coordinates of Peak			
L/R	Area	x	y	z	Extent (voxels)[a]
L	Superior temporal gyrus/ Middle temporal gyrus	−41	−32	12	595
R	Superior temporal gyrus	53	−14	3	528
L	Precuneus, middle/posterior cingulate	−14	−62	27	435
L	Angular gyrus/ Inferior parietal lobule	−41	−53	36	236
–	Anterior cingulate	−5	29	12	152
L	Middle frontal gyrus/ Superior frontal gyrus	−17	26	42	127
L	Hippocampus	−26	−32	−4	86

[a] Voxels are 3 × 3 × 3 mm, or 27 mm^3, in size.

Significant variability in the consolidation effect was accounted for by age, reading skill (LWID/WA) and vocabulary (PPVT) in the PCC/PCun. Specifically, a larger consolidation effect (greater difference between trained consolidated and trained unconsolidated items) was associated with better reading ability and vocabulary and increased age. There was no relationship between our individual difference variables and the size of the consolidation effect in any other regions. In addition, age by skill interaction terms were significant predictors of consolidation effects for reading and vocabulary in this region (plotted in Fig. 4). As can be seen in Fig. 4, this interaction appears to be driven by greater range in skill for younger individuals in our sample (e.g., the standard deviation for average percent correct for LWID and Word Attack was 6.8 for the younger half of subjects, compared to 5.2 for the older half of subjects). Nevertheless, reading and vocabulary made significant contributions, independent of age, to the size of the consolidation effect in the PCC/PCun.

4. Discussion

The present study investigated how training and overnight consolidation impacts neural representations for newly learned words that are paired with visual semantic objects. To this end, we identified a neural network associated with overnight *Consolidation* (retrieval of trained consolidated items relative to trained but unconsolidated items). Critically, we also identified a relationship between individual differences in age, reading and vocabulary and behavioral measures of learning, as well as a relationship between these measures and the size of the neural overnight consolidation effect.

With respect to individual differences in our paired associate learning task, we found that the rate of learning on Day 1, here quantified by the number of blocks participants needed to reach the 90% criterion required for task completion, was significantly correlated with all of our individual difference measures. Specifically, older participants and participants with higher scores on our word reading and vocabulary measures reached 90% criterion in fewer blocks. These findings add to a growing literature on the relationship between paired associate learning and reading and language skill (Li et al., 2009; Litt et al., 2013; Warmington and Hulme, 2012).

Our fMRI analysis revealed several effects of interest. Our initial analysis revealed a main effect of stimulus type, which modulated a core network of regions associated with spoken language processing and memory function including bilateral STG/MTG, left IPL, the PCC/PCun, left IFG/MFG and the left hippocampus. Critically, we observed *Consolidation effects* (greater activation for trained consolidated items relative to trained unconsolidated items) in a number of language-associated regions, including the left STG/MTG and left IPL, as well as the PCC/PCun, which has been implicated in learning, memory, and consolidation. No areas showed increased activation for trained unconsolidated items relative to trained consolidated items. We also observed some *Training effects* (greater activation for trained pseudowords relative to novel pseudowords) that were independent of consolidation effects; in particular, we observed reduced activity in the left hippocampus for novel pseudowords relative to both training conditions (see Fig. 3). In general, our finding of increased cortical instantiation for trained and overnight consolidated relative to trained but not overnight consolidated items is in agreement with expectations based on the complementary learning systems account, which postulates an initial period of hippocampal representation during episodic encoding and increasingly greater representation in cortex after a period of offline sleep or rest. However, in our design, we did not observe greater activation in the hippocampus for novel pseudowords that were being newly encoded during scanning, as some others have seen (cf., Davis et al., 2009), which we discuss below.

With respect to specific regions that respond to a greater extent for overnight consolidated versus unconsolidated items, both the left IPL (angular gyrus) and left STG/MTG play a role in storage of lexical semantic information; thus, increases here make good sense. Additionally, STG/MTG have previously been implicated in retrieval of overnight

Table 8
Results of regression analyses between skill measures and the size of the consolidation effect in each of the brain regions that showed a significant effect of consolidation (consolidated vs. unconsolidated trained items).

Brain Region	Predictor	*B*	*SE*	*t*	*p*
Left superior temporal gyrus	WJ-III LWID/WA average percent correct	.19	.22	.86	.40
	Age	.99	.99	1.00	.33
	WJ-III LWID/WA × Age interaction	−.01	.01	−.82	.42
Left inferior parietal lobule	WJ-III LWID/WA average percent correct	.32	.21	1.53	.14
	Age	1.60	.93	1.72	.10
	WJ-III LWID/WA × Age interaction	−.02	.01	−1.55	.13
Left precuneus & left cingulate	WJ-III LWID/WA average percent correct	.66	.20	3.28	< **.01**
	Age	3.25	.90	3.61	< **.01**
	WJ-III LWID/WA × Age interaction	−.03	.01	−3.31	< **.01**
Left superior temporal gyrus	PPVT−4 raw score	.04	.09	.43	.67
	Age	.75	.91	.83	.42
	PPVT−4 × Age interaction	< .01	< .01	−.60	.56
Left inferior parietal lobule	PPVT−4 raw score	.10	.10	1.07	.29
	Age	1.27	.95	1.33	.19
	PPVT−4 × Age interaction	−.01	< .01	−1.18	.25
Left precuneus & left cingulate	PPVT−4 raw score	.22	.10	2.25	**.03**
	Age	2.53	.97	2.62	**.01**
	PPVT−4 × Age interaction	−.01	< .01	−2.36	**.02**

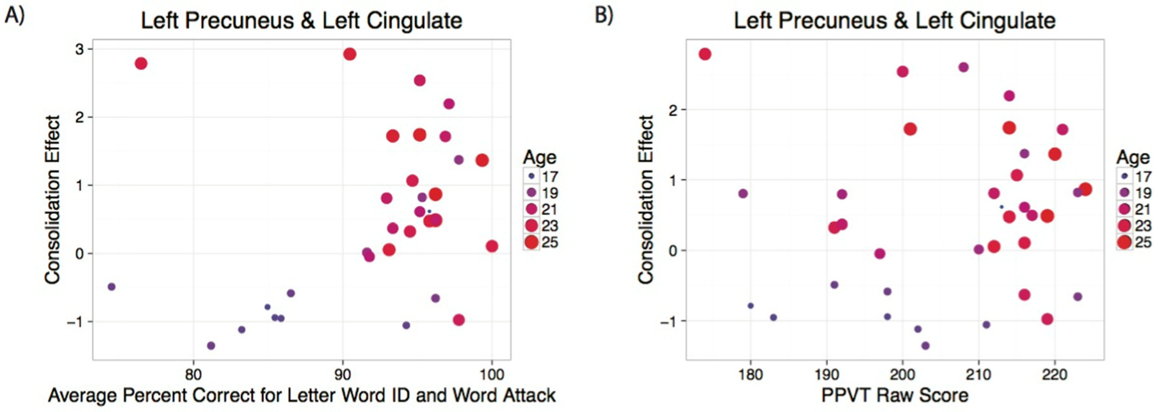

Fig. 4. Associations between reading (A) and language (B) measures and the size of the cortical consolidation effect by age. Color and dot size indicate age range; older individuals show larger consolidation effects overall, however, greater variability in reading and language skill is observed in younger individuals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

consolidated relative to unconsolidated lexical semantic items (Takashima et al., 2014). The PCC/PCun is involved in episodic memory (Eichenbaum, 2001), attention (e.g., Leech et al., 2011) semantic retrieval (Binder et al., 2009) and has been implicated in studies of dyslexia (e.g., Finn et al., 2014). Further, animal studies have shown that this region is more responsive to the retrieval of events that have taken place in the past relative to events learned just prior to test (e.g., Ries et al., 2006), and to retrieval of events that have been consolidated through rehearsal (Bird et al., 2015), potentially suggesting a more domain general role for this region in consolidation of recently learned information.

In addition to the contrast of trained consolidated versus trained unconsolidated items, several effects of stimulus type are worthy of discussion. First, activity in the hippocampus was differentially modulated for retrieval of items that had been trained relative to those that had not. However, activity was greater in this region for retrieval of all trained items (consolidated and unconsolidated) relative to both existing words and completely novel pseudowords, suggesting a main effect of training, not of overnight consolidation. This finding breaks from Davis et al. (2009), who found increased activation for trained unconsolidated items and untrained items relative to trained consolidated items, which they interpreted as evidence for the complementary learning systems (CLS) account. Specifically, they proposed that the CLS is consistent with increases in hippocampus for newly learned items, with consolidated items being represented primarily in cortex. There are several important differences between our study and the study by Davis et al. (2009): first, their study involved form-only training, not form-meaning associative learning, and second, their design was optimized for examining consolidation as well as lexicalization and thus included pseudowords that were more similar to existing words. Indeed, there is ongoing debate concerning the role of the hippocampus in retrieval, and across studies it appears that the nature of the learning and the task strongly contributes to the relative involvement of the hippocampus (i.e., Nadel and Moscovitch, 1997; Norman and O'Reilly, 2003; Yassa and Reagh, 2013). Our findings suggest that more complex (form-meaning) associations may continue to be stored (in some form) in the hippocampus even after overnight consolidation. However, although our stimuli were not optimized to explore lexicalization per se, the fact that this region was more active for pseudowords (trained or not), relative to existing words is consistent with findings of increased activation in this region for novel items.

Consistent with our interpretation of form-meaning associations, Takashima et al. (2014) report activation increases after overnight consolidation in the right hippocampus during retrieval of pseudowords that were previously paired with objects (relative to untrained pseudowords and existing words), but not for pseudowords that had been trained without an object pairing. This finding is consistent with our finding of greater hippocampal engagement for newly learned pseudowords that have semantic representations. A second difference between our findings and those of Davis et al. (2009) is that we did not observe any patterns that could be directly interpreted as increased lexicalization. That is, we did not observe any areas where overnight consolidated items elicited patterns of activation that were more similar to activation for real words than for trained unconsolidated pseudowords. This difference is most likely due to the experimental stimuli and design of the two studies. For one, our design was not optimized for exploring lexicalization, as our pseudowords were part of an artificial lexicon design that maximizes comparability of all pseudowords to each other, rather than creating pseudowords to be highly similar to existing words. Also, our paradigm focused on novel paired associate (form-meaning) learning rather than training of novel pseudoword forms only, and we did not include any existing words in our paired associate training. Thus, our design was optimized for contrasting different levels of training (trained and consolidated, trained unconsolidated, and untrained) within the set of pseudowords in our artificial lexicon.

With respect to individual differences, we examined the relationships between age, reading, vocabulary and the size of the neural consolidation effect. Age, vocabulary (PPVT) and reading skill were associated with the size of the cortical consolidation effect in the posterior cingulate/precuneus. Specifically, older participants and those with better reading and vocabulary scores showed larger cortical consolidation effects in the posterior cingulate/precuneus, a region that has been previously implicated in both retrieval and consolidation. The fact these skills were associated with the size of the cortical consolidation effect in a more domain general region and not in canonical language processing regions is somewhat surprising. However, given that our consolidation contrast was designed to isolate the influence of overnight processing, this may speak to an association between language skills and more domain general mechanisms associated with overnight consolidation. We also observed significant age by skill interactions in this region, with larger skill-related modulation of the cortical consolidation effect in younger readers. This effect is most likely driven by the fact that we had a greater range of skill in our younger readers, which may reflect greater variability in reading skill in younger readers in the population more generally. Importantly, even with age in our model, reading skill and vocabulary terms accounted for significant variance in the size of the consolidation effect in the posterior cingulate/precuneus (Table 8).

Finally, the initial goals of this work focused on gaining a better understanding of the relationship between reading and language skill and overnight consolidation effects, however, it is perhaps most interesting that age was such a strong predictor of cortical consolidation in this study. Although we did not begin this work with explicit predictions about age, the main effect of age on overnight consolidation is not unexpected given observed differences in sleep habits between adolescents (e.g., high school students who often get too little sleep) and adults (see Carskadon, 2011). As such, this finding adds to a growing literature that highlights the importance of overnight consolidation (including sleep) for learning and academic achievement in adolescents (e.g., Backman et al., 2015; Bartel et al., 2016). Further research on the relationship between word learning and overnight consolidation that includes measures of sleep duration and quality in this age range would help elucidate this relationship.

The current study should be viewed in light of several limitations. First, the Day 2 refresher of Day 1 trained items was included to avoid primacy effects associated with our overnight design. However, this resulted in participants having more encounters with the trained consolidated items overall. One potential concern is that this increased exposure may have led to greater activations in this condition regardless of consolidation. Although we cannot rule out this possibility within the current design, we note that regional variation in our findings for consolidated relative to trained unconsolidated items suggests that this simple explanation is not sufficient. To investigate this possibility further, we ran a follow-up brain-behavior analysis to examine the relationship between number of exposures, given variability in the number of blocks it took participants to reach criterion on Day 1 (also explored behaviorally above as Day 1 learning) and the size of the neural consolidation effect. This analysis revealed no modulation of the neural consolidation effect as a function of exposure (see Supplemental Table 1). Thus, it does not appear that exposure can explain the overall pattern of findings we observed. However, it is still possible that seeing items in the trained consolidated condition across two days had an unanticipated effect that we cannot rule out. Second, we do not have available data on the duration or quality of sleep for our participants; as such, we cannot definitively conclude that sleep, rather than simply duration of time between behavioral learning and in-scanner retrieval drove the pattern of observed results. Further research that includes measures of sleep duration and quality is needed to address this concern. Finally, although our study is well-designed for contrasting activation to pseudowords as a function of training and overnight consolidation, our ability to contrast these pseudowords to existing words to identify lexicalization effects was limited.

In conclusion, we have shown that a period of overnight consolidation (which included a period of sleep) is associated with greater cortical instantiation of newly learned spoken words. This finding was observed in a number of left hemisphere regions that are important for spoken word processing including the left STG and left IPL, as well as the left posterior cingulate/precuneus, which is an important cortical region for learning and memory that has previously been implicated in preclinical work on consolidation (Bird et al., 2015; Ries et al., 2006). We also provide the first demonstration that overnight cortical consolidation of spoken words is modulated by individual differences in age, reading skill, and vocabulary. Older participants and more skilled readers showed increased cortical consolidation in the posterior cingulate/precuneus. These findings have implications for the development of neural systems for word learning in both neurotypical and reading-impaired populations. Specifically, our findings suggest that there are age-related changes in the relationship between sleep and cortical consolidation, which continue into late adolescence. These age-related changes may be driven by differences in the amount of sleep or quality of sleep between younger and older adolescents, which have previously been studied in the context of adolescent mental health status and academic achievement (e.g., Backman et al., 2015; Bartel et al., 2016). Critically, our finding of reduced cortical consolidation in the posterior cingulate/precuneus for poorer readers and those with lower vocabulary scores implicates overnight cortical consolidation as one possible factor in reading and language disorders. This finding is consistent with recent findings from Earle et al., (2015, 2018), McGregor et al. (2013) and James et al. (2017), which suggest reduced sleep-associated gains for individuals with poor oral language skills.

Acknowledgements

This work was funded by the following NIH grants: R01 HD065794; P01 HD001994-46, and also received support from National Science Foundation IGERT grant DGE-1144399 to the University of Connecticut. The authors also thank Annie Stutzman, Jessica Kamzik, Nina Gumkowski and Jocelyn Springfield for assistance with data collection and manipulation.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.neuropsychologia.2018.01.011.

References

Ashworth, A., Hill, C.M., Karmiloff-Smith, A., Dimitriou, D., 2014. Sleep enhances memory consolidation in children. J. Sleep. Res. 23 (3), 304–310.

Backman, H., Laajasalo, T., Saukkonen, S., Salmi, V., Kivivuori, J., Aronen, E.T., 2015. Are qualitative and quantitative sleep problems associated with delinquency when controlling for psychopathic features and parental supervision? J. Sleep. Res. 24 (5), 543–548.

Bartel, K., Williamson, P., van Maanen, A., Cassoff, J., Meijer, A.M., Oort, F., Gradisar, M., 2016. Protective and risk factors associated with adolescent sleep: findings from Australia, Canada, and The Netherlands. Sleep. Med. 26, 97–103.

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb. Cortex 19 (12), 2767–2796.

Bird, C.M., Keidel, J.L., Ing, L.P., Horner, A.J., Burgess, N., 2015. Consolidation of complex events via reinstatement in posterior cingulate cortex. J. Neurosci. 35 (43), 14426–14434.

Bishop, D.V.M., Hsu, H.J., 2015. The declarative system in children with specific language impairment: a comparison of meaningful and meaningless auditory-visual paired associate learning. BMC Psychol. 3 (1), 3.

Bosshardt, S., Schmidt, C.F., Jaermann, T., Degonda, N., Boesiger, P., Nitsch, R.M., Henke, K., 2005. Effects of memory consolidation on human hippocampal activity during retrieval. Cortex 41 (4), 486–498.

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 (3), 45–54.

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A.F., Sommer, J., Wolbers, T., Knecht, S., 2005. Hippocampus activity differentiates good from poor learners of a novel lexicon. NeuroImage 25 (3), 958–968.

Brown, H., Weighall, A., Henderson, L.M., Gaskell, M.G., 2012. Enhanced recognition and recall of new words in 7- and 12-year-olds following a period of offline consolidation. J. Exp. Child Psychol. 112 (1), 56–72.

Carskadon, M.A., 2011. Sleep in adolescents: the perfect storm. Pediatr. Clin. North Am. 58 (3), 637–647. http://dx.doi.org/10.1016/j.pcl.2011.03.003.

Collisson, B.A., Grela, B., Spaulding, T., Rueckl, J.G., Magnuson, J.S., 2015. Individual differences in the shape bias in preschool children with Specific Language Impairment and typical language development: theoretical and clinical implications. Dev. Sci. 18 (3), 373–388.

Coutanche, M.N., Thompson-Schill, S.L., 2015. Rapid consolidation of new knowledge in adulthood via fast mapping. Trends Cogn. Sci. 19 (9), 486–488.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res., Int. J. 29 (3), 162–173.

Crowder, R.G., 1976. Principles of Learning and Memory. Erlbaum, Hillsdale, NJ.

Davis, M.H., Gaskell, M.G., 2009a. A complementary systems account of word learning: neural and behavioural evidence. Philos. Trans. R. Soc. Lond. Ser. B, Biol. Sci. 364 (1536), 3773–3800.

Davis, M.H., Gaskell, M.G., 2009b. A complementary systems account of word learning: neural and behavioural evidence. Philos. Trans. R. Soc. B 364, 3773–3800.

Davis, M.H., Di Betta, A.M., Macdonald, M.J.E., Gaskell, M.G., 2009. Learning and consolidation of novel spoken words. J. Cogn. Neurosci. 21 (4), 803–820.

Dehaene, S., Pegado, F., Braga, L.W., Ventura, P., Filho, G.N., Jobert, A., Cohen, L., 2010. How learning to read changes the cortical networks for vision and language. Science 330 (6009), 1359–1364.

Dumay, N., Gaskell, M.G., 2007. Sleep-associated changes in the mental representation of spoken words. Psychol. Sci. 18 (1), 35–39.

Dunn, D.M., Dunn, L.M., 2007. Peabody Picture Vocabulary Test: Manual. Pearson.

Earle, F.S., Myers, E.B., Gumkowski, N., Landi, N., 2015. Changes to neural sensitivity before and after overnight consolidation of phonetic training is absent in adults with language-based disorders. Poster presented at Proceedings of the Twenty-Second Annual Meeting of the Society for Cognitive Neuroscience, San Francisco, CA.

Earle, F.S., Landi, N., Myers, E.B., 2018. Adults with specific language impairment fail to consolidate speech sounds during sleep. Neurosci. Lett. 666, 58–63.

Eichenbaum, H., 2001. The hippocampus and declarative memory: cognitive mechanisms and neural codes. Behav. Brain Res. 127 (1–2), 199–207.

Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. Proc. Natl. Acad. Sci. USA 113 (28), 7900–7905.

Finn, E.S., Shen, X., Holahan, J.M., Scheinost, D., Lacadie, C., Papademetris, X., Constable, R.T., 2014. Disruption of functional networks in dyslexia: a whole-brain, data-driven analysis of connectivity. Biol. Psychiatry 76 (5), 397–404.

Friston, K.J., Zarahn, E., Josephs, O., Henson, R.N., Dale, A.M., 1999. Stochastic designs in event-related fMRI. NeuroImage 10 (5), 607–619.

Gaskell, M.G., Dumay, N., 2003. Lexical competition and the acquisition of novel words. Cognition 89 (2), 105–132.

Henderson, L., Powell, A., Gareth Gaskell, M., Norbury, C., 2014. Learning and consolidation of new spoken words in autism spectrum disorder. Dev. Sci. 17 (6), 858–871.

Henderson, L., Devine, K., Weighall, A., Gaskell, G., 2015. When the daffodat flew to the intergalactic zoo: off-line consolidation is critical for word learning from stories. Dev. Psychol. 51, 406.

James, E., Gaskell, M.G., Weighall, A., Henderson, L., 2017. Consolidation of vocabulary during sleep: the rich get richer? Neurosci. Biobehav. Rev. 77, 1–13.

Ji, D., Wilson, M.A., 2007. Coordinated memory replay in the visual cortex and hippocampus during sleep. Nat. Neurosci. 10 (1), 100–107.

Kapnoula, E.C., Packard, S., Gupta, P., McMurray, B., 2015. Immediate lexical integration of novel word forms. Cognition 134 (5), 85–99.

Leech, R., Kamourieh, S., Beckmann, C.F., Sharp, D.J., 2011. Fractionating the default mode network: distinct contributions of the ventral and dorsal posterior cingulate cortex to cognitive control. J. Neurosci. 31 (9), 3217–3224.

Li, H., Shu, H., McBride-Chang, C., Liu, H.Y., Xue, J., 2009. Paired associate learning in Chinese children with dyslexia. J. Exp. Child Psychol. 103 (2), 135–151.

Lindsay, S., Gaskell, M.G., 2013. Lexical integration of novel words without sleep. J. Exp. Psychol.: Learn., Mem., Cogn. 39 (2), 608–622.

Litt, R.A., Nation, K., 2014. The nature and specificity of paired associate learning deficits in children with dyslexia. J. Mem. Lang. 71 (1), 71–88.

Litt, R.A., de Jong, P.F., van Bergen, E., Nation, K., 2013. Dissociating crossmodal and verbal demands in paired associate learning (PAL): what drives the PAL–reading relationship? J. Exp. Child Psychol. 115 (1), 137–149.

Ludersdorfer, P., Wimmer, H., Richlan, F., Schurz, M., Hutzler, F., Kronbichler, M., 2016. Left ventral occipitotemporal activation during orthographic and semantic processing of auditory words (*Part A*). NeuroImage 124, 834–842. http://dx.doi.org/10.1016/j.neuroimage.2015.09.039.

Magnuson, J.S., Tanenhaus, M.K., Aslin, R.N., Dahan, D., 2003. The time course of spoken word recognition and learning: studies with artificial lexicons. J. Exp. Psychol. 132 (2), 202–227.

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. Psychol. Rev. 102 (3), 419–457.

McGregor, K.K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., Walker, E., 2013. Why words are hard for adults with developmental language impairments. J. Speech Lang. Hear. Res. 56 (6), 1845.

Müller, G.E., Pilzecker, A., 1900. Experimentelle beiträge zur lehre vom gedächtnis. Z. Psychol. 1, 1–300.

Nádasdy, Z., Hirase, H., Czurkó, A., Csicsvari, J., Buzsáki, G., 1999. Replay and time compression of recurring spike sequences in the hippocampus. J. Neurosci.: Off. J. Soc. Neurosci. 19 (21), 9497–9507.

Nadel, L., Moscovitch, M., 1997. Memory consolidation, retrograde amnesia and the hippocampal complex. Curr. Opin. Neurobiol. 7 (2), 217–227.

Norman, K.A., O'Reilly, R.C., 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol. Rev. 110 (4), 611–646.

Pavlides, C., Winson, J., 1989. Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. J. Neurosci. 9 (8), 2907–2918.

Preston, A.R., Wagner, A.D., 2007. The medial temporal lobe and memory. In: Kesner, R., Martinez Jr.J.L. (Eds.), Neurobiology of Learning and Memory, 2nd ed. Academic Press, pp. 305–337.

Pugh, K.R., Frost, S.J., Sandak, R., Landi, N., Rueckl, J.G., Constable, R.T., Mencl, W.E., 2008. Effects of stimulus difficulty and repetition on printed word identification: an fMRI comparison of nonimpaired and reading-disabled adolescent cohorts. J. Cogn. Neurosci. 20 (7), 1146–1160.

Ries, M.L., Schmitz, T.W., Kawahara, T.N., Torgerson, B.M., Trivedi, M.A., Johnson, S.C., 2006. Task-dependent posterior cingulate activation in mild cognitive impairment. NeuroImage 29 (2), 485–492.

Rosen, B.R., Buckner, R.L., Dale, a.M., 1998. Event-related functional MRI: past, present, and future. Proc. Natl. Acad. Sci. USA 95 (3), 773–780.

Sandak, R., Mencl, W.E., Frost, S.J., Rueckl, J.G., Katz, L., Moore, D.L., Pugh, K.R., 2004. The neurobiology of adaptive learning in reading: a contrast of different training conditions. Cogn. Affect. Behav. Neurosci. 4 (1), 67–88.

Skaggs, W.E., McNaughton, B.L., 1996. Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. Science 271 (5257), 1870–1873.

Squire, L.R., 1992. Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychol. Rev. 99 (2), 195–231.

Takashima, A., Bakker, I., van Hell, J.G., Janzen, G., McQueen, J.M., 2014. Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. NeuroImage 84, 265–278.

Tamminen, J., Payne, J.D., Stickgold, R., Wamsley, E.J., Gaskell, M.G., 2010. Sleep spindle activity is associated with the integration of new memories and existing knowledge. J. Neurosci. 30 (43), 14356–14360.

Warmington, M., Hulme, C., 2012. Phoneme awareness, visual-verbal paired-associate learning, and rapid automatized naming as predictors of individual differences in reading ability. Sci. Stud. Read. 16 (1), 45–62.

Wechsler, D., Hsiao-pin, C., 2011. WASI-II: Wechsler Abbreviated Scale of Intelligence. Pearson.

Wilson, M., McNaughton, B., 1994. Reactivation of hippocampal ensemble memories during sleep. Science 265 (5172), 676–679.

Woodcock, R.W., McGrew, K.S., Mather, N., 2001. Woodcock-Johnson Test of Achievement. Riverside Publishing, Rolling Meadows IL.

Yassa, M.A., Reagh, Z.M., 2013. Competitive trace theory: a role for the hippocampus in contextual interference during retrieval. Front. Behav. Neurosci. 7, 107.

Yoncheva, Y.N., Zevin, J.D., Maurer, U., McCandliss, B.D., 2010. Auditory selective attention to speech modulates activity in the visual word form area. Cereb. Cortex 20 (3), 622–632.