



UNIVERSITY OF LEEDS

This is a repository copy of *Consolidation of vocabulary is associated with sleep in children..*

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/94677/>

Version: Accepted Version

---

**Article:**

Henderson, LM, Weighall, AR, Brown, H et al. (1 more author) (2012) Consolidation of vocabulary is associated with sleep in children. *Developmental Science*, 15 (5). pp. 674-687. ISSN 1363-755X

<https://doi.org/10.1111/j.1467-7687.2012.01172.x>

---

© 2012 Blackwell Publishing Ltd. This is the peer reviewed version of the following article: Henderson, L. M., Weighall, A. R., Brown, H. and Gareth Gaskell, M. (2012), Consolidation of vocabulary is associated with sleep in children. *Developmental Science*, 15: 674–687, which has been published in final form at <https://dx.doi.org/10.1111/j.1467-7687.2012.01172.x>. This article may be used for non-commercial purposes in accordance with the Wiley Terms and Conditions for Self-Archiving.

**Reuse**

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

## Consolidation of vocabulary is associated with sleep in children

Lisa M. Henderson<sup>1</sup>, Anna R. Weighall<sup>2</sup>, Helen Brown<sup>1</sup>, M. Gareth Gaskell<sup>1</sup>

<sup>1</sup> Department of Psychology, University of York, York, UK, YO10 5DD

<sup>2</sup> Department of Psychology, Sociology & Politics, Sheffield Hallam University, Collegiate Crescent,  
Sheffield, UK, S10, 2BP

*Corresponding Author:*

M.G. Gaskell

Department of Psychology

University of York

York

YO10 5DD

*Telephone:* 01904 433181

*Email:* [g.gaskell@psych.york.ac.uk](mailto:g.gaskell@psych.york.ac.uk)

## Research Highlights

- We present the first evidence that sleep-associated mechanisms are involved in vocabulary acquisition in childhood and provide the first support for a complementary learning systems account of vocabulary acquisition across development.
- Using the same paradigm as in adult research, we find that novel words show evidence of engaging in lexical competition with existing items approximately 12 hours after encountering the novel words, but only if sleep has occurred.
- We also provide the first evidence that sleep is important for the strengthening of new phonological representations in children: Children's ability to recall and recognise new phonological forms improved approximately 12 hours after training, but only if sleep occurred.
- In addition, we replicate the key finding that children show sleep-associated consolidation of declarative but not procedural memory in contrast to adults, suggesting that explicit and implicit aspects of word learning (i.e., recall/recognition and changes in lexical activity, respectively) cannot simply be associated with declarative and procedural memory systems.

## **Abstract**

Although the acquisition of a novel word is apparently rapid, adult research suggests that integration of novel and existing knowledge (measured by engagement in lexical competition) requires sleep-associated consolidation. We present the first investigation of whether a similar time-course dissociation characterises word learning across development. Consistent with previous research but counter to adults, 7-12-year-olds showed sleep-associated consolidation effects in declarative but not procedural memory. Nevertheless, the relationship between sleep and word learning was remarkably similar to adults. Following exposure to nonword competitors (e.g., biscali) in the a.m. or p.m., children's ability to recognise and recall the nonwords improved only after sleep (after approximately 12-hrs for the p.m. group and 24-hrs for the a.m. group), with performance stable 1-week later. Novel nonwords only induced lexical competition effects after sleep. These findings suggest that children utilize a dual-memory system when acquiring and integrating new vocabulary and highlight sleep as integral to this process.

Vocabulary learning is a fundamental component of language acquisition. Hence, it is essential that we understand how children acquire new vocabulary so that we can inform developmental models of vocabulary acquisition and understand how best to promote word learning across development. Vocabulary acquisition has typically been studied in a domain-specific way, in isolation from the cognitive constraints that govern learning generally. In some formulations, word learning is thought of as part of a modular system (Pinker & Jackendoff, 2005). An alternative position views word learning in terms of the operation of general memory systems (Davis & Gaskell, 2009). According to the *complementary learning systems* framework (McClelland, McNaughton, & O'Reilly, 1995; Norman & O'Reilly, 2003; O'Reilly & Norman, 2002) new information is initially stored separately from existing knowledge and integrated over time. Evidence suggests that new information is initially sparsely coded in the hippocampus but that offline replay (particularly during sleep) results in strengthening of representations in long-term neocortical memory, where they are stored as overlapping distributed representations (French, 1999; Robins & McCallum, 1999). This dual memory system is proposed to provide enough plasticity to accommodate new memories quickly whilst protecting existing memories from damage. There is, understandably, a large body of research on the mechanisms of vocabulary acquisition during development, but none of this research looks at whether similar sleep-associated mechanisms are involved in vocabulary acquisition during development as they appear to be in adulthood. Word learning can be thought of as a lifelong component of the language system, but it is quite plausible that adults, having already completed the bulk of their vocabulary acquisition, rely on different learning mechanisms than children.

A key prediction of the dual memory system approach is that the competition during recognition between novel words and similar sounding well-established lexical neighbours can only occur after the new lexical entry has been integrated into the lexicon and has reshaped existing neocortical networks involved in language processing. Gaskell and colleagues (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003) have examined how lexical activity changes when adults learn fictitious novel nonwords (e.g., *cathedruke*) that are close neighbours of established words (e.g., *cathedral*). The pause detection paradigm (Mattys & Clark, 2002) was used, such that participants made speeded decisions about the

presence of a silent pause inserted toward the offset of existing words (e.g., *cathedr\_al*). Pause detection latencies in existing words became slower if participants had recently learned an onset competitor "*cathedruke*". This finding is argued to reflect the increased amount of lexical activity at pause onset once a novel competitor has been learned; that is, when there is greater lexical activity at pause onset fewer processing resources can be allocated to pause detection. Importantly, this lexical competition effect emerged 12-hours after exposure to the nonword competitors, but only if sleep occurred. Sleep also benefited free recall of the novel nonwords suggesting that off-line consolidation plays a role in the stabilisation of new phonological representations in addition to aiding the integration of novel words into the lexicon.

In support of the dual systems account of vocabulary acquisition, an fMRI study (Davis, Di Betta, MacDonald, & Gaskell, 2009) reported a strong hippocampal response to unknown nonwords compared with newly learned nonwords. Critically, only newly learned nonwords trained on the day prior to the scan elicited neocortical activation akin to existing words suggesting that the neocortical response to words requires a period of consolidation. Furthermore, sleep spindle activity (11-15 Hz oscillations lasting up to 3 seconds) has been shown to be positively associated with larger overnight increases in lexical competition for existing competitors of taught novel nonwords suggesting that sleep spindles play an active role in lexical consolidation (Tamminen, Payne, Stickgold, Wamsley & Gaskell, 2010). Sleep spindles increase in activity during the up-state of slow oscillations (Molle, Marshall, Gais & Born, 2002) and are temporally associated with hippocampal ripples (Sirota, Csicsvari, Buhl, Buzsaki, 2003), implicating them in hippocampal-neocortical consolidation (Diekelmann & Born, 2010).

Although these results are striking, the fact that the dual systems account of vocabulary acquisition has been developed using adult participants severely limits its scope. During development, lexical memory may be organised such that a sleep-associated period of adjustment is not required for complete lexicalisation of novel words. Heightened plasticity in the developing brain might fulfil the role of the hippocampus in the short term (O'Reilly, 2006). For instance, children are less susceptible than adults to interference from competing information following periods of procedural learning (Dorfberger, Adi-Japha, & Karni, 2007). However, it is unclear how long such a single system could remain viable as

knowledge accumulates. Furthermore, considering that children learn at a much greater rate and also sleep longer and more deeply than adults (Campbell & Feinberg, 2009) it is conceivable that sleep may play a different role in vocabulary acquisition.

Alternatively, it may be that adults and children make use of the same dual-system in the acquisition of new vocabulary. The finding that children operate like adults in this respect would require a significant change in our conception of vocabulary acquisition in development, which has tended to focus on the immediate consequences of novel word exposure. One crucial plank of support for this position is the finding that infants who nap after implicit learning show diminished memory for artificial language features but abstract, to a greater extent than controls, a rule-like grammatical pattern from the stimulus materials (Gomez, Bootzin, & Nadel, 2006). It is therefore important to determine whether changes in children's representations of novel words are associated with sleep. Such a finding would have clear theoretical implications for vocabulary acquisition in children, and might also have clinical implications for children with neurodevelopmental disorders such as autism where poor sleep habits have been associated with poor receptive vocabulary (Malow et al., 2009) and qualitative differences in language processing have been associated with aberrant consolidation processes linked to epileptiform discharges during sleep (Femia & Hasselmo, 2002).

A small number of studies have reported delayed improvements in children's explicit memory for new words, suggesting that children require a period of off-line consolidation before new phonological representations are stabilised. Two previous studies have reported improvements in children's ability to recall and recognise novel nonwords 24-hours after training (Brown, Weighall, Henderson & Gaskell, 2012; Henderson, Weighall, Brown & Gaskell, submitted). Importantly, Brown et al. found no such improvements 3-4 hours after training, suggesting that improvements at the 24-hour retest were not due to practice effects or repeated exposure. Studies including longer-term retests have shown that recognition of novel words improves one week after initial training in 3-6 year old children (Storkel, 2001) and up to 24 weeks later in 6-7 year olds (Dockrell, Braisby & Best, 2007) without any additional exposure. However, these previous studies cannot determine whether sleep or simply time is the key factor in stabilising new phonological representations, allowing improvements in recall and

recognition to emerge at delayed test points. In adults, greater overnight gains in recognition speed for novel nonwords are associated with longer slow wave sleep duration (Tamminen et al., 2010), a finding that is consistent with other reports that slow wave sleep is important for consolidating declarative memories (Diekelmann & Born, 2010). Together these adult studies suggest that sleep plays a role in stabilising and perhaps even enhancing new phonological representations, but this has yet to be established in children. Nonetheless, previous developmental studies demonstrating improvements in recognition of new words at delayed test points, along with evidence that 6-8 year old children show lexical competition effects for newly-learned words 24-hours (but not immediately) after exposure (Henderson et al., submitted) provide support for the position that word learning is incremental in children as it is in adults.

On the other hand, an accumulating body of evidence suggests that there are developmental changes in the involvement of sleep in the consolidation of procedural but not declarative memory (see Wilhelm, Prehn-Kristensen & Born, in press, for a review), lending some weight to the hypothesis that developmental changes in the association between sleep and word learning may also be present. Although word learning is largely dependent on declarative memory, engagement in lexical competition may be an instance of a declarative memory becoming proceduralised (such that spoken word recognition is an over-learned skill that is opaque to conscious control) (Dumay & Gaskell, 2007). Children aged 6-8-years show sleep-associated improvements in declarative memory (as measured by paired-associate learning and 2D object locations tasks) similar to adults (Wilhelm, Diekelmann & Born, 2008; Backhaus, Hoeksfeld, Born, Hohagen & Junghanns, 2008). However, unlike adults, no such sleep advantage has been found for procedural memory in children aged 6-11-years-old, as measured by finger sequence learning (Wilhelm et al., 2008) and serial response time (SRT) tasks (Fischer, Wilhelm & Born, 2007). These results suggest that sleep exerts different effects on procedural memory in children and adults. It is possible that declarative memory consolidation might be disproportionately enhanced relative to procedural memory in children because of their high amounts of slow wave sleep which has been revealed to preferentially strengthen hippocampally-encoded declarative memories (Marshall & Born, 2007; Marshall, Helgadottir, Mollé & Born, 2006). In the light of these findings, one might predict



different influences of sleep on different aspects of word learning, with children showing improvements in declarative (e.g., recall/recognition) but not procedural aspects (e.g., changes in lexical activity).

The lack of an immediate overnight gain in procedural learning does not necessarily exclude an improving influence of post-training sleep on long-term learning. Smaller improvements immediately after post-training sleep may be followed by greater or more stable improvements after a longer time period. For instance, young birds that show stronger post-sleep deterioration of song performance achieve a better final song imitation after 3 months (Deregnacourt, Mitra, Feher, Pytte, & Tchernichovski, 2005). Preliminary support for this view has been obtained by Meulemans, van der Linden, and Perruchet (1998) where SRT performance was compared in children and adults at the end of a learning session and after a 1-week delay. In the SRT task a star successively appeared at one of four locations on a computer screen; participants responded to each stimulus by pressing one of four buttons. Sequence learning (the difference in RT to sequenced versus randomly ordered trials) was closely comparable in 6-11-year-olds and adults; however, only children showed a difference between random and sequenced conditions one week later (attributed to a floor effect in adult RT). Thus, although children may not show sleep-associated improvements in procedural memory after ~12 hours, they nevertheless show retention of that skill after one week.

We tested the hypothesis that nocturnal sleep is associated with enhancement of children's ability to learn novel spoken words and integration of these words into the lexicon. Following previous adult studies (Dumay & Gaskell, 2007), children learned 16 nonsense competitor words ("novel" words) either between 07:30-09:30 (a.m. group) or between 16:30-19:00 (p.m. group). Effects of exposure on explicit memory (recognition, recall) and lexical activity (pause detection latency) were measured 0-hrs, approximately 12-hrs<sup>1</sup>, 24-hrs and 1-wk later. Critically, the ~12-hr retest occurred after a night of sleep for the p.m. group but after a day awake for the a.m. group; conversely, the 24-hr retest occurred after a day awake for the p.m. group but after sleep for the a.m. group. Close comparisons can be made between the present developmental data and the adult data collected from a previous study carried out in our laboratory (Dumay & Gaskell, 2007) since both studies adopt the same experimental design and use highly similar stimuli. To place our word learning results in the context of the existing literature, we

also re-examined the finding that children show sleep-associated improvements in declarative memory (on an 2D-object locations task) but not procedural memory (on an SRT task).

## **Method**

### *Participants*

Fifty-three children were recruited from three boarding schools in North Yorkshire. Children were tested early in the morning and later in the evening in quiet classrooms within schools or boarding houses. Children were randomly allocated to the a.m. group (n=26; 13 males) or the p.m. group (n=27; 16 males). The groups were matched on cognitive, reading and language skills, as shown in tests administered a week before the experiment (Table 1). Informed parental consent was obtained. Children were free from reported learning, neurological or sleep disabilities, had normal or corrected to normal vision and hearing, and were native English speakers. Children aged 7–12 years were recruited. Although vocabulary learning remains well-practised throughout this period this age range was deemed broad enough to detect any developmental changes. At 7-years-old children only show lexical competition effects for highly familiar words (e.g., Metsala, Stavrinos & Walley, 2009; Munson, Swenson & Manthei, 2005); however, by 12 years their lexicons begin to appear to more adult-like (Ojima, Matsuba-Kurita, Nakamura & Hagiwara, 2011). Furthermore, the amplitude and slope of slow waves as well as slow wave activity during sleep increases until approximately 12 years of age and decreases to adult-like levels thereafter (Campbell & Feinberg, 2009; Kurth et al., 2010).

### *Stimuli & Design*

Thirty-two stimulus triplets were devised following previous adult studies (Dumay & Gaskell, 2007) and developmental studies (Brown et al., 2012; Henderson et al., submitted) and comprised an existing word (biscuit), a novel competitor (biscal), and a novel foil (biscan) (see Appendix 2). Existing words were selected to be highly familiar to children and were picturable, highly familiar morphologically simple nouns with a uniqueness point before or on this segment. The 32 stimulus triplets were divided into two equal lists matched on frequency, letter and syllable length, phonological neighbourhood size and uniqueness point (as based on the CELEX database). During training, children heard 16 of the novel

words (from List 1 or List 2, counterbalanced across participants). During the lexical integration (pause detection) test, children heard all 32 existing words; half of these items had a trained competitor (competitor condition), whereas the other half did not (control condition). All stimuli were recorded on a Pioneer PDR 509 system by a female native English speaker.

### *Procedure*

The experiment spanned four sessions (Figure 1). The a.m. sessions were administered between 07:30 and 09:30 and the p.m. sessions were administered between 16:30 and 19:00 (Table 2). The 1-week retests were administered at varying points throughout the day (between 09:00-16:00hrs). There were no group differences for children's ratings of their prior night sleep (0=slept through without waking; 1=woke up once; 2=woke up more than once), sleepiness (rated from 0-10, with 0=wide awake and 10=very sleepy), or self-reported hours of sleep (all  $ps>.05$ ).

Stimuli were presented via headphones, using DMDX (Forster & Forster, 2003). During the *training phase*, children were exposed to each novel word 18 times in two phonological tasks. In the *phoneme monitoring task* children listened to each novel word and indicated whether a pre-specified phoneme (/p/, /t/, /d/, /s/, /m/, /b/) was present. Five practice trials were administered. Each novel word occurred 12 times, twice per phoneme. The inter-trial interval was 500ms. All children scored >62.5% correct (mean 89.06%, SD=6.92%). There were no significant group differences across all phoneme conditions ( $p>.05$ ) or for each phoneme condition separately ( $ps>.05$ ). In the *phoneme segmentation task* children listened to each novel word, repeated it, and then said the first (Block 1) or last sound (Block 2). Novel words were presented three times per block in a randomised order. Three practice trials were administered before each block. Repetition accuracy was at ceiling (mean 98.42%, SD=2.25%) with no group difference ( $p>.05$ ). For the initial and final segmentations, children's mean scores were 98.81% (SD=2.73%) and 93.04% (SD=8.31%) respectively, and did not differ between groups ( $ps>.05$ ).

In the *integration test* children made speeded decisions indicating whether a pause was present in each spoken stimulus by pressing one of two buttons. Stimuli comprised 32 existing words

(competitor condition n=16, control condition n=16) and 32 fillers. Half of the items (experimental and fillers) contained a 200ms pause. Four versions of the task were counterbalanced across participants so that each item was equally represented in the four cells of the design (competitor, pause present; competitor, pause absent; control, pause present; control, pause absent). Pauses were inserted before the second vowel offset if the following consonant was a voiceless plosive and was inserted just after this vowel otherwise. Latency was measured from pause onset.

In *cued recall*, children heard the first syllable (bis-) of the 16 novel words from the training phase and completed the cue using one of the new words. In the *2AFC task* children heard both the novel word (biscal) and its corresponding foil (biscan). Children listened to both items before responding with the number 1 or 2 to indicate which item had been heard during training. The order of the novel word – foil word pairs and the order of the two items within each pair were randomised. Accuracy was recorded for both tasks. No feedback was provided.

A *picture-word matching task* was administered to measure children's knowledge of the existing words. For each trial, one existing word target and three distracters (selected from [www.fotosearch.com/clip-art](http://www.fotosearch.com/clip-art) and matched on AoA to the targets, MRC Psycholinguistic Database; Wilson, 1988) were displayed in quadrants on the screen. A target word was played through headphones and the participant indicated the matching picture. Trial order was randomised but the same distracter images always occurred with the same target. Targets were equally distributed across quadrants. All children scored 100% correct on this task, confirming that the existing words were highly familiar.

To measure *declarative memory*, a 2D object location task was used, based on Wilhelm et al. (2008) and Rasch et al. (2007). Children learned 10 card-pair locations on a 4x5 matrix. The card pairs were coloured pictures of animals and everyday objects, all highly nameable with monosyllabic high frequency names. Throughout the exposure and learning blocks, all 20 possible spatial locations were shown as grey squares geometrically ordered in a 4 x 5 matrix. At the 0-hr session, children were exposed to all pairs once. The first in the pair was presented for 1000ms and then joined by the second

for a further 3000ms. The inter-trial interval was 3000ms. Participants were instructed to remember the locations. Immediately after this, children completed two learning blocks. In each trial the first card of each pair was presented and children clicked on the location of the second card with a mouse. Feedback was provided on all trials by presenting the second card at the correct location for 2000ms. At the end of each trial both cards were replaced by grey squares. A 30 second break separated the two learning blocks. The pairs were presented in a randomised orders. The second learning block determined the child's declarative memory score for the 0-hr session. In subsequent sessions, one learning block was administered using the same cued recall procedure.

An SRT task was used to measure *procedural memory*, based on Meulemans et al. (1998). Children were presented with four boxes on a computer screen and placed their index and middle fingers of each hand on four corresponding keys on a keyboard (c, v, n, m). They responded as quickly as possible to a cartoon character that appeared in one of the four boxes by pushing one of the four keys. The character remained on screen until the button press. A block of trials was composed of randomly ordered trials alternating with a repeating sequence of 8 trials (i.e., CBDACABD, with A being the left key and D the right key). This contrasts with conventional SRT tasks where participants are typically trained on a repeating sequence and are transferred to a random sequence at the end of training. The procedure adopted here avoids confounding performance on the random condition with fatigue and reduces the probability that participants will explicitly discover the task structure since repeating sequence and randomly ordered trials are mixed. Random trials were constructed such that no more than two trials were in the same order as the sequenced trials. The training session (administered at 0-hr) comprised 5 blocks of 72 trials and lasted approximately 6 minutes in total. Each block started with 2 random trials, followed by the sequence (8 trials), after which 8 random trials were presented, and so on. One block comprised 5 presentations of the repeating sequence. Because of their unpredictability, the two first stimuli of the repeating sequence were omitted from the analysis (10 trials per block). The first two trials of each block were also omitted from the analysis. Hence, the sequenced and random trials considered in the analysis comprised the same number of stimuli (n=30). In each block, the four positions appeared in the same proportion for the random trials as for the final 6 items in the sequenced trials (e.g., A and

D, 30%; and B and C, 20%). Hence, the learning of the sequence could not be attributable to an unequal distribution of the four positions. Participants were given a 30 second break between each block. At the 0-hr test, the presentation order of the five blocks was counterbalanced across participants. The subsequent retests comprised two blocks of 72 trials. The first block was considered as a 'warm-up'; the second was used for analysis. Five versions of the retest blocks were produced where the random trials were presented in reverse order from the learning blocks. Participants received a different version at each retest and versions were counterbalanced across participants.

## Results

There were no correlations between age and improvement scores on any of the experimental measures for either the a.m. or p.m. groups (Appendix 1). Hence the data are not presented by age-group.

### *Integration test*

The RT and error data for experimental items are shown in Table 3. RTs <200ms and >2.5 SDs from the condition mean were removed for each participant separately (<2%). The RT data were entered into a 2 (Group; a.m., p.m.) x 2 (Condition; competitor, control), x 4 (Session; 0-hr, ~12-hr, 24-hr, 1-wk) x 2 (List 1, List 2) mixed-design ANOVA. RTs were analysed for correct responses.

Responses were slower for the competitor than control condition (Condition,  $F_1(1, 49)=10.97$ ,  $p<.01$ ,  $\eta_p^2=.18$ ,  $F_2(1, 30)=5.88$ ,  $p<.05$ ,  $\eta_p^2=.16$ ) and RTs became slower over sessions (Session,  $F_1(3, 147)=6.97$ ,  $p<.001$ ,  $\eta_p^2=.13$ ,  $F_2(3, 90)=18.83$ ,  $p<.001$ ,  $\eta_p^2=.39$ ). There was no Group x Condition x Session interaction,  $F_1(3, 147)=0.60$ ,  $p>.05$ ,  $\eta_p^2=.01$ ,  $F_2(3,90)=2.08$ ,  $p=.11$ ,  $\eta_p^2=.07$ , but a Session x Condition interaction ( $F_1(3, 147)=4.14$ ,  $p<.01$ ,  $\eta_p^2=.08$ ,  $F_2(3, 90)=4.26$ ,  $p<.01$ ,  $\eta_p^2=.12$ ) suggested that competition effects changed over the course of the 4 sessions. Furthermore, a Group x Session interaction (marginal by participants;  $F_1(3, 147)=2.59$ ,  $p=.06$ ,  $\eta_p^2=.05$ ,  $F_2(3, 90)=4.35$ ,  $p<.01$ ,  $\eta_p^2=.13$ ) motivated more detailed analyses of the conditions in which learning a novel competitor led to slower RTs. Over the course of the first 24 hours, there was a significant Session x Condition interaction both for the a.m. group (marginal

by participants),  $F_1(2, 48)=3.01, p=.06, \eta_p^2=.11, F_2(2, 60)=4.99, p<.01, \eta_p^2=.14$ , and the p.m. group,  $F_1(2, 50)=3.60, p<.05, \eta_p^2=.13, F_2(2, 60)=4.69, p<.05, \eta_p^2=.14$ . However, difference interaction contrasts showed that the effect of competition changed only between ~12 and 24 hours for the a.m. group, (contrasting 0- and 12-hrs,  $F_1(1,24)<1, F_2(1,30)<1$ ; contrasting 12- and 24-hrs,  $F_1(1,24)=4.22, p<.05, \eta_p^2=.15, F_2(1,30)=9.50, p<.01, \eta_p^2=.24$ ) and only between 0 and ~12 hours for the p.m. group (contrasting 0- and 12-hrs,  $F_1(1,25)=6.76, p<.05, \eta_p^2=.21, F_2(1,30)=8.18, p<.01, \eta_p^2=.21$ ; contrasting 12- and 24-hrs,  $F_1(1,25)=1.60, p>.05, \eta_p^2=.06, F_2(1,30)=2.17, p>.05, \eta_p^2=.07$ ). In other words, competition effects changed overnight but not across a similar daytime period for both groups. Furthermore, at the ~12-hr retest there was a marginal Group x Condition interaction that was significant by items,  $F_1(1,49)=2.76, p=.10, \eta_p^2=.05, F_2(1,30)=6.15, p<.05, \eta_p^2=.17$ , suggesting that the p.m. group showed lexical competition at the 12-hr retest but the a.m. group did not. T-tests were performed (with Bonferroni correction) to examine sleep effects (Table 3). Lexical competition effects (significantly slower responses for competitor than control conditions) were absent in both groups prior to sleep (at 0-hr for both groups, plus at ~12-hr for the a.m. group), whereas they were present in both groups following nocturnal sleep (at ~12-hrs and 24-hrs for the p.m. group; at 24-hrs for the a.m. group) (see Figure 2). At the 1-wk retest the lexical competition effects were numerically similar to the 24-hr effects, but no longer individually significant for each group. Nonetheless, there was a main effect of condition at this time point (67ms, SD=217ms;  $t=2.23, p<.05$ ), implying that lexical competition was still present.

Although more time elapsed between 0- and ~12-hrs for the p.m. group than the a.m. group, there was no correlation between the time that elapsed (hrs) and the size of the lexical competition effects for the p.m. group,  $r=.06$ , or the a.m. group,  $r=-.01$ . This suggests that it was sleep, not time, that was associated with the emergence of the ~12-hr lexical competition effect for the p.m. group. Furthermore, despite clear group differences at the ~12-hr test, there were no group differences on pause detection latency or accuracy at 0- or 24-hrs (all  $F_s<1$ , for competitor and control conditions separately), suggesting that circadian effects are unlikely to account for the results.

Pause detection errors were low and hence were not entered into statistical analysis. However, it is notable that pause detection errors increased significantly from the 24-hr retest to the 1-wk retest

for the a.m. group ( $p < .05$ ) but not the p.m. group ( $p > .05$ ) (see Table 3). It is possible that the greater time between training and sleep had a negative effect on performance for the a.m. group, such that it impeded longer-term lexical memory. Importantly, this result does not undermine our interpretation of the RT data since the key interactions concerned the ~12-hr and 24-hr retests, for which there were no group differences in pause detection errors.

#### *Explicit tests of novel word learning*

The accuracy data for *cued recall* and *2AFC* (Figure 3) were entered into separate 2 (Group; a.m., p.m.) x 4 (Session; 0-hr, ~12-hr, 24-hr, 1-wk) x 2 (List 1, List 2) mixed-design ANOVAs.

*Cued recall* performance improved across sessions (Session,  $F_1(3, 147) = 196.73, p < .001, \eta_p^2 = .80, F_2(3, 90) = 189.47, p < .001, \eta_p^2 = .86$ ) and the p.m. group recalled more novel words than the a.m. group (Group,  $F_1(1, 49) = 11.33, p < .01, \eta_p^2 = .19, F_2(1, 30) = 62.15, p < .001, \eta_p^2 = .67$ ). There was also a Session x Group interaction ( $F_1(3, 147) = 27.61, p < .001, \eta_p^2 = .36, F_2(3, 90) = 38.88, p < .001, \eta_p^2 = .56$ ): The a.m. group showed no improvement after wake between 0-hrs and ~12-hrs (mean 3.85%, SD=10.76%;  $t(25) = -1.82, p > .05$ ) but significant improvements after sleep between ~12-hrs and 24-hrs (mean 27.88%, SD=17.61%;  $t(25) = -8.07, p < .001$ ) and between 24-hrs and 1-wk (mean 17.07%, SD=13.64%;  $t(25) = -6.38, p < .001$ ). The p.m. group showed significant improvements after sleep between 0-hrs and ~12-hrs (mean 35.88%, SD=16.41%;  $t(26) = -11.36, p < .001$ ), and between ~12-hrs and 24-hrs (mean 4.86%, SD=9.23%;  $t(26) = -2.74, p < .01$ ) and 24-hrs and 1-wk (mean 6.72%, SD=8.99%;  $t(26) = -3.88, p < .01$ ).

*2AFC* performance improved across sessions (Session,  $F_1(3, 147) = 11.47, p < .001, \eta_p^2 = .12, F_2(3, 90) = 13.15, p < .001, \eta_p^2 = .31$ ). There was a Session x Group interaction ( $F_1(3, 147) = 4.09, p < .01, \eta_p^2 = .08, F_2(3, 90) = 6.77, p < .001, \eta_p^2 = .18$ ): The a.m. group showed no improvements after being awake between 0-hrs and ~12-hrs ( $t(25) = 0.92, p > .05$ ) but showed a significant improvement after sleep between ~12-hrs and 24-hrs ( $t(25) = -4.37, p < .001$ ) and maintained similar performance at 24-hrs and 1-wk ( $t(25) = -0.27, p > .05$ ). The p.m. group showed a significant improvement after sleep between 0-hrs and ~12-hrs ( $t(26) = -2.83, p < .01$ ) but no improvement after being awake between ~12-hrs and 24-hrs ( $t(26) = 0.13, p > .05$ ) and maintained similar performance between 24-hrs and 1-wk ( $t(26) = 0.84, p > .05$ ).



There were no significant correlations between 0-hr-12-hr improvements in cued recall or 2AFC and the time elapsing between these sessions: Cued recall (a.m.  $r(26)=.12$ ; p.m.  $r(27)=-.07$ ), 2AFC (a.m.  $r(26)=.34$ ; p.m.  $r(27)=.08$ ). Similarly, there were no correlations between the ~12-hr-24-hr improvements in cued recall or 2AFC and the time elapsing between these sessions: Cued recall (a.m.  $r(26)=.04$ ; p.m.  $r(27)=.15$ ), 2AFC (a.m.  $r(26)=-.07$ ; p.m.  $r(27)=-.08$ ). There were no a.m./p.m. group differences for cued recall ( $F(1, 51)=1.18, p>.05$ ) or 2AFC ( $F(1,51)=.05, p>.05$ ) at the 0-hr test suggesting that circadian effects are unlikely to account for the pattern of results.

### *Declarative Memory*

There were no group differences in accuracy for the first training block ( $F_1(1, 51)=1.79, p>.05, F_2(1,18)<1$ ) or second training block ( $F_1(1, 51)<1, F_2(1,18)<1$ ) at 0-hrs, suggesting the two groups started from a similar baseline.

The accuracy data for the object locations task are shown in Figure 4. The data were entered into a 2 (Group; a.m., p.m.) x 4 (Session; 0-hr, 12-hr, 24-hr, 1-wk) mixed-design ANOVA. Accuracy improved across sessions (Session,  $F_1(3, 153)=7.92, p<.001, \eta_p^2=.13, F_2(3,54)=10.42, p<.001, \eta_p^2=.37$ ). There was a trend for a Session x Group interaction ( $F_1(3, 147)=2.04, p=.11, \eta_p^2=.04, F_2(3,54)=2.50, p=.07, \eta_p^2=.12$ ): The a.m. group showed lower accuracy at 12-hrs than 0-hrs ( $t(25)=3.27, p<.01$ ) whereas the p.m. group showed no difference ( $t(26)=-0.68, p>.05$ ). Both groups showed significantly better accuracy at 24-hrs than 12-hrs (a.m.,  $t(25)=-4.76, p<.001$ ; p.m.,  $t(26)=-2.20, p<.05$ ). Neither group showed a difference between 24-hrs and 1-wk (a.m.,  $t(25)=0.48, p>.05$ ; p.m.,  $t(26)=1.12, p>.05$ ). The Session x Group interaction was significant when excluding the 1-wk data from analysis,  $F_1(2, 102)=3.61, p<.05, \eta_p^2=.07, F_2(2,36)=3.53, p<.05, \eta_p^2=.16$ .

### *Procedural Memory*

In the following section, the results of the five training blocks of the SRT task administered in the 0-hr session are presented first followed by an analysis of the sleep effects comparing performance across the four sessions for the a.m. and p.m. groups.

### *Performance during the training blocks in the 0-hr session*

#### *RT*

RT data from the 5 training blocks (Figure 5) were entered into a mixed-design ANOVA with Group (a.m., p.m.), Condition (random, sequenced) and Block (1, 2, 3, 4, 5) as factors. There was no significant main effect of Group,  $F(50)=0.65, p>.05, \eta_p^2=.004$ . There was a significant main effect of Block,  $F(4, 200)=23.30, p<.001, \eta_p^2=.32$ : Children showed faster responses for block 2 than block 1 (mean difference 45ms,  $SD = 74ms, p<.001$ ), and block 3 than block 2 (mean difference 36ms,  $SD=50ms, p<.001$ ) but RT for blocks 3 and 4 (mean difference 9ms,  $SD=54ms, p>.05$ ) and blocks 4 and 5 (mean difference -3ms,  $SD=74ms, p>.05$ ) did not differ. There was a significant main effect of Condition,  $F(1, 50)=33.93, p<.001, \eta_p^2=.40$ , and a significant Block x Condition interaction,  $F(4, 200)=2.51, p<.05, \eta_p^2=.05$ . The Block x Condition x Group interaction approached significance<sup>2</sup>,  $F(4, 200)=2.08, p=.09, \eta_p^2=.04$ : Both groups showed a procedural learning effect in blocks 2 and 3 and this effect diminished for the p.m. group (but not the a.m. group) in blocks 4 and 5. The p.m. group may have become fatigued during the final blocks which may account for their diminishing procedural learning effect during training.

#### *Accuracy*

Accuracy scores during initial training were also analysed using a mixed-design ANOVA. There were significant main effects of Block,  $F(1, 50)=3.55, p<.01, \eta_p^2=.07$ , and Condition,  $F(1, 50)=47.20, p<.001, \eta_p^2=.49$ , and a significant Block x Condition interaction,  $F(4, 200)=3.79, p<.01, \eta_p^2=.07$ : Sequenced trials were responded to more accurately than random trials and this difference increased across blocks. The Block x Condition x Group interaction and the main effect of Group were not significant,  $F(4, 200)=0.61, p>.05, \eta_p^2=.012$ , and  $F(50)=0.87, p>.05, \eta_p^2=.005$ , respectively.

#### *Sleep effects*

#### *RT*

The 0-hr, 12-hr, 24-hr and 1-wk RT data for the SRT task are shown in Figure 6. A mixed-design ANOVA was performed with Group (a.m., p.m.), Condition (random, sequenced) and Session (0-hr, 12-

hr, 24-hr, 1-wk). The 0-hr retest data was calculated by averaging RT for the final three training blocks for each condition. Thus, 0-hr RT and accuracy scores did not differ between groups for random ( $F_s < 1$ ) or sequenced ( $F_s < 1$ ) conditions.

Responses were faster to sequenced than random conditions (Condition,  $F(1,48)=121.85$ ,  $p < .001$ ,  $\eta_p^2=.72$ ) and this effect increased across sessions (Session x Condition,  $F(3, 144)=15.17$ ,  $p < .001$ ,  $\eta_p^2=.24$ ): 0-hrs (mean difference 33ms,  $SD=47ms$ ,  $t(51)=4.98$ ,  $p < .001$ ), 12-hrs (52ms,  $SD=56ms$ ,  $t(50)=6.63$ ,  $p < .001$ ), 24-hrs (74ms,  $SD=58ms$ ,  $t(50)=9.04$ ,  $p < .001$ ), 1-wk (85ms,  $SD=58ms$ ,  $t(51)=10.60$ ,  $p < .001$ ). RT became faster across sessions (Session,  $F(3, 144)=105.13$ ,  $p < .001$ ,  $\eta_p^2=.69$ ), particularly for the a.m. group (Session x Group,  $F(3, 144)=3.81$ ,  $p < .05$ ,  $\eta_p^2=.07$ ). Both groups showed significantly faster RT at 12-hrs than 0-hrs (a.m. group mean difference 98ms,  $SD=81ms$ ,  $t(24)=6.03$ ,  $p < .001$ ; p.m. group mean difference 113ms,  $SD=64ms$ ,  $t(25)=8.99$ ,  $p < .001$ ), suggesting that time rather than sleep was responsible for this improvement. Only the a.m. group showed a significant improvement in RT at 24-hrs than 12-hrs (a.m. group difference 61.02,  $SD=44ms$ ,  $t(24)=6.97$ ,  $p < .001$ ; p.m. group difference -3ms,  $SD=69$ ,  $t(24) < 1$ ). Whilst this could be due to sleep experienced by the a.m. group prior to their 24-hr test, it is more likely due to a circadian dip in performance for the p.m. group who received their 24-hr test in the evening. Finally, both groups showed significantly faster RT at 1-wk than 24-hrs (a.m. mean difference 33ms,  $SD=49ms$ ,  $t(25)=3.41$ ,  $p < .01$ ; p.m. mean difference 52ms,  $SD=75ms$ ,  $t(24)=3.51$ ,  $p < .01$ ). In sum, both groups showed clear evidence of procedural learning across the week of the experiment reflected by an increase in RT and a progressively larger difference between sequenced and random conditions. However, there was no evidence that sleep enhanced this effect.

### *Accuracy*

The same mixed-design ANOVA was performed to analyse accuracy. Response accuracy was significantly greater for sequenced than random conditions (Condition,  $F(1,48)=107.70$ ,  $p < .001$ ,  $\eta_p^2=.69$ ). No other main effects or interactions were significant.

### **Discussion**

This study presents the first developmental evidence that sleep is associated with an opportunity for newly learned spoken words to be integrated into long-term lexical memory: Engagement in lexical competition was not observed shortly after exposure or after a day awake, but was consistently found after a night of sleep. Both groups were retested 12-hours after exposure, but only the group that had slept showed significant lexical competition effects for similar sounding existing words. Crucially, these results build on previous developmental (Brown et al., 2012; Dockrell et al., 2007; Gomez et al., 2006; Henderson et al., submitted; Storkel, 2001; Storkel & Lee, 2011) and adult studies (Dumay & Gaskell, 2007) and provide the first evidence that it is sleep, not simply the passing of time that allows novel spoken words to be incorporated into children's mental lexicons. In addition to linking vocabulary acquisition to sleep, our data suggest the complementary learning systems framework can also provide a good account of vocabulary acquisition across development. This framework predicts that a delay in the integration of a novel spoken word is necessary to prevent new words from overwriting existing ones.

The second novel contribution of our study is that children's ability to recognise and recall the novel spoken nonwords was associated with sleep. Children accurately recognised the novel nonwords immediately post-exposure, consistent with studies showing that children and infants are able to learn novel phonological forms rapidly (Church & Fischer, 1998) and also adult studies where recognition immediately after training is typically near ceiling (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003). However, significant improvements were seen on the recognition task after ~12-hrs, but only for the p.m. group who had slept: for the a.m. group, significant improvements were observed at 24-hrs. Significant improvements in cued recall were also seen after ~12-hrs, but only if that period included sleep: the a.m. group showed significant improvements after 24-hrs (after sleep). The highest performance on the cued recall task was seen after 1-wk. We would argue that recall and recognition tasks may measure only episodic knowledge prior to sleep (since the novel spoken form has not yet been integrated into the lexicon), but that these tasks may measure both episodic and lexical memory after sleep.

Whilst the present study reveals an association between sleep and consolidation of novel words, it does not provide causal evidence. Nonetheless, our results add to a growing body of converging evidence including neuroimaging and polysomnography data that implicate sleep in word learning. The closest indication of causality arises from the finding that sleep spindles and slow wave sleep duration are associated with overnight changes in lexical competition and improvements in novel word recognition speed, respectively (Tamminen et al., 2010). This implies an active role for sleep in memory consolidation (Ellenbogen, Payne, & Stickgold, 2006), and fits with an integrative role for sleep spindles in the consolidation of declarative memory (Walker, 2009).

A potential caveat of our sleep-associated improvements in explicit memory and integration is the influence of circadian effects, namely performance may simply be better in the morning than the evening (Schmidt, Collete, Cajochen, Peigneux, 2007). Indeed, there was some evidence of circadian effects in the SRT data, as discussed below. However, if circadian effects were responsible for the above pattern of data then there should have been group differences at the 0-hr and 24-hr time points which took place at the same circadian time (i.e., the morning for the a.m. group and the evening for the p.m. group). This was not the case for any of the word learning measures at the 0-hr test or for the integration and recognition tests at the 24-hr test. Furthermore, if the a.m. group were more tired at the 12-hr test, this should have been reflected in the participants' subjective evaluations of tiredness; however, no difference was found between the a.m. and p.m. groups in any of the test sessions. Finally, there was no influence on the amount of time that elapsed between sessions and the improvements shown. Together these findings provide a strong case against circadian, fatigue or timing effects.

A significant correlation has previously been reported between post-sleep improvements in recall and post-sleep lexical competition effects ( $r(48)=-.35, p<.02$ ), suggesting that greater improvements in recall are associated with lexical integration (Dumay & Gaskell, 2007). We found only a weak positive correlation between the same two variables ( $r(53)=.23, p=.09$ ). It is possible that greater variability in lexical competition in children accounts for our weaker correlation. Alternatively, there may be additional factors leading to increased lexical competition effects after sleep in children. For instance, since children are likely to have significantly slower pause detection latencies than adults, post-sleep

lexical competition may reflect a combination of lexical competition during the early (automatic) processes involved in word recognition as well as more strategic (conscious) processes involved in noticing the overlap between existing and novel words.

Despite showing clear similarities between sleep-associated consolidation of vocabulary in children and adults, we replicated the previous finding that children show sleep-associated consolidation of declarative but not procedural memory. This situates our word learning data in the context of the sleep literature and supports the view that there are no simple one-to-one mappings between explicit and implicit aspects of word learning (i.e., recall/recognition and changes in lexical activity, respectively) and declarative and procedural memory systems. This view is further strengthened by the lack of correlations between improvements in recall/recognition and declarative learning effects and between increases in lexical competition and procedural learning effects (see Appendix 1).

The results of our declarative memory task replicate and extend previous findings showing that children's retention of declarative memories is more robust following a period of sleep rather than a period of wakefulness after learning (Backhaus et al., 2008; Wilhelm et al., 2008). Namely, the a.m. group showed a significant decline in retention of object pair locations after ~12-hrs of wake whereas the p.m. group showed no decline in performance after ~12-hrs which included sleep (Figure 4). Hence, rather than sleep having an enhancing role on declarative memory in children, sleep appears to have a protective role and recovers performance to the original level (Fenn, Nusbaum, & Margoliash, 2003). The contrast between sleep-associated enhancement of explicit word memory (recall/recognition) and the more protective role of sleep on our declarative memory task further supports the notion that the explicit aspects of word learning cannot simply be associated with declarative memory.

In contrast to previous developmental studies we also included a 24-hr retest in the declarative task to examine whether comparable sleep-associated improvements are observed after ~24-hrs in the a.m. group. As predicted, we found that the a.m. group showed significant improvements in retention after 24-hrs. However, improvements were also observed for the p.m. group (similar to cued recall), suggesting that sleep may have an initial benefit for the retention of new declarative memories but

thereafter, sleep, time and repeated exposure to stimuli may all play a role. Finally, we found that performance remained stable after one week, suggesting that the newly learned object locations had been retained.

This study also supports the finding that children do not show sleep-associated improvements in procedural memory (Fischer et al., 2007; Wilhelm et al., 2008). The use of an SRT task allowed us to examine changes in overall RT in addition to changes in response to sequenced versus random trials, each of which may arguably tap different aspects of procedural learning (motor automaticity versus abstraction of statistical probability, respectively). Similar to Meulemans et al. (1998), children showed significantly faster responses to sequenced versus random trials during training as well as 1-wk later (Figure 6). Moreover, this effect increased after ~12-hrs and 24-hrs similarly for both groups. The absence of a group difference at ~12-hrs suggests that procedural learning is not influenced by sleep in children. However, previous studies using similar tasks have produced slightly different results. For instance, Fischer et al. (2007) used an SRT task with children and reported a decline in sequence learning after sleep and no difference after a period of wake. Wilhelm et al. (2008) reported significantly greater accuracy gains in finger sequence learning after ~12-hrs of being awake than ~12-hrs including sleep, whereas we found no group differences in overall accuracy or RT between 0-hrs and ~12-hrs. Although the present and previous studies share the common theme of no evidence for sleep-associated improvements in procedural memory in children, the disparity in the results is difficult to reconcile and likely reflects methodological differences. Our sample size was substantially larger than previous studies suggesting that the results are unlikely to be due to a lack of statistical power.

It is important to note that circadian effects were evident for the SRT data. First, the a.m. group showed a significant improvement in overall RT between ~12-hrs and 24-hrs but the p.m. group showed no significant improvement (see Figure 6). This could be interpreted as a sleep effect since only the a.m. group slept between 12-24-hrs. However, it is more likely that circadian factors are responsible since the 24-hr test for the p.m. group was administered in the evening (during which there is typically a circadian dip; Schmidt et al., 2007). Although participants did not report being more tired in the p.m. sessions than then a.m. sessions it is possible that minimal differences in attention could have influenced

performance on the SRT task since this task arguably places greater demands on sustained attention than the lexical tasks. Furthermore, both groups showed equivalent improvements in overall RT at ~12-hrs (after wake for the a.m. group and after sleep for the p.m. group), implying that the 12-24-hr improvement for the a.m. group is unlikely to be associated with sleep. Second, children trained in the p.m. showed a significantly smaller difference (in ms) between sequenced and random trials for the last block of training than the a.m. group (see Figure 5). Thus, time of day appears to influence the ability of children to express what they have learned on SRT tasks (Keisler, Ashe & Willingham, 2007).

Extensive research has accumulated suggesting that sleep is important for processes of memory consolidation in adults. In children and infants, a greater capacity to learn and to memorise coincides with longer and deeper sleep (see Wilhelm et al., in press, for a review). This study provides the first evidence that the transformation in memory that occurs during sleep has the consequence of enhancing new representations and integrating these representations into the mental lexicons of children without additional training. Lexical integration, an essential process in language development, gives rise to automaticity in word recognition whilst sustaining existing knowledge. The crucial effects were striking and reliable and suggest that consolidation plays a key role in learning new phonological forms. Sleep appears to be instrumental in this process in children and supports the consolidation of new lexical information in memory thereby generating stable and long-lasting memory representations. Our data provide the first evidence that a dual memory system works to protect the developing lexicon from “catastrophic interference” (the phenomenon of new information from overwriting existing information) similar to evidence from adults. Many previous developmental studies of vocabulary learning have focussed on the immediate consequences of word learning and have characterised word learning as a “relatively simple affair” (Plunkett & Wood, 2005, p. 165). Studies have claimed that young children can create novel entries in the lexicon for a newly heard word given only a 3-second exposure to that word and its referent (Halberda & Goldman, submitted). In contrast, our results suggest that this ability, called “fast mapping” (Carey & Bartlett, 1978), should be considered as the start of a prolonged consolidation process and point towards a major theoretical shift in our understanding of language development. In addition to these theoretical implications, our results may have pedagogical



ramifications since they implicate sleep as an integral part of learning and development and highlight the remarkable similarity between word learning in children and adults.

Future work will be needed to determine the influence of different types of training on the time course of word learning across development (e.g., semantic versus phonological, explicit versus implicit). It will also be important to establish whether the same effects can be obtained for the acquisition of real rather than fictitious novel nonwords (e.g., science words; Henderson, Weighall, & Gaskell, submitted), and when using more naturalistic training techniques (e.g., using storybooks, videos). Additionally, the present findings need to be validated with evidence concerning the precise aspects of sleep that are involved (e.g. sleep stages and spindles). Such future studies will go further in strengthening our case that the complementary learning systems framework can provide a strong account of vocabulary acquisition across development and open up new dimensions of research. Developmental studies of sleep and word learning represent a highly promising approach for furthering our understanding of the neurobiological mechanisms underpinning memory consolidation during sleep.

## **Acknowledgements**

This research was funded by the Leverhulme Trust (grant F/00 224/AO). We would like to thank all of the children who took part in this study and the parents and schools for making this possible.

## References

- Backhaus, J., Hoecksfeld, R., Born, J., Hohagen, F., & Junghanns, K. (2008). Immediate as well as delayed post learning sleep but not wakefulness enhances declarative memory consolidation in children. *Neurobiology of Learning and Memory, 89*(1), 76-80.
- Brown, H., Weighall, A., Henderson, L.M., & Gaskell, M.G. (2012). Enhanced recognition and recall of new words in 7- and 12-year old children following a period of offline consolidation. *Journal of Experimental Child Psychology, 112*, 56-72.
- Campbell, I.G., & Feinberg, I. (2009). Longitudinal trajectories of non-rapid eye movement delta and theta EEG as indicators of adolescent brain maturation. *Proceedings of the National Academy of Science, USA, 1096*, 5177 – 5180.
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference, 15*, 17–29.
- Church, B. A., & Fischer, C. (1998). Long-term auditory word priming in preschoolers: Implicit memory support for language acquisition. *Journal of Memory and Language, 39*, 523-542.
- Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21*, 803-820.
- Davis, M.H., & Gaskell, M.G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society B, 364*, 3773-3800.
- Diekelmann, S. & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11*, 114-126.
- Deregnaucourt, S., Mitra, P.P., Feher, O., Pytte, C., & Tchernichovski, O. (2005). How sleep affects the developmental learning of bird song. *Nature, 433*, 710 – 716.
- Dockrell, J. E., Braisby, N., & Best, R. M. (2007). Children's acquisition of science terms: Simple exposure is insufficient. *Learning and Instruction, 17*(6), 577-594.
- Dorfberger, S., Adi-Japha, E., & Karni, A. (2007). Reduced susceptibility to interference in the consolidation of motor memory before adolescence. *PLoS ONE, 2*(2), e240.
- Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*(1), 35-39.
- Dunn, L.M. & Dunn, D.M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition*. Pearson Education.

- Ellenbogen, J.M., Payne, J. D., & Stickgold, R. (2006). The role of sleep in declarative memory consolidation: passive, permissive, active or none? *Current Opinions in Neurobiology*, *16*, 716-722.
- Femia, L.A., & Hasselmo, M.E. (2002). Is autism partly a consolidation disorder? *Behaviour and Cognition Neuroscience Review*, *1* (4), 251 – 263.
- Fenn, K.M., Nusbaum, H.C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature*, *425*, 614–616.
- Fischer, S., Wilhelm, I., & Born, J. (2007). Developmental differences in sleep's role for implicit off-line learning: Comparing children with adults. *Journal of Cognitive Neuroscience*, *19*, 214-227.
- French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, *3*(4), 128-135.
- Forster, J. C., & Forster, K. I. (2003). DMDX: A Windows display program with millisecond accuracy. *Behaviour Research Methods, Instruments, & Computers*, *35*, 116-124.
- Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*, 105-132.
- Gomez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps Promote Abstraction in Language-Learning Infants. *Psychological Science*, *17*, 670-67.
- Halberda, J. & Goldman, J. (submitted). One-trial learning in 2-year-olds: Children learn new nouns in 3 seconds flat.
- Henderson, L.M., Weighall, A., Brown, H. & Gaskell, M. G. (submitted). On-line lexical competition during spoken word recognition and word learning in children and adults.
- Henderson, L.M., Weighall, A., & Gaskell, M.G. (submitted). Learning new vocabulary in childhood: Effects of semantic training on consolidation.
- Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sounds patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1-23.
- Keisler, A., Ashe, A., & Willingham, D.T. (2007). Time of day accounts for overnight improvement in sequence learning. *Learning and Memory*, *14*, 669 – 672.
- Kurth, S., Jenni, O.G., Riedner, B.A., Tononi, G., Carskadon, M.A., Huber, R. (2010). Characteristics of sleep slow waves in children and adolescents. *Sleep*, *33*, 475 – 480.

- Malow, B.A., Crowe, C., Henderson, L., McGrew, S.G., Wang, L., Song, Y., & Stone, W.L. (2009). A sleep habits questionnaire for children with autism spectrum disorders. *Journal of Child Neurology, 24* (1), 19 – 24.
- Marshall, I., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Science, 11*, 442 – 450.
- Marshall, I., Helgadottir, H., Molle, M., & Born, J. (2006). Boosting slow oscillations during sleep potentiates memory. *Nature, 444*, 610 – 613.
- Mattys, S. L., & Clark, J. H. (2002). Lexical activity in speech processing: Evidence from pause detection. *Journal of Memory and Language, 47*(3), 343-359.
- 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*(3), 419-457.
- Meulemans, T., Van der Linden, M., & Perruchet, P. (1998). Implicit sequence learning in children. *Journal of Experimental Child Psychology, 69*, 199 – 221.
- Metsala, J. L., Stavrinos, D., & Walley, A. C. (2009). Children's spoken word recognition and contributions to phonological awareness and nonword repetition: A 1-year follow-up. *Applied Psycholinguistics, 30*, 101-121.
- Molle, M., Marshall, I., Gais, S., & Born, K. (2002). Grouping of spindle activity during slow oscillations in human non-rapid eye movement sleep. *Journal of Neuroscience, 22*, 10941-10947.
- Munson, B., Swenson, C. L., & Manthei, S. C. (2005). Lexical and phonological organization in children: Evidence from repetition tasks. *Journal of Speech Language and Hearing Research, 48*(1), 108-124.
- Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review, 110*(4), 611-646.
- Ojima, S., Matsuba-Kurita, H., Nakamura, N., & Hagiwara, H. (2011). The acceleration of spoken-word processing in children's native-language acquisition: an ERP cohort study. *Neuropsychologia, 49*(5), 790-799.
- O'Reilly, R.C. (2006). Biologically Based Computational Models of High-Level Cognition. *Science, 314*, 91-94.

- 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(12), 505-510.
- Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, 95, 201 – 236.
- Plunkett, K., & Wood, C. (2003). *The development of children's understanding of grammar*. In J. Oates, & A. Grayson (Eds.), *Cognitive and language development in children*. Oxford: Blackwell.
- Rasch, B., Buchel, C., Gais, S., & Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science*, 315, 1426 – 1429.
- Robins, A., & McCallum, S. (1999). The consolidation of learning during sleep: comparing the pseudorehearsal and unlearning accounts. *Neural Networks*, 12, 1191 – 1206.
- Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology*, 24, 755-789.
- Sirota, a., Csicsvari, J., Buhl, D., Buzsaki, G. (2003). Communication between neocortex and hippocampus during sleep in rodents. *Proceedings of the National Academy of Science USA*, 100, 2065-2069.
- Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech Language and Hearing Research*, 44(6), 1321-1337.
- Storkel, H. L., & Lee, S. Y. (2011). The independent effects of phonotactic probability and neighbourhood density on lexical acquisition by preschool children. *Language and Cognitive Processes*, 26(2), 191-211.
- 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. *Journal of Neuroscience*, 30(43), 14356-14360.
- Torgesen, J., Wagner & Rashotte (1999). *Test of Word Reading Efficiency*. London: The Psychological Corporation.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing (CTOPP)*. Austin, TX: PRO-ED
- Walker, M.P. (2009). The role of sleep in cognition and emotion. *NYAS*; 1156:168-197.
- Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: Harcourt Assessment.

Wilhelm, I., Diekelmann, S., & Born, J. (2008). Sleep in children improves memory performance on declarative but not procedural tasks. *Learning & Memory, 15*(5), 373-377.

Wilhelm, I., Prehn-Kristensen, A., & Born, J. (in press). Sleep-dependent memory consolidation – What can be learnt from children? *Neuroscience and Biobehavioural Reviews*.

Wilson, M. (1988). MRC Psycholinguistic Database – machine-usable dictionary, version 2.00. *Behaviour Research Methods Instruments & Computers, 20*, 6 – 10.

## Footnotes

<sup>1</sup>It was not always possible to ensure the time elapsing between the 0-hr and 12-hr retests was precisely 12 hours due to the difficulties of testing children early in the morning and late at night (i.e., many children had a bedtime of around 19:00). Consequently, the time elapsing between the 0-hr and ~12-hr retests was significantly longer for the p.m. group than the a.m. group ( $p<.001$ ) and the time elapsing between the ~12-hr and 24-hr retests was longer for the a.m. group than the p.m. group ( $p<.001$ ). To guard against this potential confound correlations are reported between the time elapsing between sessions and the improvement scores between sessions for each group (see Results section).

<sup>2</sup> The a.m. group did not show a significant difference between random and sequenced trials in block 1 (mean difference -2ms,  $SD=68ms$ ,  $p>.05$ ), but showed slower responses to random than sequenced trials for the remaining four blocks: 2 (mean difference 38ms,  $SD=94ms$ ,  $p<.05$ ), 3 (35ms,  $SD=59ms$ ,  $p<.01$ ), 4 (28ms,  $SD=74ms$ ,  $p=.06$ ), 5 (51ms,  $SD=58ms$ ,  $p<.001$ ). In contrast, the p.m. group showed significantly faster responses to sequenced than random trials for blocks 1 (29ms,  $SD=43ms$ ,  $p<.01$ ), 2 (61ms,  $SD=93ms$ ,  $p<.01$ ), and block 3 (49ms,  $SD=49ms$ ,  $p<.001$ ) but did not show a significant difference between random and sequenced conditions for block 4 (15ms,  $SD=66ms$ ,  $p>.05$ ) or 5 (19ms,  $SD=92ms$ ,  $p>.05$ ).



## Appendix 1

Correlations on improvement scores from 0-hrs – 12-hrs and 12-hrs – 24-hrs for the a.m. group and p.m. group. *Note.* Correlations were largely weak none of them survived Bonferonni correction. Furthermore, there were no significant correlations between improvements on any measure and chronological age.

<i>0-hrs – 12-hrs</i>					
	Cued Recall	2AFC	Lexical Comp	Declarative	Procedural
2AFC	.26 / .19				
Lexical Comp	-.19 / .04	-.14 / -.22			
Declarative	.01 / -.14	.20 / .18	-.07 / -.10		
Procedural	.06 / .59	.31 / .06	-.50 / .03	-.08 / .09	
Age	-.38 / .01	-.22 / -.21	.45 / .09	-.24 / -.21	-.20 / -.23
Nonword	.37 / .08	.37 / -.08	-.11 / -.31	.44 / -.09	.02 / .27
Word Reading	.34 / .11	.32 / -.05	-.16 / -.09	.39 / .11	.07 / .44
NVIQ	.22 / -.13	.33 / -.37	.18 / -.23	.20 / .02	.17 / -.31
Phoneme Elis.	.04 / .12	.32 / .10	.22 / -.35	.37 / .22	-.06 / .30
Recept. Vocab	-.15 / -.15	.03 / -.20	.37 / -.04	-.23 / -.01	.01 / -.27
<i>12-hrs – 24-hrs</i>					
2AFC	-.05 / -.40				
Lexical Comp	.07 / .37	.05 / -.10			
Declarative	.04 / -.30	.11 / .24	.12 / -.13		
Procedural	-.11 / -.01	.06 / -.43	-.30 / .13	.08 / .26	
Age	.08 / -.06	-.20 / .26	-.36 / .13	-.09 / -.01	-.02 / -.04
Nonword	-.21 / -.04	-.14 / -.07	.02 / .10	.03 / -.01	.07 / .14
Word	-.18 / -.03	.01 / -.18	.21 / .02	-.01 / -.05	.09 / .18
NVIQ	.16 / -.05	-.15 / .36	-.42 / .13	-.22 / .16	.16 / .13
Phoneme Elis.	-.21 / -.12	-.24 / .11	-.13 / .14	-.07 / .03	.06 / .05
Vocabulary	.26 / .01	-.15 / .13	-.48 / .18	.08 / .04	.07 / .12

## Appendix 2

Existing Word	Novel Competitor	Novel Foil
LIST 1		
apricot	aprickel	apricken
baboon	babeel	babeen
blossom	blossail	blossain
bracelet	bracelop	braceliff
bramble	brambooce	bramboof
caravan	caravoth	caravol
crocodile	crocodiss	crocodin
dolphin	dolp heg	dolphess
fountain	fountel	fouted
lantern	lantobe	lanto ke
mermaid	mermiff	mermod
octopus	octopoth	octopol
parade	parafe	parane
potato	potabu	potago
skeleton	skeletobe	skeleto pe
walnut	walnog	walnep
LIST 2		
biscuit	biscal	biscan
breakfast	breakfal	breakfem
cardigan	cardigite	cardigile
costume	costuke	costupe
daffodil	daffadat	daffadan
dungeon	dungeill	dungeic
napkin	napkem	napkess
onion	oniot	onioff
parachute	parasheff	parashen
pyramid	pyramon	pyramotch
sardine	sardiss	sardike
squirrel	squirrome	squirrope
tulip	tulode	tulome
volcano	volcater	volcador
walrus	walrick	walreg
yoghurt	yogem	yogell

## Figure Legends

Figure 1. Schematic representation of the experiment.

Figure 2. Lexical competition effects ( $=$  competitor RT – control RT) for the a.m. and p.m. groups at 0-hr, ~12-hr, 24-hr and 1-wk retests (error bars show SEM). Lexical competition was observed after ~12 hours but only for the p.m. group who had slept prior to the ~12-hr retest.

Figure 3. Mean percent correct on explicit tests at 0-hr, ~12-hr, 24-hr and 1-wk retests, for the a.m. and p.m. groups (error bars show SEM). Significant gains in recognition and recall were observed after ~12 hours but only for the p.m. group who had slept prior to the ~12-hr retest.

Figure 4. Proportion of objects located on the declarative memory task as a function of session and group (error bars show SEM). There were clear sleep effects: the a.m. group showed significantly lower accuracy at ~12-hrs than 0-hrs (after being awake) whereas the p.m. group showed no such reduction in performance (after sleep).

Figure 5. Response time for random and sequenced conditions on the five training blocks of the SRT task for the a.m. group (upper panel) and p.m. group (lower panel) (error bars show SEM).

Figure 6. Response time for random and sequenced conditions on the SRT task at the 0-hr (averaged across blocks 3-5 of training), ~12-hr, 24-hr and 1-wk retests for the a.m. (upper panel) and p.m. (lower panel) groups (error bars show SEM). Both groups showed clear evidence of procedural learning across the week of the experiment reflected by an increase in RT and a progressively larger difference between sequenced and random conditions. However, there was no evidence that sleep enhanced this effect.

## Table Legends

Table 1. Descriptive statistics for age, cognitive, and language skills. *Note.* ss = standard score (mean 100, normal range 85 – 115), T = T score (mean 50, normal range 40 – 60), sc = scaled score (mean 10, normal range 8 – 12). Standardised tests used: <sup>1</sup> *Test of Word Reading Efficiency* (Torgesen, Wagner & Rashotte, 1999), <sup>2</sup> *Matrix Reasoning* from the Wechsler Abbreviated Scales of Intelligence (Wechsler, 1999), <sup>3</sup> *Phoneme Elision* and *Memory for Digits* (from the Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999), <sup>4</sup> Peabody Picture Vocabulary Test 4<sup>TH</sup> Edition (Dunn & Dunn, 2007).

Table 2. Mean (SD) time elapsing between sessions for the a.m. and p.m. groups.

Table 3. Pause detection latencies, % errors and planned contrasts for competitor and control conditions (SDs in parentheses).