Generalisation in language learning can withstand total sleep deprivation

Jakke Tamminen*, Chloe R. Newbury, Rebecca Crowley, Lydia Vinals, Benedetta Cevoli, Kathleen Rastle

Department of Psychology, Royal Holloway, University of London, Egham, TW20 0EX, United Kingdom

ABSTRACT

Research suggests that sleep plays a vital role in memory. We tested the impact of total sleep deprivation on adults' memory for a newly learned writing system and on their ability to generalise this knowledge to read untrained novel words. We trained participants to read fictitious words printed in a novel artificial orthography, while depriving them of sleep the night after learning (Experiment 1) or the night before learning (Experiment 2). Following two nights of recovery sleep, and again 10 days later, participants were tested on trained words and untrained words, and performance was compared to control groups who had not undergone sleep deprivation. Participants showed a high degree of accuracy in learning the trained words and in generalising their knowledge to untrained words. There was little evidence of impact of sleep deprivation on memory or generalisation. These data support emerging theories which suggest sleep-associated memory consolidation can be accelerated or entirely bypassed under certain conditions, and that such conditions also facilitate generalisation.

1. Introduction

Increasing evidence suggests that sleep benefits memory (Diekelmann & Born, 2010). Two broad mechanisms are thought to give rise to this benefit. Firstly, according to the active consolidation theory, sleep *after* learning strengthens new declarative memories through re-activation of the neuronal traces of those memories (McClelland, McNaughton, & O'Reilly, 1995). This reactivation originates in the hippocampus, a region of the brain thought to encode new memories rapidly and to represent them separately from each other and existing memory stores. It is through repeated reactivation, primarily during slow-wave sleep (SWS), that new memories become represented in the neocortex, a region of the brain that allows integration of new memories with existing knowledge as well as representation of similarities across newly encoded memories.

According to a recent extension to the active consolidation theory (Saletin & Walker, 2012), sleep *before* learning also supports hippocampal-dependent declarative encoding capacity. Research suggests that one night of total sleep deprivation results in decreased activation of the bilateral posterior hippocampus while encoding episodic memory (Yoo, Hu, Gujar, Jolesz, & Walker, 2007). Further, even a brief daytime nap restores encoding capacity for declarative memories, with the magnitude of this restoration associated with specific sleep parameters (stage 2 sleep and sleep spindle activity; Mander, Santhanam, Saletin, &

Walker, 2011). Mander et al. (2011) argued that one implication of the sparse representation of new memories in the hippocampus is that encoding capacity reduces as a result of learning over the course of a day. Sleep is needed to allow new memories to become independent of the hippocampus, thus restoring hippocampal encoding capacity. Without this restoration, subsequent learning would be impaired. The synaptic homeostasis hypothesis (Tononi & Cirelli, 2014) also emphasises the role of sleep before learning in restoring declarative memory encoding capacity. According to this theory learning is instantiated in the brain through increased synaptic strength. However, as synaptic strength increases, neurons become less selective in discriminating between inputs, leading to a gradual reduction in encoding ability over the course of a day of wake. During subsequent SWS those synapses that were not consistently activated during learning weaken, thus allowing the restoration of encoding ability and optimising later learning.

Taken together, the theories reviewed above point to a symbiotic role for sleep both before and after learning. Sleep after learning consolidates hippocampal-dependent *recently encoded memories,* while sleep before learning prepares the hippocampus for encoding of *future memories.* These mechanisms apply to the types of learning that depend on the medial temporal lobe (MTL, which includes the hippocampus) memory systems, that is, declarative memory (sometimes referred to as explicit memory). Declarative memory is considered to consist of knowledge that is available to conscious recall, such as memory for

facts and events, while nondeclarative memory (sometimes referred to as implicit memory) consists of knowledge that is not necessarily available to conscious recall but is expressed through performance (see Squire & Dede, 2015, for a review). Examples of nondeclarative memory include classical conditioning, habit learning, perceptual learning, and priming. These forms of learning do not depend on the medial temporal lobe structures and are often partially or wholly preserved in amnesic patients with extensive damage to the MTL (Squire & Dede, 2015).

The broad consolidation theory outlined earlier makes clear predictions about the detrimental behavioural consequences of sleep deprivation. Sleep deprivation after learning should disrupt declarative memory consolidation, resulting in poorer memory performance compared to a night of undisturbed sleep. Several studies of episodic memory support this prediction. For example, Gais et al. (2007) taught participants word-pairs in the evening, followed by one night of sleep deprivation, or undisturbed sleep in the control condition. Participants were tested 48 h later, allowing for a night of recovery sleep for the sleep deprived participants. Sleep deprived participants were found to have forgotten significantly more word-pairs than the controls (see also Gais, Lucas, & Born, 2006, for similar results). Research suggests that sleep deprivation before learning has similar consequences. Yoo et al. (2007) deprived participants of sleep for one night before they studied a set of pictures. In a recognition memory test following two nights of recovery sleep, these sleep deprived participants showed significantly poorer memory for the pictures than control participants who had not been sleep deprived before learning. Mander et al. (2011) had participants encode face-name pairs where the encoding was preceded by a daytime nap or wake. Napping before encoding benefited memory for the face-name pairs but did not benefit nondeclarative memory in a motor skill learning task (see also Drummond et al., 2000 for similar results).

While these studies suggest that sleep deprivation before and after memory encoding has a significant detrimental effect on episodic memory, little is known about its impact on learning of new semantic knowledge, and other forms of declarative memory. Previous work has shown that generalisation (the extraction of general knowledge from individual learning episodes) benefits from offline memory consolidation (Tamminen, Davis, Merkx, & Rastle, 2012; Tamminen, Davis, & Rastle, 2015). This form of learning is critically important for our ability to apply previously encoded information in new situations. Participants who were asked to learn the meanings of new affixes attached to familiar words (e.g., *sleepnule is a participant in a sleep experiment*) exhibited semantic priming of the new affixes (e.g., *nule*) only after a 24-hour consolidation opportunity when those affixes were attached to a new, untrained word context during a generalisation test (e.g., *sailnule*; Tamminen et al., 2015). The consolidation effects observed in these studies could reflect consolidation during either sleep or wake, or a combination of both. Other studies looking at sleep and generalisation in language learning have sought to isolate the role of sleep. Most of this work has concentrated on grammar learning and has suggested that sleep facilitates the extraction of abstract grammar rules from exposure to an artificial language (Gomez, Bootzin, & Nadel, 2006; Nieuwenhuis, Folia, Forkstam, Jensen, & Petersson, 2013; Batterink, Oudiette, Reber, & Paller, 2014; Mirković & Gaskell, 2016; Batterink & Paller, 2017).

While our ability to generalise is fundamental to successful grammar learning, other domains of language learning also rely on generalisation. Learning to read is one such domain. Learning to read requires the learner to encode mappings between letters and sounds. Importantly, these mappings need to be generalised to allow the skilled reader to read previously unseen words. The role of sleep in acquiring these letter-sound mappings is unknown. In order to understand the different ways sleep might facilitate generalisation in this type of learning we trained participants to read novel words printed in an artificial orthography, and deprived them of one night of sleep either immediately after training (Experiment 1) or immediately before training (Experiment 2). Participants were tested following two nights of recovery sleep (to eliminate any confounds due to fatigue at the time of testing) and then again 10 days later. We compared performance with controls who were not deprived of sleep. Critically, we tested participants not only on the novel words they were trained on but also on untrained words they had never encountered before in order to assess their ability to generalise what they had learned about the artificial orthography.

At the heart of learning to read and of our artificial orthography task is the ability to learn the arbitrary pairing between a symbol and a sound, and to express this pairing explicitly when asked to read or spell words. Our task therefore should yield declarative knowledge (and potentially also some degree of nondeclarative knowledge) and involve a role for sleep-associated consolidation, as predicted by the theories reviewed earlier. Following the active consolidation theory, we predicted that sleep deprivation after learning (Experiment 1) should disrupt participants' declarative memory for the symbol-sound mappings as well as their ability to generalise these mappings to read previously unseen words. At training participants read and spelled new words printed in the new orthography. At test they read and spelled these words again, as well as new, untrained words to test generalisation. As using the same task at encoding and retrieval is known to facilitate retrieval due to the same cues being present at both times (encoding specificity principle; Thomson & Tulving, 1970) we added another task as a stronger test of generalisation. In the phoneme knowledge task participants indicated the sound associated with symbols shown in isolation. We expected both tasks to primarily (although possibly not exclusively) tap into hippocampally-dependent declarative memory, as accurate responding requires access to newly formed arbitrary associations between novel symbols and sounds, particularly in the phoneme knowledge task which the participants have not experienced during training. Finally, we also included an old-new test of the trained words to assess not only recall but also recognition memory of the trained words.

Turning to the potential role of sleep in restoring learning capacity after extended wake (Experiment 2), we predicted that sleep deprivation before learning should make learning more effortful in the training session and might not lead to sufficiently strong memory to allow for normal generalisation. We therefore hypothesised that compared to controls, sleep deprived participants would need more training trials to reach a pre-specified learning criterion, would show poorer reading and spelling of untrained words at test, and poorer knowledge of the sound-symbol mappings of the new orthography. If sleep deprivation before learning also leads to accelerated forgetting, we expected poorer recognition memory of the trained words, as well as poorer reading and spelling of those words at test.

2. Experiment 1: Sleep deprivation after learning

2.1. Methods

2.1.1. Participants

47 healthy monolingual native speakers of British English (13 males), aged 18 to 35, took part in Experiment 1 (23 in sleep deprivation, 24 in control group). To avoid experimenter bias in allocating participants to conditions while at the same time allowing for fully informed consent, participants signed up to one of the two groups, but the two groups were recruited at different times to prevent participants choosing which condition to volunteer for. None of the participants had visual or hearing impairments, or known sleep, neurological, or psychiatric disorders, and none were taking medication influencing sleep or had travelled across time zones in the past four weeks. Participants provided their written informed consent and were paid to take part in the study. In choosing the sample size we were guided by the sample sizes reported in published sleep deprivation studies. The sample sizes

Table 1
Sample characteristics in Experiment 1. Values indicate means, with standard error in parentheses.

	Sleep deprivation group	Rested control group	Difference between groups
Digit span	6.61 (0.20)	6.13 (0.24)	p = .12
Chronotype (total MEQ score)	43 (1.55)	45 (1.87)	p = .30
Spelling (proportion of words spelt correctly)	0.43 (0.04)	0.42 (0.04)	p = .82
Vocabulary (proportion of words recognised)	0.75 (0.01)	0.74 (0.02)	p = .68
Reading (reading efficiency index)	109 (0.18)	113 (0.14)	p = .34[†]
Sleep duration on the night before experiment (in hours)	6.89 (0.24)	6.85 (0.21)	p = .91
Sleep duration on the experimental night (in hours)	N/A	6.38 (0.29)	N/A
Sleep duration on the first recovery night (in hours)	8.18 (0.51)	6.09 (0.27)	p < .001
Sleep duration on the second recovery night (in hours)	6.33 (0.35)	7.04 (0.19)	p = .08

[†] Comparison used Kruskal-Wallis tests of differences between groups as the data are ordinal.

for both experiments were further motivated by taking the effect size reported in Lau, Alger, and Fishbein (2011) in a nap vs. no-nap manipulation using a language learning generalisation task, and entering this into a power analysis which resulted in an estimated 22 participants per group (when f = 0.38, α = 0.05, power = 0.80).

We collected a number of learning and language related background measures to characterise our sample and to check whether participants in the two conditions were closely matched. The measures included working memory capacity (forward visual digit span), chronotype (using the Morningness-Eveningness Questionnaire, Horne & Östberg, 1976), spelling to dictation accuracy (Burt & Tate, 2002), vocabulary knowledge (using the Vocabulary sub-scale of the Shipley Institute of Living Scale; Shipley, 1940), and reading fluency (using The Test of Word Reading Efficiency-Second Edition; Torgesen, Wagner, & Rashotte, 2012). Table 1 shows the descriptive statistics associated with each measure, and results of statistical comparison between the groups using independent-samples t-tests.

2.1.2. Materials

Six lists of 36 monosyllabic consonant–vowel-consonant pseudowords were created (e.g., "bav"). The six lists were rotated across the following uses: pseudowords used as training items, pseudowords used as untrained items for the old-new decision task, pseudowords used as untrained items to test reading aloud generalisation, and pseudowords used as untrained items to test spelling generalisation. Eighteen alphabetic symbols were selected from the Hungarian Runes archaic orthography and mapped 1:1 to 18 phonemes (Taylor, Davis, & Rastle, 2017). Fig. 1 shows examples of items used at training and at test.

2.1.3. Procedure

Participants were asked to have a normal night of sleep the night before the training session, and to refrain from napping or consuming caffeine or alcohol on the day of the training session (Fig. 2). Compliance with these sleep requirements was monitored by actigraphy until the first test session. On the day of the training session, participants came to the lab at 8 pm or 10 pm to complete the training session.
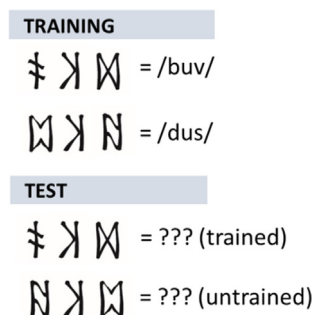

Fig. 1. Examples of training stimuli printed in the artificial orthography and their associated pronunciations, and examples of untrained stimuli used at test.

Participants in the control group then left the lab and slept at home. Participants in the sleep deprivation group stayed in the lab and were kept awake until they left the lab at 8am the following morning. During the night, these participants were continuously monitored, and asked to complete digit span and psychomotor vigilance tests (PVT; Dinges & Powell, 1985) every hour. Results from these two tasks are presented in the Supplementary Materials. Upon leaving the lab, participants were asked to remain awake for as long as possible and stick to their normal bedtime. Compliance in both groups was checked against actigraphy data and sleep diaries. Inspection of these data showed that most participants either napped during the day after the deprivation session or went to bed earlier than usual. However, correlational analyses showed that the amount of time that elapsed between leaving the lab and the first sleep period had no impact on test performance, and therefore no participants were excluded due to sleeping before normal bedtime. Participants in both groups returned to the lab for their first test session after two nights of sleep (three for the control group), and 10 days later for the second test session.

At the beginning of the training session, the orthographic and spoken forms of each of the 36 training items were presented once simultaneously. Participants then began active training blocks. Each block comprised a reading aloud task and a spelling task. In the reading aloud task, the orthographic forms of each of the training items were presented once in a randomised order. Participants were asked to read aloud the pronunciation in the new language, as quickly and accurately as possible. Participants then heard the correct pronunciation for the orthographic form before moving on to the next trial.

In the spelling task, participants heard the phonological form of each training item once, while viewing all alphabetic symbols. Participants selected the three symbols that spelled that training item in the correct order. After each trial, feedback was given, and the correct spelling was displayed. Participants completed a minimum of 4 training blocks consisting of both tasks, and continued up to a maximum of 8 blocks, until they reached a criterion of 70% accuracy on the spelling task. If the criterion was not reached after eight blocks, the participant was dismissed and replaced.

In the test sessions, participants completed four tasks in fixed order as follows: old-new decision task, reading aloud, spelling, phoneme knowledge. In the old-new decision task, the orthographic forms of 36 trained and 36 untrained items were presented in the centre of the screen. Participants were asked to categorise each as trained (old) or untrained (new). In the reading aloud task, participants were asked to read the 36 trained items as well as 72 untrained items. Participants had 9 s to make a response. In the spelling task, participants were asked to spell the 36 trained items as well as 72 untrained items (different from those used in the reading aloud task). There was no time limit in this task. Finally, in the phoneme knowledge task, each of the 18 alphabetic symbols was presented individually on the screen and participants were asked to produce the sound corresponding to each symbol. Participants had 9 s to make a response. No feedback was provided. Symbols were presented in a randomised order for each participant, and
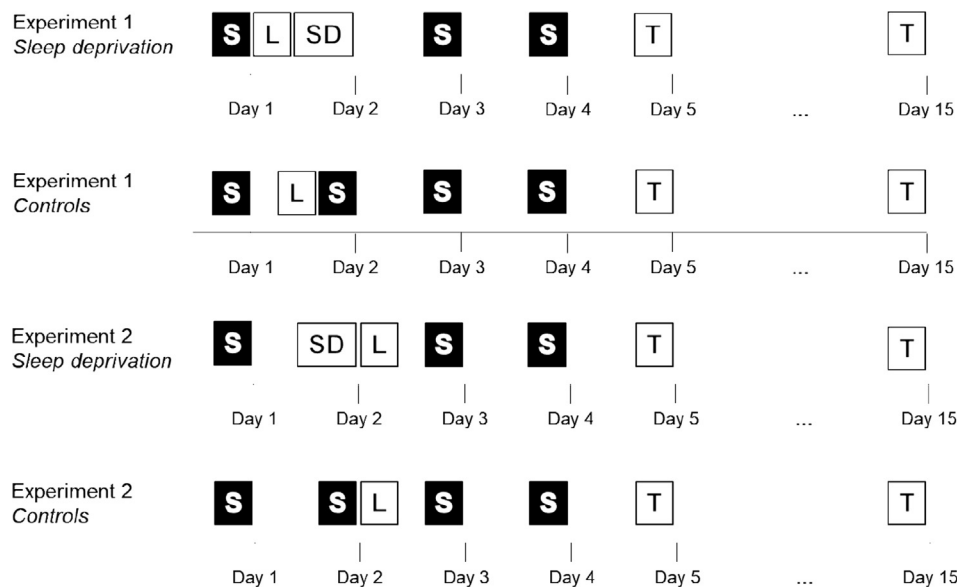
Fig. 2. Experimental design in Experiments 1 and 2. S = night of sleep, L = learning in the training session, SD = night of sleep deprivation, T = test session.

their vocal response recorded. In all tasks the order of stimuli was randomised.

2.2. Results

The stimuli, raw data, and data analysis files for both experiments are publicly available at https://osf.io/2kyrd/. Vocal responses were marked manually via visual inspection of the speech waveform using CheckVocal software (Protopapas, 2007). Response time (RT) data were analysed by fitting (generalised) linear mixed effects models to log or inverse-transformed data, depending on which resulted in a more normal distribution in each task. Accuracy data were analysed using logistic mixed effects regression models with a binomial distribution. Recognition performance on the old-new decision task was analysed by calculating signal detection scores (d') in order to account for response bias. The fixed effects in each model comprised group (sleep deprivation vs. control), time of testing (test 1 vs. test 2), and familiarity (trained stimuli vs. untrained stimuli), unless stated otherwise. Participants and items were entered as random effects, unless stated otherwise. We sought to fit maximal random effects structure in each model (Barr, Levy, Scheepers, & Tily, 2013). When the model failed to converge, the random effects structure was simplified until the model converged. The statistics associated with the fixed effects in the test tasks are presented in Table 2 (Experiment 1) and Table 4 (Experiment 2) rather than in the main text. When an interaction between fixed effects reached significance, results of subsequent contrasts breaking down the interaction are reported in the main text.

2.2.1. Training

The mean number of training blocks required to achieve the criterion level of > 70% accuracy on the spelling task was slightly higher in the control (mean = 4.58, SD = 1.19) than in the sleep deprivation group (mean = 4.04, SD = 0.20) (t = 18.24, p < .001). Mean accuracy on the spelling task in the last training block did not differ between groups (93% and 92% for deprivation and control groups, respectively) (z = −1.11, p = .267), nor did mean accuracy on the reading aloud task for participants' highest-scoring block (79% and 77% for deprivation and control groups, respectively) (z = −1.15, p = .252). These data suggest that the groups learned the materials equally well by the end of training.

2.2.2. Test tasks

Old-New Decision. One participant was removed due to scoring close to zero correct in the first test but close to ceiling in the second test, suggesting they may have pressed the wrong buttons in the first test. Accuracy was analysed using signal detection measures (d') to control for response bias. Fig. 3 shows the accuracy and RT data. RTs were analysed only for correct responses.

One-sample t-tests on d' scores indicated that performance was above chance (higher than zero) for both groups at both tests (all ps < 0.001). A mixed effects regression model on the d' values with group (sleep deprivation vs. control) and time of testing (test 1 vs. test 2) as fixed effects showed no significant effects of group or time of testing on accuracy. RT analysis with group (sleep deprivation vs. control), time of testing (test 1 vs. test 2), and familiarity (trained vs. untrained stimuli) as fixed effects revealed a significant main effect of time of testing and a significant main effect of familiarity. Responses were faster in the second test, and responses to trained words were faster than those to untrained words. No other effects or interactions reached significance.

Reading Aloud. One participant had a mean reading aloud accuracy score across Test 1 and Test 2 of just 2%, thus they were removed from the analyses. Accuracy (Fig. 4) was analysed using a mixed effects logistic regression model with group (sleep deprivation vs. control), time of testing (test 1 vs. test 2), and familiarity (trained vs. untrained stimuli) as fixed effects. Only the effect of familiarity was significant. Participants were more accurate in reading trained compared to untrained words. No other effects or interactions were significant. In the RT analysis on correct responses with the same fixed effects, the results again revealed a significant main effect of familiarity. Participants were overall faster at reading aloud trained compared to untrained items. The main effect of time of testing was also significant, with faster responding in the second test. No other effects reached significance.

Spelling. The accuracy analysis using a mixed effects logistic regression model with group (sleep deprivation vs. control), time of testing (test 1 vs. test 2), and familiarity (trained vs. untrained stimuli) as fixed effects revealed a significant main effect of familiarity. Trained words were spelled more accurately than untrained words (Fig. 5). The interaction between sleep group and test session was also significant. Subsequent multiple comparisons with a Bonferroni correction revealed that participants in the sleep deprivation group were more accurate at Test 2 than Test 1 (z = 3.13, p = .007), whereas the control group were

Table 2

Predictors of accuracy and response time for each of the four test tasks in Experiment 1. Accuracy in the old new decision task is calculated as d′.

Fixed effect	Accuracy*				Response Time			
	b	SE	t/z	p	b	SE	t	p
Old New Decision Task[†]								
Group	-0.27	0.20	−1.33	0.189	-0.22	0.11	−1.97	0.055
Time of testing	-0.08	0.12	-0.71	0.479	**-0.08**	**0.04**	**−2.30**	**0.026**
Training	–	–	–	–	**0.17**	**0.01**	**11.42**	**< 0.001**
Group × Time of testing	0.02	0.17	0.12	0.904	-0.003	0.07	-0.04	0.972
Group × Training	–	–	–	–	-0.01	0.03	-0.48	0.633
Time of testing × Training	–	–	–	–	-0.003	0.03	-0.12	0.902
Group × Time of testing × Training	–	–	–	–	0.02	0.06	0.43	0.667
Reading Aloud Task[†]								
Group	-0.43	0.42	−1.04	0.299	-0.05	0.07	-0.68	0.502
Time of testing	0.12	0.13	0.97	0.333	**-0.07**	**0.01**	**−5.42**	**< 0.001**
Training	**-0.24**	**0.06**	**−4.28**	**< 0.001**	**0.04**	**0.01**	**5.61**	**< 0.001**
Group × Time of testing	0.08	0.25	0.33	0.743	-0.02	0.03	-0.66	0.512
Group × Training	0.10	0.11	0.87	0.384	-0.03	0.02	−1.71	0.090
Time of testing × Training	0.06	0.11	0.56	0.578	-0.003	0.02	-0.16	0.873
Group × Time of testing × Training	-0.10	0.22	-0.47	0.638	-0.03	0.03	-0.95	0.345
Spelling Task[‡]								
Group	-0.64	0.40	−1.60	0.110	-0.01	0.01	−1.34	0.188
Time of testing	0.06	0.13	0.49	0.624	< -0.03	< 0.01	−1.82	0.075
Training	**-0.16**	**0.07**	**−2.25**	**0.024**	**< 0.01**	**< 0.01**	**3.02**	**0.003**
Group × Time of testing	**-0.50**	**0.25**	**−2.03**	**0.042**	< -0.01	< 0.01	-0.83	0.412
Group × Training	-0.01	0.14	-0.06	0.956	< -0.01	< 0.01	-0.46	0.647
Time of testing × Training	0.05	0.14	0.38	0.703	< 0.01	< 0.01	0.55	0.580
Group × Time of testing × Training	0.12	0.27	0.43	0.664	< 0.01	< 0.01	0.86	0.388
Phoneme Task[‡]								
Group	-0.58	0.58	−1.00	0.319	**-0.10**	**0.04**	**−2.51**	**0.016**
Time of testing	0.03	0.28	0.09	0.925	**-0.06**	**0.01**	**−4.11**	**< 0.001**
Training	–	–	–	–	–	–	–	–
Group × Time of testing	**−1.21**	**0.47**	**−2.59**	**0.010**	0.01	0.03	0.24	0.809
Group × Training	–	–	–	–	–	–	–	–
Time of testing × Training	–	–	–	–	–	–	–	–
Group × Time of testing × Training	–	–	–	–	–	–	–	–

[†] RT data were log transformed for the statistical analysis

[‡] RT data were inverse transformed for the statistical analysis

more accurate at Test 1 than Test 2, (z = − 4.12, p < .001), but there was no significant difference between the sleep deprivation and control group at either Test 1 (z = − 0.73, p = 1), or at Test 2 (z = − 2.34, p = .077). RT analysis on correct responses with the same fixed effects revealed a significant main effect of familiarity. Responses to trained words were faster than to untrained words. No other effects were significant.

Phoneme Knowledge. Analysis of accuracy rates using a mixed effects logistic regression model with group (sleep deprivation vs. control) and time of testing (test 1 vs. test 2) as fixed effects (Fig. 6) showed that the interaction between sleep group and time of testing was significant, reflecting the change from the first to the second test session going in

opposite directions in the two groups. Subsequent multiple comparisons with Bonferroni correction revealed that there was no significant difference between the control and deprivation groups at either Test 1, (z = − 0.04, p = 1), or Test 2, (z = − 1.77, p = .306). RT analyses on correct responses with the same fixed effects revealed a significant main effect of sleep group, with faster responses for the control group. There was also a significant effect of time of testing. Participants were faster to respond at Test 2 than at Test 1. No other effects reached significance.
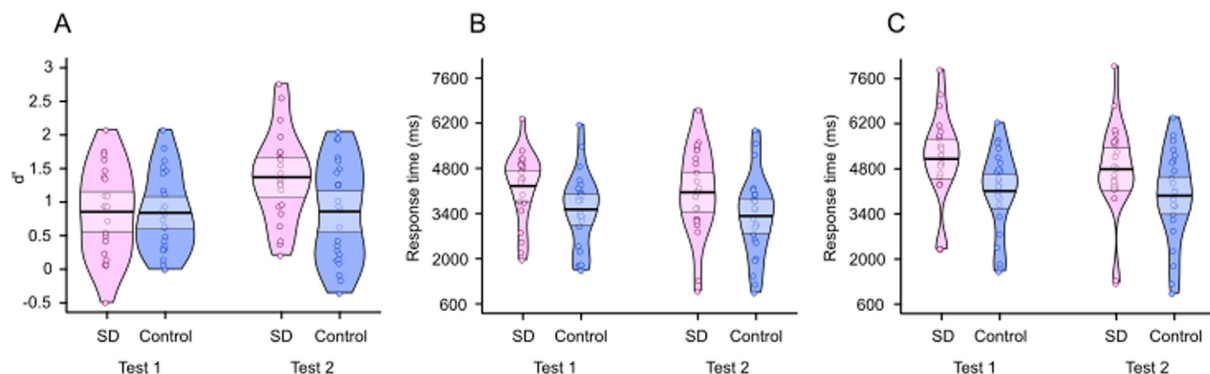

Fig. 3. Accuracy rates (A) and response times for trained (B) and untrained (C) words in the old-new decision task in Experiment 1. In this figure and all subsequent figures the horizontal line represents the mean, the box around the mean represents 95% confidence interval, the borders around data points are smoothed density curves.

Fig. 4. Accuracy rates for trained (A) and untrained (B) words, response times for trained (C) and untrained (D) words in the reading aloud task in Experiment 1. The horizontal line represents the mean, the box around the mean represents 95% confidence intervals, the borders around data points are smoothed density curves.
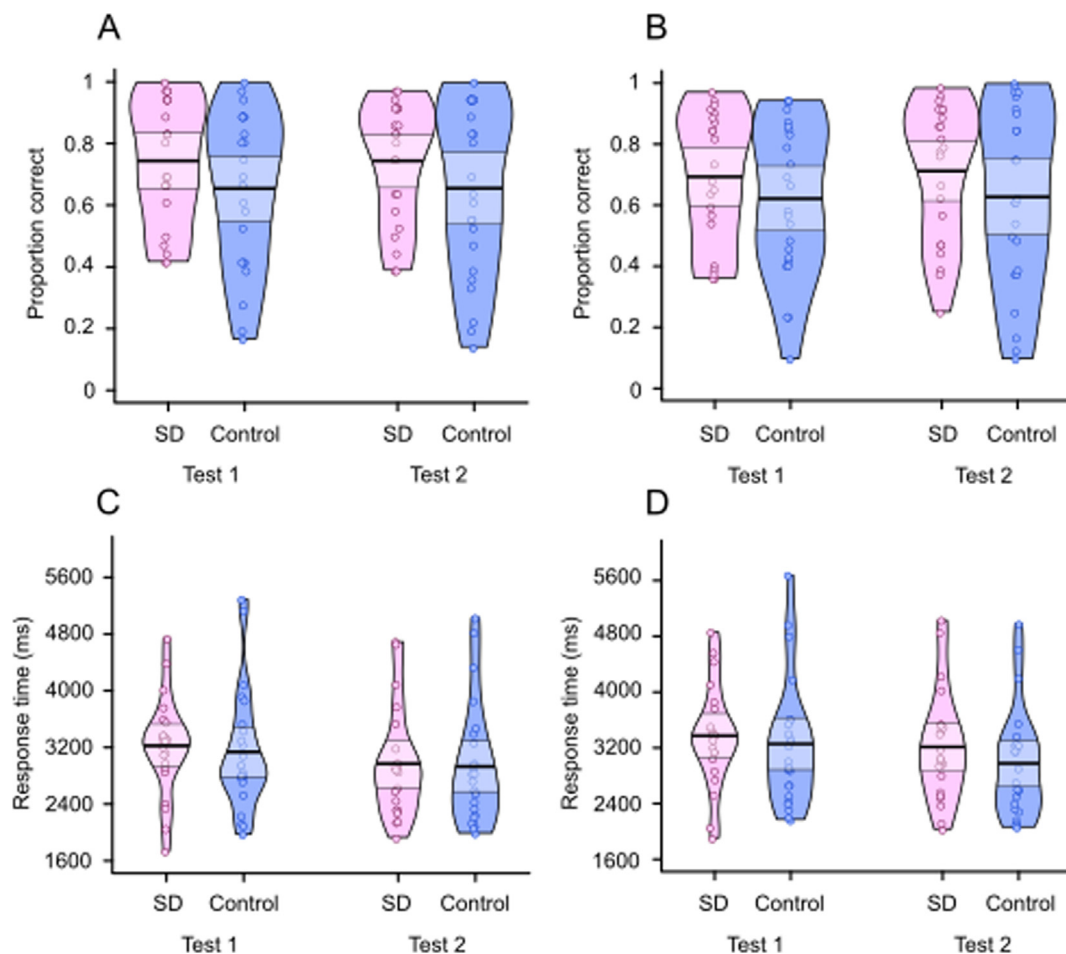
3. Experiment 2: Sleep deprivation before learning

3.1. Methods

3.1.1. Participants

46 native speakers of British English (12 males), aged 18 to 35, took part in Experiment 2 (24 in sleep deprivation, 22 in rested control group). The same inclusion and exclusion criteria were applied as in Experiment 1. The study was approved by the university's Research Ethics Committee. Participants provided their written informed consent and were paid to take part in the study. Sample characteristics are shown in Table 3.

3.1.2. Design and procedure

All tasks and procedures were exactly the same as in Experiment 1. However, participants carried out the training session at 8am after a night of normal sleep at home, or after a night of monitored sleep deprivation in the laboratory (Fig. 2). Upon leaving the lab, participants were asked to remain awake for as long as possible and stick to their normal bedtime. Compliance in both groups was checked against actigraphy data and sleep diaries. As in Experiment 1 inspection of these data showed that most participants either napped during the day after the deprivation session, or went to bed earlier than usual but correlational analyses showed that the amount of time that elapsed between leaving the lab and the first sleep period had no consistent statistically significant impact on test performance.

3.2. Results

3.2.1. Training

In the sleep control group, all participants reached the criterion level of > 70% after the first four blocks, while participants in the sleep deprivation group required on average 4.62 blocks (SD = 1.15) to meet the criterion, with 7 participants needing > 4 blocks to reach criterion. This difference was significant ($t = -21.66, p < .001$). Mean accuracy on the spelling task in the last training block and on the reading aloud task for the block on which participants received the highest score were calculated. Although both groups exceeded the criterion, there was a significant difference in accuracy scores between the sleep deprivation and the control group on the reading aloud task (75% vs. 85%; $z = 5.57, p < .001$). There was also a significant difference in accuracy scores in the spelling task (89% vs. 95%; $z = 4.25, p < .001$). In both instances the sleep control group performed significantly more accurately than the sleep deprivation group.

3.2.2. Test

Old-New decision. One participant was removed due to scoring zero correct on untrained words in the first test suggesting they were not engaging with the task and instead pressing "old" on each trial. One-sample t-tests on d' scores indicated that performance was above chance for both groups at both tests (Fig. 7; all ps < 0.001). A mixed effects regression model on the d' values with group (sleep deprivation vs. control) and time of testing (test 1 vs. test 2) as fixed effects showed no significant main effects or interactions. Here and in the other tasks p-values for the fixed effects and associated statistics are presented in
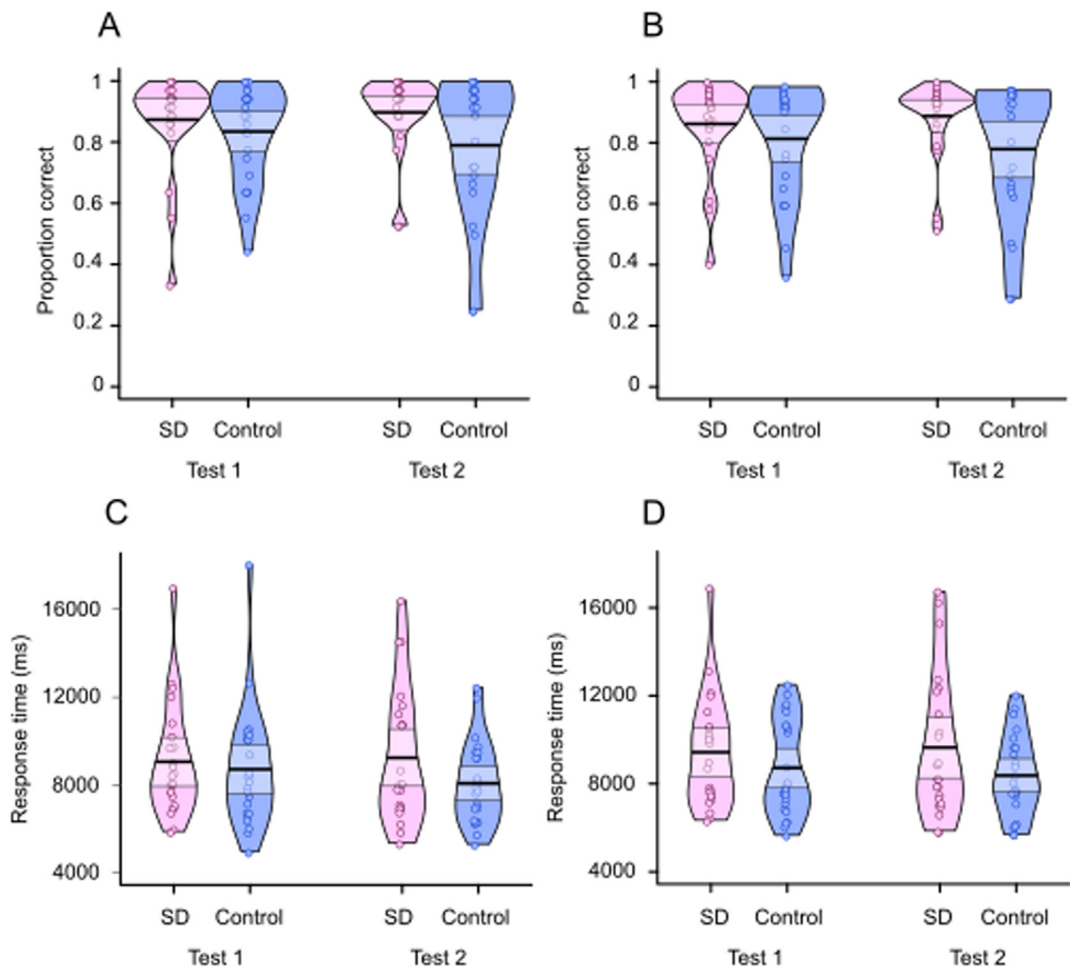
Fig. 5. Accuracy rates for trained (A) and untrained (B) words, response times for trained (C) and untrained (D) words in the spelling task in Experiment 1. The horizontal line represents the mean, the box around the mean represents 95% confidence intervals, the borders around data points are smoothed density curves.

Table 4. RT data on correct responses were analysed by fitting a model with group (sleep deprivation vs. control), time of testing (test 1 vs. test 2), and familiarity (trained vs. untrained stimuli) as fixed effects. This revealed a significant effect of testing session. Participants were faster to respond in test 2 than in test 1. There was also a significant effect of familiarity, with responses to trained words faster than those to untrained words. No other effects reached significance.

Reading Aloud. Accuracy rates are presented in Fig. 8. A mixed

effects logistic regression model with group (sleep deprivation vs. control), time of testing (test 1 vs. test 2), and familiarity (trained vs. untrained stimuli) as fixed effects showed that only the effect of familiarity was significant with more errors made to untrained words. No other significant main effects or interactions were found. RT results on correct responses using the same fixed factors revealed a significant effect of familiarity. Participants were overall faster at reading aloud trained compared to untrained items. The main effect of time of testing
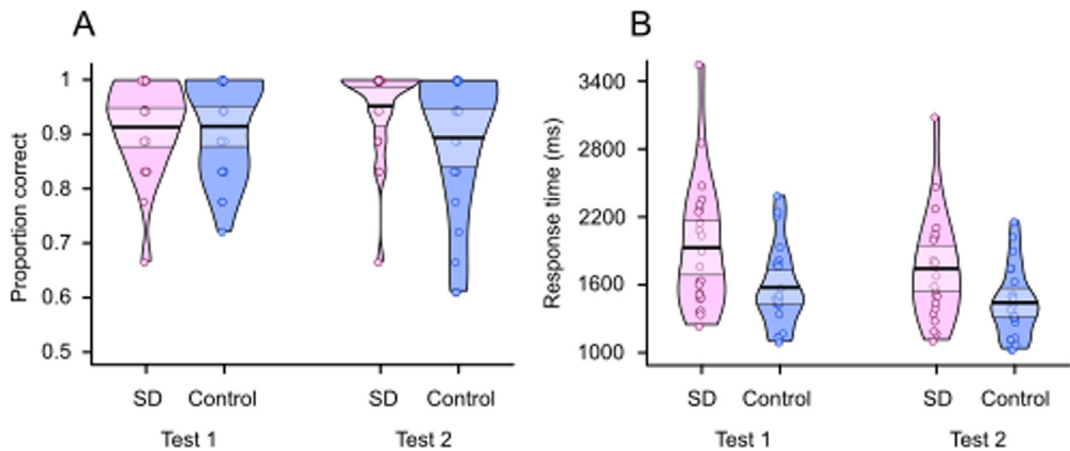

Fig. 6. Accuracy rates (A) and response times (B) in the phoneme knowledge task in Experiment 1. The horizontal line represents the mean, the box around the mean represents 95% confidence intervals, the borders around data points are smoothed density curves.

Table 3
Sample characteristics in Experiment 2. Values indicate means, with standard error in parentheses.

	Sleep deprivation group	Rested control group	Difference between groups
Digit span	6.38 (0.21)	7.10 (0.23)	p = .02
Chronotype (total MEQ score)	44 (1.68)	49 (2.02)	p = .09
Spelling (proportion of words spelt correctly)	0.41 (0.04)	0.44 (0.04)	p = .63
Vocabulary (proportion of words recognised	0.76 (0.01)	0.77 (0.02)	p = .79
Reading (reading efficiency index)	108 (0.24)	94 (0.43)	p = .18[†]
Sleep duration on the night before experiment (in hours)	6.12 (0.29)	7.22 (0.35)	p = .02
Sleep duration on the experimental night (in hours)	N/A	5.37 (0.26)	N/A
Sleep duration on the first recovery night (in hours)	7.25 (0.47)	7.19 (0.28)	p = .91
Sleep duration on the second recovery night (in hours)	6.35 (0.40)	7.14 (0.33)	p = .14

[†] Comparison used Kruskal-Wallis tests of differences between groups as the data are ordinal.

was also significant. Participants were overall faster on Test 2 compared to Test 1.

Spelling. One participant in the sleep deprivation group had a mean accuracy rate of just 1% and was thus removed from analyses; another in the control group was removed for responding extremely slowly (above 20 s on average). Accuracy was analysed using a mixed effects logistic regression model with group (sleep deprivation vs. control), time of testing (test 1 vs. test 2), and familiarity (trained vs. untrained stimuli) as fixed effects (Fig. 9) and revealed a significant effect of familiarity. Trained words were more accurately spelled than untrained words. There was also a significant effect of sleep group, with the control group responding more accurately than the sleep deprivation group. The RT data on correct responses, analysed using the same fixed

factors, revealed a significant effect of time of testing, with faster responding at Test 2 than at Test 1. The results also showed a significant effect of familiarity, with responses to trained words faster than those to untrained words.

Phoneme Knowledge. Analysis of accuracy rates using a mixed effects logistic regression model with group (sleep deprivation vs. control) and time of testing (test 1 vs. test 2) as fixed effects (Fig. 10) revealed a significant effect of sleep group. The control group performed significantly more accurately than the sleep deprivation group. The effect of time of testing was also significant, with participants performing more accurately at Test 2 than Test 1. The RT data on correct responses, analysed with the same fixed factors, also revealed a significant effect of time of testing, with faster responding at Test 2 than at Test 1.

Table 4
Predictors of accuracy and response time for each of the four test tasks in Experiment 2. Accuracy in the old new decision task is calculated as d'.

Fixed effect	Accuracy*				Response Time			
	b	SE	t/z	p	b	SE	t	p
Old New Decision Task[†]								
Group	0.23	0.20	1.35	0.261	0.14	0.14	0.99	0.329
Time of testing	0.001	0.12	0.01	0.992	**-0.22**	**0.07**	**−3.29**	**0.002**
Training	–	–	–	–	**0.16**	**0.02**	**9.57**	**< 0.001**
Group × Time of testing	-0.008	0.18	-0.045	0.964	0.18	0.13	1.33	0.190
Group × Training	–	–	–	–	-0.02	0.03	-0.58	0.565
Time of testing × Training	–	–	–	–	-0.02	0.03	-0.56	0.575
Group × Time of testing × Training	–	–	–	–	-0.01	0.06	-0.21	0.837
Reading Aloud Task[†]								
Group	1.08	0.57	1.87	0.061	-0.01	0.07	-0.18	0.857
Time of testing	0.07	0.15	0.48	0.628	**-0.05**	**0.02**	**−2.92**	**0.005**
Training	**-0.13**	**0.06**	**− 2.03**	**0.042**	**0.05**	**0.08**	**5.97**	**< 0.001**
Group × Time of testing	-0.03	0.29	-0.09	0.930	-0.02	0.04	-0.55	0.582
Group × Training	0.04	0.13	0.35	0.725	0.01	0.02	0.74	0.463
Time of testing × Training	-0.05	0.12	-0.44	0.662	-0.02	0.02	-0.98	0.328
Group × Time of testing × Training	-0.08	0.23	-0.35	0.723	0.01	0.03	0.37	0.715
Spelling Task[‡]								
Group	**0.82**	**0.40**	**2.06**	**0.040**	-0.01	0.01	− 1.33	0.191
Time of testing	-0.01	0.19	-0.08	0.938	**-0.005**	**0.002**	**− 2.44**	**0.018**
Training	**-0.25**	**0.07**	**− 3.57**	**< 0.001**	**0.003**	**0.007**	**2.70**	**0.007**
Group × Time of testing	0.48	0.38	1.27	0.206	0.001	0.004	0.17	0.866
Group × Training	-0.03	0.14	-0.19	0.847	0.001	0.002	0.39	0.700
Time of testing × Training	0.14	0.14	0.97	0.332	< 0.001	0.002	0.004	0.997
Group × Time of testing × Training	0.37	0.28	1.31	0.190	-0.008	0.007	− 1.68	0.093
Phoneme Task[‡]								
Group	**1.26**	**0.48**	**2.65**	**0.008**	-0.06	0.05	− 1.28	0.208
Time of testing	**0.70**	**0.29**	**2.36**	**0.018**	**-0.07**	**0.02**	**− 3.55**	**0.001**
Training	–	–	–	–	–	–	–	–
Group × Time of testing	0.74	0.47	1.56	0.118	-0.03	0.04	-0.68	0.499
Group × Training	–	–	–	–	–	–	–	–
Time of testing × Training	–	–	–	–	–	–	–	–
Group × Time of testing × Training	–	–	–	–	–	–	–	–

[†] RT data were log transformed for the statistical analysis
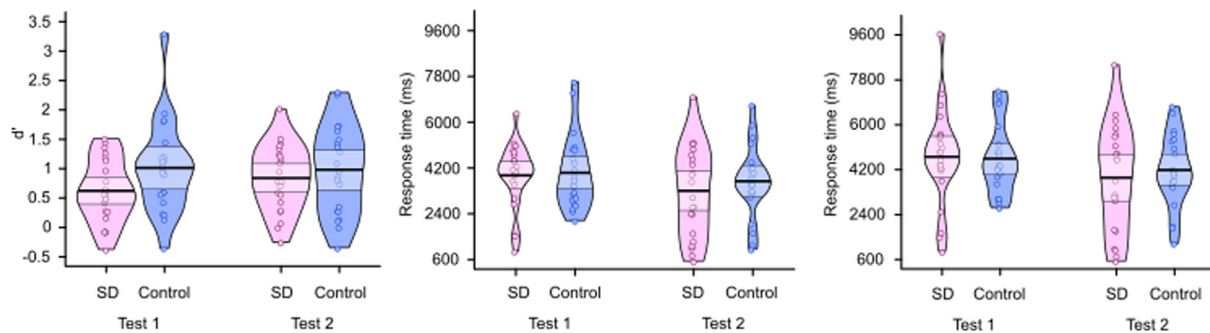[‡] RT data were inverse transformed for the statistical analysis

Fig. 7. Accuracy rates (A) and response times for trained (B) and untrained (C) words in the old-new decision task in Experiment 2. The horizontal line represents the mean, the box around the mean represents 95% confidence intervals, the borders around data points are smoothed density curves.

4. Discussion

Active consolidation theory led us to predict that generalisation in language learning would be substantially disadvantaged by a full night of sleep deprivation, both before (Saletin & Walker, 2012) and after (McClelland et al., 1995) learning. Results did not straightforwardly support this prediction (see Table 5 for a summary of findings). Experiment 1 tested the impact of sleep deprivation after learning. Across three generalisation tasks with two measures each (RT and accuracy), we observed an effect of sleep deprivation in only the phoneme knowledge task in the RT measure. This result is consistent with poorer generalisation in the sleep deprivation group (or at least poorer access to general knowledge). We discuss below possible reasons why poorer

generalisation or access to general knowledge would manifest only in this way, and not in spelling or reading aloud of untrained items.

Experiment 2 tested the impact of sleep deprivation before learning. Sleep deprived participants required more training to reach criterion. This could reflect decreasing hippocampal capacity following extended wake but could also reflect increased lapses in attention known to arise in sleep deprived individuals (Lim & Dinges, 2008). Furthermore, the sleep deprived group had a slightly but statistically significantly lower working memory capacity (as measured by digit span) which may also have contributed to this difference. Turning to the test tasks, across six sets of analyses probing generalisation, we observed an effect of sleep deprivation only in the accuracy measure of the spelling and phoneme knowledge tasks. We note though that our control group got about two
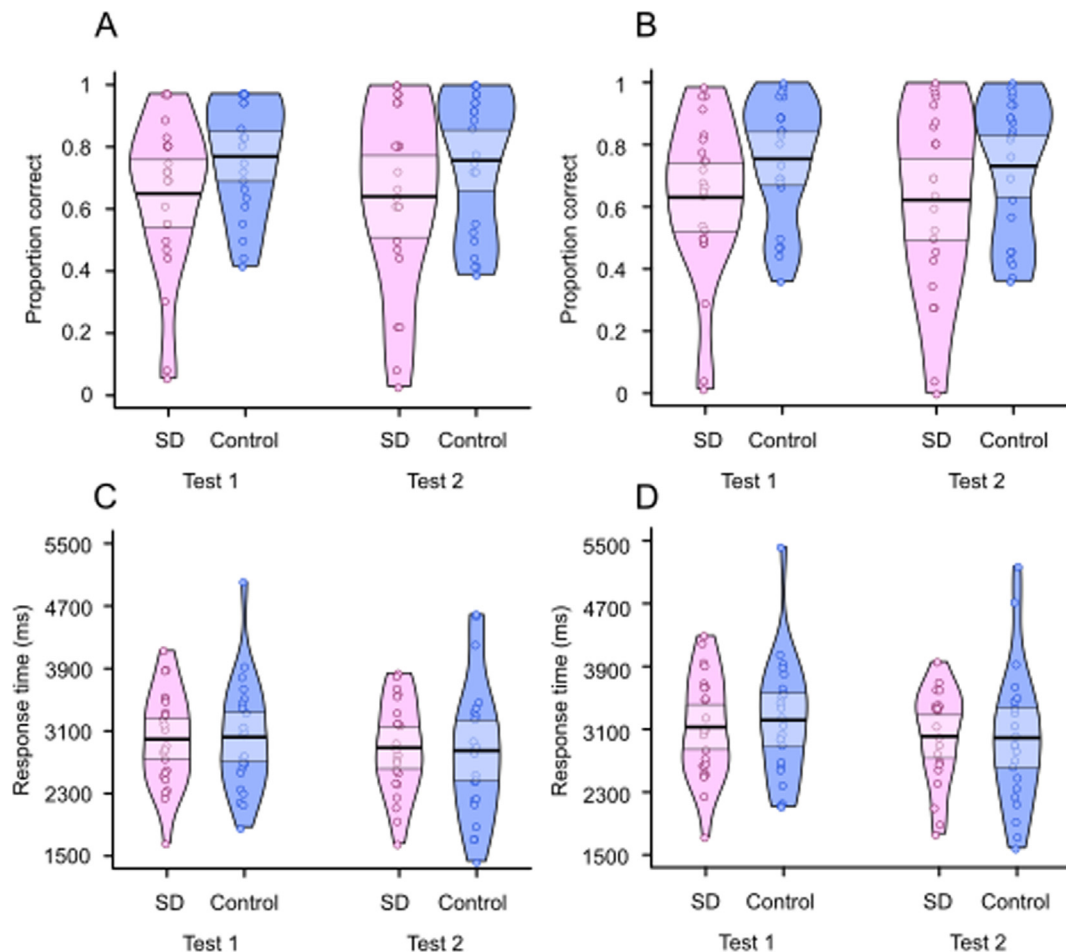

Fig. 8. Accuracy rates for trained (A) and untrained (B) words, response times for trained (C) and untrained (D) words in the reading aloud task in Experiment 2. The horizontal line represents the mean, the box around the mean represents 95% confidence intervals, the borders around data points are smoothed density curves.
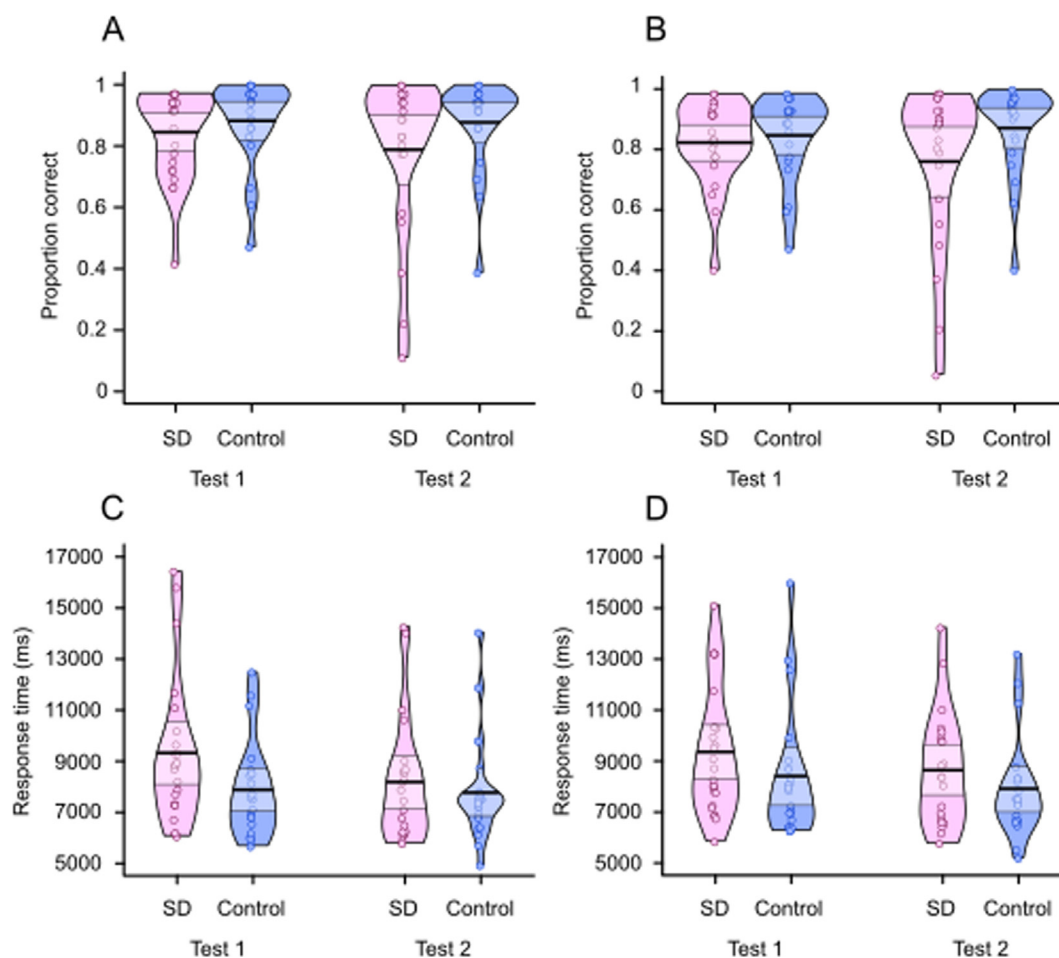
Fig. 9. Accuracy rates for trained (A) and untrained (B) words, response times for trained (C) and untrained (D) words in the spelling task in Experiment 2. The horizontal line represents the mean, the box around the mean represents 95% confidence intervals, the borders around data points are smoothed density curves.

Fig. 10. Accuracy rates (A) and response times (B) in the phoneme knowledge task in Experiment 2. The horizontal line represents the mean, the box around the mean represents 95% confidence intervals, the borders around data points are smoothed density curves.
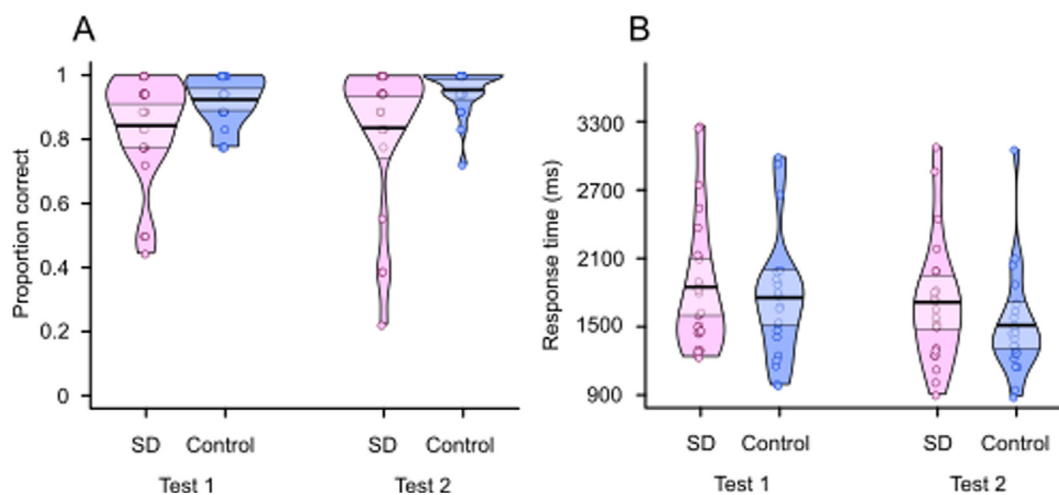
hours less sleep on the experimental night compared to the night before or after (Table 3). This is likely due to the need to arrive in the lab relatively early in the morning for the training session. This slight sleep restriction may have attenuated differences between the groups. The phoneme knowledge results could reflect poorer generalisation following sleep deprivation, and we discuss below possible reasons why these impacts might manifest only in this task.

Our findings of limited impact of sleep deprivation on generalisation appear to be at odds with the other language learning studies cited in the Introduction which have shown that generalisation in grammar learning benefits from sleep-associated consolidation (Gomez et al., 2006; Nieuwenhuis et al., 2013; Batterink et al., 2014; Mirković & Gaskell, 2016; Batterink & Paller, 2017). However, there are some fundamental differences between the type of learning employed in

Table 5
Summary of significant and non-significant main effects across the two experiments.

	Sleep deprivation vs. rested controls	Trained words vs. untrained words	First test vs. second test
Experiment 1			
Old-New decision	No	Yes (RT only)	Yes (RT only)
Reading aloud	No	Yes (Acc & RT)	Yes (RT only)
Spelling	No	Yes (Acc & RT)	No
Phoneme knowledge	Yes (RT only)	N/A	Yes (RT only)
Experiment 2			
Old-New decision	No	Yes (RT only)	Yes (RT only)
Reading aloud	No	Yes (Acc & RT)	Yes (RT only)
Spelling	Yes (Acc only)	Yes (Acc only)	Yes (RT only)
Phoneme knowledge	Yes (Acc only)	N/A	Yes (Acc & RT)

these studies and our study. In the grammar learning studies, the generalised knowledge is not directly observable and needs to be extracted on the basis of statistical regularities in the artificial language. This is in contrast to our task, where each symbol is paired with one sound consistently all of the time. In other words, even though the symbols occur in multiple different contexts (i.e. within different words), the symbol-sound mappings in our paradigm are deterministic and there are no statistical patterns that need extracting for generalisation to occur. Our data therefore might support a view whereby consolidation benefits the extraction of statistical patterns in the environment but plays no role, or a minor role, in strengthening direct associations observed in multiple contexts during wake. Indeed, there is evidence to suggest that high degree of contextual variability facilitates generalisation (e.g., Gomez, 2002).

Even taking the above into consideration, our results seem inconsistent with the Complementary Learning Systems (McClelland et al., 1995) account which proposes that the hippocampus uses discrete representations of newly learned memories, unable to integrate information across multiple learning episodes. If such representations correspond to the trained words in our paradigm, it is not clear how these representations would support extraction of context-independent symbol-sound mappings from across discrete word representations. However, a recent update to the original CLS account (Kumaran, Hassabis, & McClelland, 2016; Koster et al., 2018) suggests that the hippocampus acts a big-loop recurrent circuit, whereby activation of a discrete episodic memory can be recirculated as new input into the hippocampus which allows the selective activation of a number of other memories related to the original memory. This selective activation of related memories allows the integration of discrete learning episodes to give rise to generalisation and inference without support from the neocortex. According to this view it should be possible to observe at least some simple forms of generalisation (perhaps ones that do not require statistical learning) in the absence of consolidation.

A second possible reason why consolidation effects might not be observed in our learning paradigm can be found in the recent literature on schema learning. Tse and colleagues (Tse et al., 2007, 2011) have shown in animal studies that neocortical consolidation is accelerated when the information being encoded fits into an existing schema. In these studies, rats learned paired associations in one environment and were later able to learn new associations in the same environment while encoding the new information in parallel in the hippocampus and the neocortex, showing that the rate of learning in the neocortex is similar to that of the hippocampus when the new information is being assimilated into an existing schema. Our learners were all skilled readers of English, and therefore have existing symbol-sound mappings which may act as existing schemas. Under the schema learning view, the artificial orthography task requires a modification to these existing mappings rather than the learning of completely new information and

may consequently result in accelerated consolidation.

A third mechanism that has recently been suggested to allow acceleration of consolidation is retrieval-mediated learning (sometimes referred to as the "testing effect"). Antony, Ferreira, Norman, and Wimber (2017) have argued that retrieval of newly acquired memories has a similar impact to sleep-associated consolidation. According to this theory retrieval (unlike study) is imprecise: retrieval of one memory coactivates other memories that are semantically or episodically related to the retrieved memory. Thus, retrieval might not only strengthen memory of the learning episodes but might also facilitate integration of information across learning episodes, leading to generalisation. Given that in our training paradigm each training trial required retrieval, it is possible that generalisation was achieved already during training without need for further consolidation processes[1].

Though we sought primarily to study the impact of sleep deprivation on generalisation, it is interesting that we did not find robust effects on memory for trained exemplars in either experiment. Only the accuracy measure of the spelling task in Experiment 2 provided some indication of disadvantaged learning of exemplars. Instead, sleep deprived participants learned novel words to a high degree of accuracy, and performance at test 10 days later showed stable or improved access to memories in line with control participants. One potential explanation for the lack of a sleep deprivation effect here comes from the growing literature suggesting that sleep benefits weakly encoded memories more than strongly encoded memories (e.g., Kuriyama, Stickgold, & Walker, 2004; Peters, Smith, & Smith, 2007; Djonlagic et al., 2009; Diekelmann, Born, & Wagner, 2010; Sio, Monaghan, & Ormerod, 2013; Creery, Oudiette, Antony, & Paller, 2015). For example, Drosopoulos, Schulze, Fischer, and Born (2007) had participants encode word pairs in a strong or weak encoding condition. Strongly encoded pairs showed no benefit of sleep, while weakly encoded pairs did. In the current studies we intentionally trained participants to a high criterion, as confirmed by high accuracy in the reading and spelling tasks at the end of training. This was done to ensure that any potential difficulties in generalisation would not be due to insufficient encoding of the training stimuli or encoding level differences between groups. However, it does leave open the possibility that we failed to see sleep deprivation effects for the trained items simply because they were very strongly encoded.

Finally, it is important to note that we did not test exclusive sleep and exclusive wake conditions: sleep deprivation participants were tested following two nights of recovery sleep in the first instance. Recent theorising suggests that the hippocampus can act as a buffer during wake following learning and the consolidation processes can thus be delayed until the first sleep opportunity occurs (Schönauer, Gratsch, & Gais, 2015). Our findings from Experiment 1 would be consistent with this delayed sleep hypothesis. Similarly, recent work suggests that recovery sleep compensates for poor encoding due to disrupted sleep before learning. Baena, Cantero, Fuentemilla, and Atienza (2020) had participants encode face-pairs after either a normal night of sleep or a night of restricted sleep. A normal night of recovery

---

[1] We conducted an exploratory analysis to evaluate the influence of retrieval practice on our observed test performance by calculating correlations between number of retrievals during training (measured as number of training blocks completed) and accuracy in the phoneme judgement test task (as a test of generalisation) and in the reading aloud of trained words test task (as a test of memory for trained stimuli). The retrieval hypothesis predicts a positive correlation between number of training blocks and test performance. In both experiments these correlations were statistically significant but negative: higher number of training blocks was associated with *lower* accuracy at test, possibly reflecting poorer learners needing more training to reach criterion. These analyses are reported in full in Supplemental Materials. Without testing a training regime with no retrieval at all it is impossible to definitively rule out the role of retrieval in our data. However, the lack of a positive correlation suggests that retrieval hypothesis alone is unlikely to be a sufficient explanation for our findings.

sleep followed the encoding session. EEG measures during the recovery night showed higher slow oscillation-spindle coupling in the restricted sleep condition, suggesting that consolidation processes involving hippocampal-neocortical dialogue were increased to compensate for impaired encoding after sleep restriction. No difference in recall between the groups was found after the recovery night. Our findings from Experiment 2 are consistent with this form of the delayed sleep hypothesis.

One interesting possibility is that selection biases in our study contributed to the failure to observe robust effects of sleep deprivation on memory and learning. Our decision to train participants to criterion and dismiss participants who could not complete the initial learning meant that our sample may have comprised individuals of high learning and memory ability. It might be argued that sleep-related consolidation is more important for individuals of lower learning and memory ability. This is however inconsistent with the few studies that have looked at individual differences is sleep-associated memory consolidation. Fenn and colleagues have shown that those with high intelligence (Fenn & Hambrick, 2015) and high working memory capacity (Fenn & Hambrick, 2012) show *larger*, not smaller, sleep benefits on declarative memory. Similarly, there are large and stable individual differences in the vulnerability to the impact of sleep deprivation on cognitive processing (see Van Dongen, Vitellaro, & Dinges, 2005 for a review). It could be that individuals who volunteered for our sleep deprivation study were particularly resilient to the effects of sleep deprivation; there was some evidence for this in Experiment 2 in which participants' PVT latencies did not increase through the night (see supplementary materials). These individual differences may have unknown consequences for memory consolidation or restoration of encoding ability, for example less vulnerable individuals may be better able to delay consolidation processes until a later sleep opportunity or may be more resilient against reducing encoding ability over extended periods of wake. More research is needed to understand why some individuals benefit less from sleep-associated memory processes, and what alternative cognitive and neural mechanisms they might be employing.

4.1. Conclusions

Participants in both experiments showed strong generalisation: their accuracy rates in reading and spelling untrained words in the artificial orthography were high, and there were no consistent differences between sleep deprived and control participants on accuracy or RT measures of these tasks across experiments. We did find effects of sleep deprivation on the phoneme knowledge measure across both experiments, albeit on RT in one experiment and accuracy in the other. Conversely, we observed reliable effects of training and time of testing across test tasks and across experiments, suggesting our experiments were sufficient powered to detect some learning effects.

These findings show similarities to other recent data that have looked at the impact of sleep vs. wake following learning on generalisation. Schapiro et al. (2017) found that memory of shared properties of a set of newly learned stimuli (fictitious satellites) improved during sleep after learning, but the application of this knowledge to untrained exemplars did not benefit from sleep. Our data showed a similar pattern: in the phoneme knowledge task in Experiment 1 participants in the sleep deprivation group were on average 23% (339 ms) slower than control participants. In Experiment 2 participants in the sleep deprivation group were on average 11% less accurate than controls in this task. Therefore, memory for shared properties in the form of sound-symbol mappings in the phoneme knowledge task was somewhat impaired by sleep deprivation after learning, but the application of these mappings to untrained exemplars in reading aloud and spelling was unaffected by lack of sleep. It is however also important to consider whether reading and spelling rely on the same memory systems as the phoneme knowledge task. Reading and spelling are tasks that skilled readers are highly practised in and are more likely to reflect normal online language processing than the phoneme knowledge task. Reading and spelling may therefore involve a larger implicit memory component (with contributions from memory processes akin to conditioning, priming, and perceptual learning, for example) than the phoneme task which is the closest of our tasks to classical hippocampus-dependent explicit memory tasks.

Whatever the relative contributions of explicit and implicit memory turn out to be, it seems clear none of these tasks are process-pure. A related point applies to our old-new categorisation task. Within the recognition memory literature there is controversy about the role of the hippocampus on recognition memory. According to the dual-process view, recognition memory responses can be made either on the basis of familiarity or recollection. These responses may rely on different neural structures, with responses made on the basis of recollection relying on the hippocampus, and responses based on familiarity relying on other structures, such as the perirhinal system (e.g., Brown & Aggleton, 2001), although others have suggested that both familiarity and recollection rely on the hippocampus (e.g., Jeneson, Kirwan, Hopkins, Wixted, & Squire, 2010). Future studies should either use tasks that clearly rely on recollection (such as free recall) or ask participants in recognition memory tasks whether a response was made on the basis of familiarity or recollection.

Our results show that eliminating sleep on the first night before learning (Experiment 2) or the first night after learning (Experiment 1) has little impact on the acquisition and generalisation of a new writing system. Our results from Experiment 1 are inconsistent with the original formulation of the CLS account which suggested that generalisation is not possible prior to memory consolidation and that such consolidation occurs preferentially during sleep. However, these data can be accommodated by more recent theories that allow for accelerated sleep-independent consolidation as a result of schemas or retrieval-induced learning, and by the recent extensions to the CLS account that suggest at least some forms of generalisation are possible prior to consolidation based solely on hippocampally-stored memory. Our results from Experiment 2 are somewhat consistent with theories that suggest prolonged wake impairs encoding ability (e.g., (Tononi & Cirelli, 2014) in that our sleep deprived participants required more training to reach the learning criterion, but no long-term disadvantage on episodic memory or generalisation was found. These data could be accommodated by recent theories that propose that the first night of sleep before or after learning is not critical, but that new information can be buffered in the hippocampus until the first sleep opportunity (e.g., Baena et al., 2020). Alternatively, sleep before or after learning may be unnecessary in the learning and generalisation of strongly encoded materials. More research is needed to elucidate the impact of sleep on generalisation, if any, and to inform revisions to existing theories, which may currently assume too broad a role for sleep in memory consolidation.

CRediT authorship contribution statement

Jakke Tamminen: Conceptualization, Methodology, Formal analysis, Writing - original draft, Visualization, Supervision, Project administration, Funding acquisition. **Chloe R. Newbury:** Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Rebecca Crowley:** Investigation, Writing - original draft. **Lydia Vinals:** Methodology, Software, Formal analysis, Investigation, Data curation, Writing - review & editing. **Benedetta Cevoli:** Methodology, Software, Investigation, Writing - review & editing. **Kathleen Rastle:** Conceptualization, Methodology, Writing - original draft, Supervision, Project administration, Funding acquisition.

Acknowledgements

Declaration of Competing Interest

None.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nlm.2020.107274.

References

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in Cognitive Science, 21*, 573–576.

Baena, D., Cantero, J. L., Fuentemilla, L., & Atienza, M. (2020). Weakly encoded memories due to acute sleep restriction can be rescued after one night of recovery sleep. *Scientific Reports, 10*, 1449.

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

Batterink, L. J., Oudiette, D., Reber, P. J., & Paller, K. A. (2014). Sleep facilitates learning a new linguistic rule. *Neuropsychologia, 65*, 169–179.

Batterink, L. J., & Paller, K. A. (2017). Sleep-based memory processing facilitates grammatical generalization: Evidence from targeted memory reactivation. *Brain & Language, 167*, 83–93.

Brown, M., & Aggleton, J. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience, 2*, 51–61.

Burt, J. S., & Tate, H. (2002). Does a reading lexicon provide orthographic representations for spelling? *Journal of Memory and Language, 46*, 518–543.

Creery, J. D., Oudiette, D., Antony, J. W., & Paller, K. A. (2015). Targeted memory reactivation during sleep depends on prior learning. *Sleep, 38*, 755–763.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11*, 114–126.

Diekelmann, S., Born, J., & Wagner, U. (2010). Sleep enhances false memories depending on general memory performance. *Behavioral Brain Research, 208*, 425–429.

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers, 17*, 652–655.

Djonlagic, I., Rosenfeld, A., Shohamy, D., Myers, C., Gluck, M., & Stickgold, R. (2009). Sleep enhances category learning. *Learning & Memory, 16*, 751–755.

Drosopoulos, S., Schulze, C., Fischer, S., & Born, J. (2007). Sleep's function in the spontaneous recovery and consolidation of memories. *Journal of Experimental Psychology: General, 136*, 169–183.

Drummond, S. P., Brown, G. G., Gillin, J. C., Stricker, J. L., Wong, E. C., & Buxton, R. B. (2000). Altered brain response to verbal learning following sleep deprivation. *Nature, 403*, 655–657.

Fenn, K. M., & Hambrick, D. Z. (2012). Individual differences in working memory capacity predict sleep-dependent memory consolidation. *Journal of Experimental Psychology: General, 141*, 404–410.

Fenn, K. M., & Hambrick, D. Z. (2015). General intelligence predicts memory change across sleep. *Psychonomic Bulletin & Review, 22*, 791–799.

Gais, S., Albouy, G., Boly, M., Dang-Vu, T. T., Darsaud, A., Desseilles, M., ... Peigneux, P. (2007). Sleep transforms the cerebral trace of declarative memories. *Proc. Natl. Acad. Sci. U.S.A. 104*, 18778–18783.

Gais, S., Lucas, B., & Born, J. (2006). Sleep after learning aids memory recall. *Learning & Memory, 13*, 259–262.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431–436.

Gomez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychological Science, 17*, 670–674.

Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology, 4*, 97–100.

Jeneson, A., Kirwan, C. B., Hopkins, R. O., Wixted, J. T., & Squire, L. R. (2010).

Recognition memory and the hippocampus: A test of the hippocampal contribution to recollection and familiarity. *Learning & Memory, 17*, 63–70.

Koster, R., Chadwick, M. J., Chen, Y., Berron, D., Banino, A., Düzel, E., ... Kumaran, D. (2018). Big-loop recurrence within the hippocampal system supports integration of information across episodes. *Neuron, 99*, 1342–1354.

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends in Cognitive Science, 20*, 512–534.

Kuriyama, K., Stickgold, R., & Walker, M. P. (2004). Sleep-dependent learning and motor-skill complexity. *Learning & Memory, 11*, 705–713.

Lau, H., Alger, S. E., & Fishbein, W. (2011). Relational memory: A daytime nap facilitates the abstraction of general concepts. *PloS ONE, 6*, Article e27139.

Lim, J., & Dinges, D. F. (2008). Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences, 1129*, 305–322.

Mander, B. A., Santhanam, S., Saletin, J. M., & Walker, M. P. (2011). Wake deterioration and sleep restoration of human learning. *Current Biology, 21*, R183–R184.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419–457.

Mirković, J., & Gaskell, M. G. (2016). Does sleep improve your grammar? Preferential consolidation of arbitrary components of new linguistic knowledge. *PLoS ONE, 11*, Article e0152489.

Nieuwenhuis, I. L. C., Folia, V., Forkstam, C., Jensen, O., & Petersson, K. M. (2013). Sleep promotes the extraction of grammatical rules. *PLoS ONE, 8*, Article e65046.

Peters, K, Smith, V, & Smith, C (2007). Changes in sleep architecture following motor learning depend on initial skill level. *Journal of Cognitive Neuroscience*.

Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods, 39*, 859–862.

Saletin, J. M., & Walker, M. P. (2012). Nocturnal mnemonics: Sleep and hippocampal memory processing. *Frontiers in Neurology, 3*, 59.

Schapiro, A. C., McDevitt, E. A., Chen, L., Norman, K. A., Mednick, S. C., & Rogers, T. T. (2017). Sleep benefits memory for semantic category structure while preserving exemplar-specific information. *Scientific Reports, 7*, 14869.

Schönauer, M., Gratsch, M., & Gais, S. (2015). Evidence for two distinct sleep-related long-term memory consolidation processes. *Cortex, 63*, 68–78.

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *The Journal of Psychology: Interdisciplinary and Applied, 9*, 371–377.

Sio, U. N., Monaghan, P., & Ormerod, T. (2013). Sleep on it, but only if it is difficult: Effects of sleep on problem solving. *Memory & Cognition, 41*, 159–166.

Squire, L, & Dede, A (2015). Conscious and unconscious memory systems. *Cold Spring Harbor Perspectives in Biology*.

Tamminen, J., Davis, M. H., Merkx, M., & Rastle, K. (2012). The role of memory consolidation in generalisation of new linguistic information. *Cognition, 125*, 107–112.

Tamminen, J., Davis, M. H., & Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive Psychology, 79*, 1–39.

Taylor, J. S. H., Davis, M. H., & Rastle, K. (2017). Comparing and validating methods of reading instruction using behavioural and neural findings in an artificial orthography. *Journal of Experimental Psychology: General, 146*, 826–858.

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology, 86*, 255–262.

Tononi, G, & Cirelli, C (2014). Sleep and the price of plasticity: From synaptic and cellular homeostasis to memory consolidation and integration. *Neuron*.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of Word Reading Efficiency-Second Edition.* Austin, TX: Pro-Ed.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., ... Morris, R. G. (2007). Schemas and memory consolidation. *Science, 316*, 76–82.

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., ... Morris, R. G. (2011). Schema-dependent gene activation and memory encoding in neocortex. *Science, 333*, 891–895.

Van Dongen, H. P. A., Vitellaro, K. M., & Dinges, D. F. (2005). Individual Differences in Adult Human Sleep and Wakefulness: Leitmotif for a Research Agenda. *Sleep, 28*, 479–495.

Yoo, S., Hu, P., Gujar, N., Jolesz, F., & Walker, M. (2007). A deficit in the ability to form new human memories without sleep. *Nature Neuroscience, 10*, 385–392.