Journal of Experimental Psychology:
Learning, Memory, and Cognition
2013, Vol. 39, No. 2, 608–622

0278-7393/13/$12.00 DOI: 10.1037/a0029243

Lexical Integration of Novel Words Without Sleep

Shane Lindsay
University of Dundee

M. Gareth Gaskell
University of York

Learning a new word involves integration with existing lexical knowledge. Previous work has shown that sleep-associated memory consolidation processes are important for the engagement of novel items in lexical competition. In 3 experiments we used spaced exposure regimes to investigate memory for novel words and whether lexical integration can occur within a single day. The degree to which a new spoken word (e.g., *cathedruke*) engaged in lexical competition with established phonological neighbors (e.g., *cathedral*) was employed as a marker for lexical integration. We found evidence for improvements in recognition and cued recall following a time period including sleep, but we also found lexical competition effects emerging within a single day. Spaced exposure to novel words on its own did not bring about this within-day lexical competition effect (Experiment 2), which instead occurred with either spaced or massed exposure to novel words, provided that there was also spaced exposure to the phonological neighbors (Experiments 1 and 3). Although previous studies have indicated that sleep-dependent memory consolidation may be sufficient for lexical integration, our results show it is not a necessary precondition.

Keywords: spaced learning, testing effect, lexical competition, memory consolidation, sleep

Although traditionally a topic that has been neglected in cognitive psychology, memory consolidation has been intensively studied in recent years, bolstered by findings from cognitive and behavioral neuroscience. Research on consolidation has shown that after an initial learning episode, memory processes continue without further exposure, as unstable memory traces are transformed from an initial unstable state to a stronger and more permanent form. Recent work has shown that sleep has an important role in memory consolidation (see Walker & Stickgold, 2006, for a review), with increasing evidence for sleep-based consolidation in a variety of memory tasks, including both implicit learning of motor skills and more explicit declarative knowledge (Gais & Born, 2004). Studies of sleep-associated memory consolidation have typically demonstrated that newly acquired memories become more resilient to interference or decay following sleep (Walker & Stickgold, 2006). Strikingly, in some cases performance can actually improve following sleep, despite no further practice or rehearsal, such as in the case of motor skills (Fischer, Hallschmid, Elsner, & Born, 2002). Another important aspect of memory consolidation, which we know less about, is how new memories are integrated with old memories and the role for sleep in this integration process (Walker & Stickgold, 2010).

This article was published Online First July 9, 2012.

Shane Lindsay, School of Psychology, University of Dundee, Dundee, United Kingdom; M. Gareth Gaskell, Department of Psychology, University of York, York, United Kingdom.

Experiment 1 was previously presented at the 31st annual meeting of the Cognitive Science Society, Amsterdam, the Netherlands, July 2009 (Lindsay & Gaskell, 2009). The research was supported by U.K. Economic and Social Research Council Research Grant RES-063-27-0061, awarded to M. Gareth Gaskell. We thank Alan Garnham for his help with data collection at the University of Sussex.

Correspondence concerning this article should be addressed to Shane Lindsay, School of Psychology, University of Dundee, Dundee DD1 4HN, United Kingdom. E-mail: shane.lindsay@gmail.com

One of the few lines of research looking at the integration of new memories with old memories is one that has investigated the time course of word learning in adults. When a new word is learned, it can be said to be fully lexically integrated when its representation is able to interact with representations of established words in the lexicon. One marker for such integration is the establishment of lexical competition, a central construct for theories of spoken word recognition. When one listens to a spoken word, the incoming acoustic information unfolds over time, leading to the activation of matching word candidates in one's mental lexicon (Marslen-Wilson, 1993). For example, hearing the first part of the word *captain* activates competitors such as *capsule* and *captive* before acoustic information eventually distinguishes the incoming word from its competitors. Gaskell and Dumay (2003) taught participants the phonological forms of fictitious novel words that were derived from real words with early uniqueness points (the point at which a word diverges from competitors). Once a novel word has been integrated into the phonological lexicon it should engage in lexical competition, which would be demonstrated by slower recognition of the novel word's phonological competitors. Participants in Gaskell and Dumay's experiment were able to recognize these novel words accurately immediately after training. However, there was no immediate evidence for these novel words engaging in lexical competition. Instead, it appeared that the integration of the form of these new words into the lexicon had a time course that spanned several days. This measure of lexical integration was purely form based, because meanings were not explicitly provided for the words. Conceivably, a lexical entry without a meaning may be represented in a qualitatively different way than one with a meaning, although there is some evidence that participants will generate their own meaning for a word if one is not provided (Tamminen, 2010). In any case, studies of lexical competition where meanings were provided, for both adults and children, suggest that the time course of the emergence of lexical competition is not influenced by the presence or absence of this meaning (Dumay, Gaskell, & Feng, 2004; Henderson, Weighall, & Gaskell, 2012).

Dumay and Gaskell (2007) investigated the time course of lexical integration by focusing on the possible role of sleep in the lexical integration process. They trained participants on novel words either in the morning or in the evening. Results of a lexical competition test using existing neighbors immediately after familiarization were compared with a second test after a 12-hr gap (with an additional 24-hr test to control for time-of-day confounds). When this 12-hr interval contained nocturnal sleep, lexical competition effects were found, yet there was no evidence for lexical competition following the 12-hr gap that did not include sleep. A special role for sleep in lexical integration is supported by a recent study (Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010) in which enhanced lexical competition for existing neighbors of novel words after sleep was found to be associated with an aspect of sleep physiology called sleep spindles, which are thought to be a marker of hippocampal–neocortical interaction. These findings support the idea that nocturnal sleep is important for lexical integration (as defined by the impact of a novel word on the recognition of an established neighbor). Similar studies have also found little evidence of lexical competition effects prior to sleep (e.g., Davis, Di Betta, Macdonald, & Gaskell, 2009; Dumay & Gaskell, 2012; Dumay et al., 2004; Tamminen & Gaskell, 2008), although the spindle study described above (Tamminen et al., 2010) found some evidence for lexical competition prior to sleep. Although participants are able to recognize new words pre-sleep, representations of novel words do not seem to be in a suitable form for lexical competition effects to emerge. The question addressed in the current work was whether enhanced training before sleep could lead to more immediate lexicalization. In order to enhance learning of the novel words, we implemented a spaced training and testing regime.

In spaced or distributed learning, exposure to items during learning is spread out over time. Spaced learning is usually contrasted with massed learning, in which the equivalent amount of exposure is given all at once. The enhancement in memory with spaced learning, known as the spacing effect, has been one of the most studied topics in memory research and has been shown to occur across different time periods, different learning strategies, materials, and species (see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006, for a review). Spaced learning studies have demonstrated improved performance in the learning of words in young children (Schwartz & Terrell, 1983). For adults, the benefits of spaced learning in vocabulary acquisition have long been known (Dempster, 1987), and this method is often applied in second language teaching. A phenomenon related to spaced learning is the testing effect. Effortful retrieval involved in testing item knowledge has been shown to considerably enhance memory compared with study (Carrier & Pashler, 1992), and like spaced learning, spaced retrieval practice can lead to significant benefits for memory (Karpicke & Roediger, 2008; Landauer & Bjork, 1978). These findings suggest that spaced exposure and retrieval of new words should enhance performance in explicit tests of memory (e.g., recall or recognition). However, given the evidence of the importance of sleep in the emergence of lexical competition, it remains an open question as to whether spacing may accelerate the engagement of novel words in the lexical competition process to occur in a single day.

Spaced learning is often studied by varying the number of intervening items in a list learned in a single study session. In other cases, the effect is studied with intervals across days, weeks, or longer. However, there has been relatively little investigation of spaced learning (or spaced testing) with learning sessions spread across a single day (although cf. Shea, Lai, Black, & Park, 2000, in the domain of motor learning). Alongside the general evidence for memory benefits following spacing of study and test, there are reasons to think that spaced exposure within a single day may be particularly beneficial for lexical integration. One theoretical account of the lexical integration process (Davis & Gaskell, 2009; Lindsay & Gaskell, 2010) is based on the complementary learning systems model (McClelland, McNaughton, & O'Reilly, 1995). This theory proposes that overnight sleep provides an opportunity for the integration of rapidly acquired hippocampal information with existing knowledge stored in a slower learning neocortical system. This dual-system approach is a response to the problem of "catastrophic interference" that can occur in a single distributed connectionist network, whereby newly acquired knowledge overwrites existing distributed knowledge. The solution offered by the complementary systems account is to provide a secondary hippocampal store that uses sparse representations and a high level of plasticity to support initial learning. Integration into the neocortex then happens offline (e.g., during sleep) such that the hippocampal network "teaches" the neocortical network while interleaving the new information with the existing knowledge more gradually (Ans, Rousset, French, & Musca, 2004; Robins, 1996; see G. A. Carpenter & Grossberg, 1991, for an alternative solution).

Sleep is an ideal state for this interleaving, as our cognitive system is offline and is not required for processing new information (McClelland et al., 1995). However, it is not the only means of interleaving. McClelland et al. also argued that interleaving could occur during waking hours as consequence of interleaved exposure to new and old information. Such interleaved presentation could lead to neocortical integration of novel representations during the day. In the domain of word learning, this would predict that some (perhaps weak) evidence of lexical integration (i.e., lexical competition effects for existing words) should be found during the day when training provides a more spaced and/or interleaved mode of exposure. In turn, this could lead to less dramatic changes in the lexical competition environment overnight.

In Experiment 1, participants were intensively trained and tested on novel words in spaced sessions on a single day and tested once again on the following day. Each training session involved both familiarization of novel phonological forms and explicit retrieval practice via cued-recall and recognition memory tests. Engagement of the new words in lexical competition was also repeatedly tested by measuring lexical decision latency to existing phonological neighbors of the newly learned words. This spacing regime allows interleaving with existing words throughout the course of the day. In fact, interleaving occurred in two ways. First, the test materials were as a whole interleaved with the participant's typical daily language environment outside the lab. Second, there was explicit interleaving of novel words and their existing neighbors in the training and testing materials. In Experiment 1 we did not address the question of what kind of interleaving might be necessary for lexical integration. Instead, we focused on whether sleep is a necessary precondition for integration into the phonological lexicon. Along with providing interleaving, the combined use of spaced learning and retrieval practice should maximize participants' encoding of the new words, perhaps increasing the likelihood of pre-sleep lexical integration. On the other hand, if sleep is necessary for lexical integration, then we would not expect participants to show lexical competition effects on the first day of learning, no matter how well the new words had been learned or how much interleaving had occurred, though we should expect to find it when we test the next day.

Experiment 1

Method

Participants. Thirty-seven participants from the University of York were tested. All were native English speakers without visual or auditory impairments. Participants received course credit or were paid £15 for participation. One participant did not complete the sessions, leaving 36 participants in total.

Materials and stimulus construction. Forty monomorphemic words, listed in the Appendix, were chosen to act as what we call "base words" (e.g., *cathedral*) and were selected from a larger set used by Davis et al. (2009). Base words were bi- or trisyllabic and contained six to 11 phonemes ($M = 6.7$). Frequencies ranged from 2 occurrences per million to 18. All base words had an early uniqueness point, located before the final vowel. Fictitious novel word competitors were derived from these base words (e.g., *cathedruke*) and diverged from the base word at their final vowel. The intent was to create a new close competitor, which once integrated in the lexicon would lead to the extension of the uniqueness point of the base word and slow down its recognition (cf. Gaskell & Dumay, 2003). For each novel word, a foil was also created for use in an explicit recognition test (e.g., *cathedruce*). This diverged from the novel word only at the final phoneme. The 40 fictitious novel words were split into two lists for counterbalancing, with each participant learning one list. In order to prevent confusion in the stem completion task, we designed each list so that no items shared the same initial phoneme.

Filler words and nonwords for the lexical decision task were chosen to have properties similar to those of the experimental items. Nonwords were created by changing one phoneme of a real word. Participants heard all 40 base words in the lexical decision task, with 20 items potentially having a new competitor and the remaining 20 items acting as controls, for which no new novel competitor had been learned. Stimuli were recorded by a male native speaker of British English.

Design and procedure. The experiment was split into five sessions. Participants started the first session (S1) at 09:00, 09:30, or 10:00 and came back to the laboratory for the second (S2), third (S3), and fourth (S4) sessions at 2.5-hour intervals; the fifth session (S5) took place 24 hours after the final session of the first day, in order to avoid circadian confounds when comparing performance between S4 and S5.

Four tasks were used in the experiment. Phoneme monitoring served to familiarize participants with the phonological form of the words. Stem completion always followed phoneme monitoring and had a dual role. In terms of learning, it provided retrieval practice that (particularly when combined with feedback) was expected to lead to better retention (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991). Furthermore, it provided a direct test of retention of the novel words. The familiarity decision and lexical decision tasks were primarily used to assess the strength of representation and the integration of the novel words into the phonological lexicon, although these tasks also provided limited additional exposure. Participants did each task four times in total. The structure of sessions and order of tasks in Experiment 1 are shown in Table 1.

Although our main focus was the results from the lexical decision task, we were also interested in how performance in the other tasks changed across the sessions, which is why we included an additional repetition of the phoneme monitoring and stem completion tasks in the final session, by which time further training was no longer necessary. As we were particularly concerned with responses to the lexical decision task, this was always done at the beginning of a session, so that responses would not be affected by immediately preceding tasks. In addition, participants did not perform further repetitions of the phoneme monitoring and stem completion tasks in Session 4, as we wanted to examine changes in lexical competition in Session 5 as a function of passage of time and/or sleep, which could not be attributable to the additional exposure provided by the phoneme monitoring and stem completion tasks.

All tasks were run using DMDX software (Forster & Forster, 2003), using Beyerdynamic DT 234 Pro headsets for auditory presentation of stimuli and recording of responses in the stem completion task. Participants used a USB gamepad to respond in every task except stem completion. Sessions with two tasks took approximately 15–20 minutes and approximately 35 minutes for all four tasks.

Phoneme monitoring. Participants had to pay attention to the phonological form of the 20 novel words that each participant learned and monitor which phonemes they contained to provide repeated exposure to the novel words. Five target phonemes were used (/n/, /t/, /d/, /s/ or /p/, in separate blocks). Phoneme targets appeared across all positions of the novel words, with no particular focus on a part of the word. At the start of each block the target

Table 1
Example Schedule for a 09:00 Start in Experiment 1

Session	Time	Day	Tasks (in test order)			
S1	09:00	1	Phoneme monitoring	Stem completion		
S2	11:30	1	Lexical decision	Familiarity decision	Phoneme monitoring	Stem completion
S3	14:00	1	Lexical decision	Familiarity decision	Phoneme monitoring	Stem completion
S4	16:30	1	Lexical decision	Familiarity decision		
S5	16:30	2	Lexical decision	Familiarity decision	Phoneme monitoring	Stem completion

phoneme was specified, with example existing words given in which that phoneme appears (e.g., "listen for the t sound, as in caT or Town"). This was done to encourage a phoneme encoding strategy for the target phonemes rather than letter matching. The participant was asked to indicate whether the target sound was present in the novel word by responding as quickly and accurately as possible, using the left button on the joypad for phoneme absent and the right button for phoneme present. On each trial the target word was presented over headphones (e.g., *cathedruke*) concurrent with the onscreen presentation of a reminder of how to respond in each block (e.g., "no t" appearing on the left, and "t" on the right). Trials timed out after 4,000 ms. To help improve participants' motivation, we reported the mean response time (RT) and number of errors at the end of each block. At the end of the task we reported the overall accuracy and mean RT. Participants were also given visual feedback on correct responses and errors (a "smiley face" or sad face), with additional auditory feedback on errors, hearing an *uh-oh* sound (which may be familiar to viewers of the U.K. television show *Family Fortunes*).

The ordering of blocks was randomized per participant and per session. Within each block, participants made a decision to each of the 20 novel words twice, with the order of presentation of each set of 20 items randomized. Overall, participants encountered 10 presentations of each novel word in each phoneme monitoring session (2 per block across 5 blocks/phonemes).

Stem completion. Participants carried out a stem completion task to provide a measure of explicit recall and at the same time provide further opportunity for strengthening novel word representations. Participants were told that on each trial they would hear the initial part of a novel word they had been learning as a cue. Their task was to say out loud the matching word, if they could remember it, or otherwise to say nothing. Cue sequences were presented over headphones and constituted the initial CV or CCV of a target newly learned word (e.g., for the novel word *cathedruke*, the cue was *ca*). These cues were excised from the novel word recordings. Participants were instructed to vocalize their responses as quickly and accurately as possible. On each trial a fixation cross appeared in the center of the screen, and simultaneously a 240-ms alert sound was presented to prepare participants for the cue presentation. After 1,000 ms, the cross disappeared and the cue sound was presented. After 4,000 ms, participants were given auditory feedback of the novel word that they should have produced, regardless of whether their response was correct or not. Participants received two practice trials at the beginning with existing words to familiarize themselves with responding. They could skip these trials in later repetitions of the task.

Familiarity decision. To assess recognition of the newly learned words, we had participants make speeded old/new decisions to auditory presentations of target novel words pseudorandomly intermixed with the same number of foils (20 targets and 20 foils) that differed from target novels words only by their final phoneme. Participants pressed the left button on a joypad for a stimulus they thought they had not encountered in training and pressed the right button for a stimulus that they thought they had heard previously. Participants were instructed to be as "quick as possible without making too many mistakes." Two counterbalanced lists were created so that targets and foils appeared in the first or second half of each task in different orders. Trials timed out after 3,000 ms, and were separated by a 1,000-ms interval in which a blank screen was presented. There was a break at the halfway point. Our previous studies have used a two-alternative forced choice task (2AFC) to assess recognition memory (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003). The task differs here to avoid ceiling effects that can occur with the 2AFC task.

Lexical decision. An auditory lexical decision task was used to index the participation of the novel words in lexical competition during the recognition of similar-sounding words. We compared participants' responses to the 20 test base words for which a novel competitor had been presented during training with their responses to the 20 control base words for which the novel competitor had not been presented. The base words were encountered only in the lexical decision task. Along with these key items, 28 real word fillers with properties similar to the base words and 68 nonwords were included, leading to a 50/50 ratio of words to nonwords. Participants were instructed that on each trial they should press a button to indicate whether they thought an item was a word or not. They were to respond as quickly and accurately as possible, with a left response on the joypad indicating a nonword and a right response on the joypad indicating a real word. Participants were given visual feedback for 500 ms after each trial, with either their RT in ms or the text *too slow,* if their RT was longer than 1,200 ms, appearing in the center of the screen. The first 12 items were practice items (half words and half nonwords, with similar characteristics to the main items), and participants had to exceed 80% accuracy on these items to continue. They could skip the practice trials in subsequent repetitions of the task. Trials timed out after 3,000 ms and were separated by a 1,000-ms interval in which a blank screen was presented. There was a break at the halfway point.

Results

In the analyses of variance (ANOVAs) reported below, word list (2 levels) was included as a dummy variable to increase statistical power (Pollatsek & Well, 1995). For lexical decision and familiarity decision tasks we also controlled for which order participants saw the stimuli across sessions, which was included as a dummy variable. Main effects and interactions involving these variables are not reported. As we used phoneme monitoring primarily for training, we do not report the results in detail here. However, descriptive statistics are provided for RTs in Table 2 and errors in Table 3. These show that error rates were low, with a mean of 5.7% for the first session, reducing to just below 4% for the final session, which was coupled with a reduction in RT across sessions. Very similar phoneme monitoring performance was found in the subsequent experiments.

Stem completion. We determined vocalization onsets using manual inspection of the waveform and spectrogram for each utterance. Errors were counted as either omissions or incorrect productions, with the latter typically involving the final syllable being replaced by the final syllable of another novel word or the final syllable of the base word. Overall, omissions and incorrect productions each constituted approximately 25% of the total responses, with omissions shifting from 47% of responses in the first session to only 5% by the time of the final session, whereas incorrect productions shifted from 32% of responses to 14%.

The data from two participants were not included due to recording errors in at least one of their sessions, leaving 34 participants.

Table 2
Mean Latencies (in ms) and Standard Deviations (in Parentheses) for Experiments 1, 2, and 3

Session	Phoneme monitoring	Stem completion	Familiarity decision	Lexical decision (base words)	Lexical decision (controls)
Experiment 1					
S1	1,171 (204)	1,105 (425)			
S2	950 (164)	933 (258)	1,156 (97)	811 (73)	812 (56)
S3	934 (147)	931 (208)	1,145 (112)	796 (56)	776 (54)
S4			1,118 (106)	787 (77)	771 (69)
S5	814 (124)	713 (176)	977 (96)	783 (76)	747 (68)
Experiment 2					
S1	1,320 (219)	1,098 (263)			
S2	1,218 (239)	1,044 (264)			
S3	1,125 (248)	1,100 (243)			
S4			1,115 (115)	840 (56)	843 (40)
S5	840 (117)	753 (141)	979 (89)	830 (65)	814 (69)
Experiment 3					
S1–1	1,195 (164)	1,126 (300)			
S1–2	1,075 (191)	976 (199)			
S1–3	989 (189)	970 (260)			
S2			1,091 (144)	815 (60)	830 (70)
S3			1,120 (86)	803 (69)	791 (58)
S4			1,114 (117)	804 (72)	783 (61)
S5	839 (132)	720 (166)	1,002 (88)	788 (56)	762 (78)

Production latencies greater than 3,000 ms were omitted (< than 1% of total data). RTs are shown in Table 2, error performance is shown in Table 3, and the results are graphed in Figure 1. Recall was poor in the initial session (79% errors) but improved dramatically by the time of the final session (20% errors). We interpret this as being due partly to improvements in associating the cue with the target (aided by the feedback) but also due to a strengthening of the target representation through further training.

Unsurprisingly, given these large shifts in performance, analysis of error rates showed a highly significant effect of session, $F1(1, 33) = 45.9$, $p < .001$, $MSE = 10{,}399$, $\eta^2_G = .34$; $F2(1, 38) = 264.8$, $p < .001$, $MSE = 23{,}438.17$, $\eta^2_G = .64$. The error data showed a substantial decrease in errors between S1 and S2, $t1(33) = 7.7$, $p < .001$, $d = 1.3$,[1] $t2(39) = 10.98$, $p < .001$, $d = 1.74$; a small and marginally significant improvement between S2 and S3, $t1(33) = 2.3$, $p = .029$, $d = 0.39$; $t2(39) = 2.6$, $p = .013$, $d = 0.41$; and another very large reduction overnight from S3 to S5, $t1(33) = 13$, $p < .001$, $d = 2.2$; $t2(39) = 14.58$, $p < .001$, $d = 2.31$.

Due to the large change in error rate across sessions, a statistical analysis of correct response production latencies would have led to a severely unbalanced design. Numerically, however, production latencies for correct responses showed a similar pattern to error data, with large reductions in latency (though with large variability) between S1 and S2 ($d = 0.39$), no change between S2 and S3 ($d = 0.01$), and substantial change between S3 and S5 ($d = 1.16$).

Familiarity decision. We analyzed performance in the familiarity decision task using signal detection analysis, calculating d' for each subject ($Z_{\text{hits}} - Z_{\text{false alarms}}$; Macmillan & Creelman, 2005). Trials with RT less than 300 ms or greater than 2,500 ms were excluded (<1% of the data). Very similar results were obtained using standard accuracy measures (percentage errors). RTs for correct responses are shown in Table 2, and d' scores and percentage errors for targets ("misses") are shown in Table 3. Latencies and d' scores are both graphed for all sessions across our experiments in Figure 2. The initial d' scores of S2 indicate that early on in training participants had sufficiently strong phonological representations to discriminate between similar sounding versions of the learned forms, given that values over 1 for d' are typically taken to indicate good performance.

An ANOVA on d' scores revealed discrimination between learned words and foils changed across sessions, $F1(3, 34) = 14.1$, $p < .001$, $MSE = .0610$, $\eta^2_G = .27$; $F2(3, 38) = 13$, $p < .001$, $MSE = .0656$, $\eta^2_G = .12$. There were no significant differences in mean d' between S2 and S3, $t1(35) = 1$, $d = 0.16$; $t2(39) < 1$, $d = 0.10$, or between S3 and S4, $t1(35) = 1.3$, $d = 0.21$; $t2(39) < 1$, $d = 0.12$, but there was a highly significant increase in discrimination between S4 and S5, $t1(35) = 4.75$, $p < .01$, $d = 0.79$; $t2(39) = 3.6$, $p < .01$, $d = 0.46$. These increases in d' scores showed that participants improved their discrimination abilities across sessions, which we again attribute to task-related improvements coupled with strengthening representations for the novel words.

We also analyzed RTs for correct responses to target novel words ("hits"), which revealed that along with making more accurate decisions, correct decisions to targets were quicker across sessions, $F1(3, 34) = 30.6$, $p < .001$, $MSE = 187{,}979$, $\eta^2 = .28$;

[1] There is considerable disagreement in the literature as to what is the correct denominator for calculation of the standardized effect size measure commonly referred to as Cohen's d, both for between-participants and within-participants designs. Here we use the mean difference score across each condition divided by the standard deviation of the difference score (Morris & DeShon, 2002). Assuming a normally distributed effect, an effect size of 1.64 suggests that 95% of the population would show a change in that direction. This within-participants measure can lead to higher values when compared with a between-participants calculation (Dunlap, Cortina, Vaslow, & Burke, 1996), particularly when the correlation between tests is high. Between-subjects effect sizes can be calculated from the means and standard deviations provided in the tables.

Table 3
Mean Percentage Errors and Standard Deviations (in Parentheses) for Experiments 1, 2, and 3, With Additional d′ Scores for Familiarity Decision

Session	Phoneme monitoring	Stem completion	Familiarity decision	Familiarity decision (d′)	Lexical decision (base words)	Lexical decision (controls)
Experiment 1						
S1	5.7 (3.1)	79 (16)				
S2	5.1 (2.6)	57 (22)	20.5 (9.1)	1.68 (.64)	6.3 (12.5)	6.4 (9)
S3	4.5 (2.4)	51 (23)	19 (9.3)	1.79 (.69)	5.5 (7)	5.5 (8.3)
S4			17.3 (9.9)	1.93 (.78)	4.9 (6.9)	5.1 (7.4)
S5	3.9 (2.5)	20 (15)	11 (9.1)	2.53 (.87)	7 (6.8)	4.9 (7.5)
Experiment 2						
S1	5.25 (2.4)	74 (20)				
S2	4 (2.4)	56 (23)				
S3	3.7 (2.1)	52 (27)				
S4			12.1 (6.9)	2.37 (.67)	5.8 (7.6)	5.1 (9.7)
S5	4 (1.9)	20 (2)	6.1 (6)	3.02 (.71)	6.4 (8.1)	4.4 (4.6)
Experiment 3						
S1–1	5.6 (3)	79 (15)				
S1–2	3.6 (2.3)	48 (20)				
S1–3	3.5 (2.2)	34 (19)				
S2			15.9 (13.7)	2.13 (.96)	3.9 (3.9)	2.9 (4.2)
S3			17.8 (8.7)	1.90 (.63)	4.3 (3.8)	2.3 (3.7)
S4			18.3 (11.3)	1.82 (.90)	4.4 (5.1)	2.9 (4.4)
S5	4.1 (2.8)	21 (17)	12.7 (10.8)	2.45 (.94)	5.7 (5.3)	3.9 (4.7)

$F2(3, 38) = 111.08$, $p < .001$, $MSE = 233{,}019$, $\eta^2_G = .31$. As with discrimination scores, there was no significant improvements in performance on the first day, between S2 and S3, $t1(35) < 1$, $d = 0.09$; $t2(39) < 1$, $d = 0.01$, or between S3 and S4, $t1(35) = 1.22$, $p = .16$, $d = 0.2$; $t2(39) = 1.8$, $p = .08$, $d = 0.29$, but there was a large and significant reduction in RT between the last session of Day 1 and the final test on Day 2 (S4-S5), $t1(35) = 8.03$, $p < .001$, $d = 1.33$; $t2(39) = 13$, $p < .001$, $d = 2.1$.

Lexical decision. RTs to the base words in the lexical decision tasks were analyzed to examine the extent to which acquisition of the novel items led to heightened lexical competition in the recognition of well-established words. Our measure of lexical competition was an RT comparison between the 20 test base words for which a novel competitor had been learned and the 20 matched, counterbalanced control base words for which no novel competitor had been presented. Emergence of lexical competition should be associated with slower RTs to the test items relative to the controls (cf. Gaskell & Dumay, 2003). This contrasts with the other two tasks, as here evidence for the development of new lexical representations is shown by deterioration in performance compared with controls, as potential performance improvements due to task and item repetition are counteracted (or even reversed) by the slowing down of the test item processing due to heightened lexical competition.

Trials with RT less than 400 ms or greater than 2,000 ms were excluded (< 1% of the data). Mean RTs are shown in Table 2. As can been seen from Table 2, RTs start off at a similar level for both test and control words, and latencies consistently reduce across sessions both word types, but this change appeared greater for the controls. An ANOVA with factors of session and word type (test vs. control base words) revealed an interaction, $F1(3, 102) = 5.24$, $p < .001$, $MSE = 7{,}645$, $\eta^2_G = .01$; $F2(3, 117) = 4.2$, $p < .001$, $MSE = 14{,}828$, $\eta^2_G = .001$, indicating that the size of the lexical competition effect changed over time. The size of the mean lexical competition effect in ms for each session in Experiment 1 (and the other two experiments) is illustrated in Figure 3. Initially there was no difference between test base words and controls (S2; $t1$ and $t2 < 1$, $d < 0.01$). However, at the second test (S3), we found a significant lexical competition effect of 20 ms ($SD = 41$ ms), $t1(35) = 2.9$, $p < .01$, $d = 0.45$; $t2(39) = 3.3$, $p < .01$, $d = 0.52$. At the end of the first day (S4) there was a slightly weaker effect that was significant by participants but not by items ($M = 16$ ms, $SD = 40$ ms), $t1(35) = 2.4$, $p = .022$, $d = 0.39$; $t2(39) = 1.6$, $p = .12$, $d = 0.25$. In the final session the next day, the lexical competition effect increased to 37 ms, $t1(35) = 5.78$, $p < .001$, $d = 0.96$; $t2(39) = 4.5$, $p < .001$, $d = 0.71$. The change between these last two sessions was significant by participants, $t1(35) = 2.2$, $p = .036$, $d = 0.36$, and marginally significant by items, $t2(39) = 1.9$, $p = .065$, $d = 0.34$. The reliable effect found in the second lexical decision task (S3) suggests that given certain circumstances, a lexical competition effect can emerge without the need for sleep.

We also found a main effect of session, $F1(3, 35) = 13.8$, $p < .001$, $MSE = 52{,}543$, $\eta^2_G = .05$; $F2(1, 39) = 16.9$, $p < .001$, $MSE = 18{,}022$, $\eta^2_G = .04$, with Table 2 indicating participants were generally speeding up regardless of word type. This increase was most marked between S2 and S3 and from S4 to S5. There was also a main effect of word type, with RTs to base words slower than to control words overall, $F1(1, 35) = 25.5$, $p < .001$, $MSE = 43{,}229$, $\eta^2_G = .02$; $F2(1, 39) = 15.1$, $p < .001$, $MSE = 85{,}558$, $\eta^2_G = .016$. Analysis of errors (shown in Table 3) revealed no significant differences across session or word type nor an interaction between them (Fs < 1).

Discussion

Experiment 1 addressed whether sleep was a necessary condition for the integration of phonological forms into lexical memory,

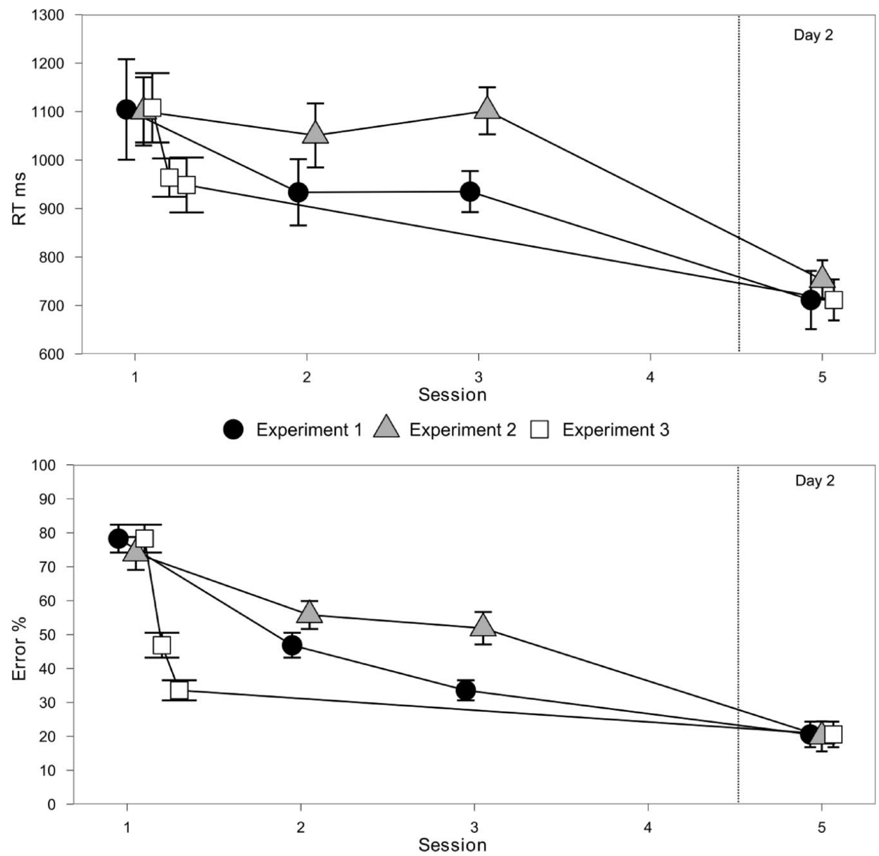

Figure 1. Mean RT (top) and errors (bottom) across experiments for stem completion. Error bars show within-participant 95% confidence intervals (Loftus & Masson, 1994). RT = response time.

given previous results implying a role of sleep in this process (Dumay & Gaskell, 2007; Tamminen et al., 2010). Results from the lexical competition task showed reliable evidence for emergence of lexical competition in the second lexical decision task on Day 1 (S3), indicating that in some circumstances lexical integration can occur within a single day. What then are the circumstances that allowed such integration?

One possibility is that the enhanced training on the novel words may be the reason that lexical competition effects were observed in two sessions on the first day of exposure (after 5–8 hours). Experiment 1 combined spaced learning and testing of the novel words with exposure to and testing on the neighboring base words. In many ways, the learning regime was near optimal, with both spacing of exposure sessions (in the phoneme monitoring tests) and spaced retrieval practice enhanced by feedback (in the stem completion tests). These conditions should lead to particularly robust representations of the novel words (Karpicke & Blunt, 2011; Roediger & Butler, 2011) in comparison with typical learning circumstances in other experiments (e.g., Gaskell & Dumay, 2003) that provide only massed exposure and no retrieval practice.

An alternative possibility is that the repeated exposure to and/or testing on the existing word neighbors in the lexical decision task may have been important to allow lexical competition to occur on the first day. In terms of a complementary memory systems account, we have argued that the interleaving of new words with old may be crucial to allow lexical integration in a system that guards against catastrophic interference (Davis & Gaskell, 2009; McClelland et al., 1995). Experiment 1 used two forms of interleaving. Participants went about their daily business in between sessions, which in most cases would have involved considerable exposure to language, both in input and output forms. This general interleaving contrasts with the much more specific interleaving provided through exposure to the existing neighbors of the novel words in repeated iterations of the lexical decision task. Conceivably, lexical competition effects for existing words (e.g., *cathedral*) require alteration of the lexical representations of those well-established words, which might occur as a consequence of their activation (alongside the activation of the related novel words such as *cathedruke*) when they are presented during the lexical decision task.

In order to help distinguish between these possible explanations, in Experiment 2 we examined the role of repeated exposure to the related base words by withholding the lexical decision test until the end of the first day. A lexical competition effect at the end of Day 1 in Experiment 2 would suggest that spaced learning and testing of the novel words, alongside more general interleaving, is sufficient to demonstrate lexical competition effects without the need for sleep. Alternatively, a lack of an effect in Experiment 2 would suggest that merely having an optimal training regime with general interleaving is not enough to lead to lexical integration, and that more specific interleaving is necessary.

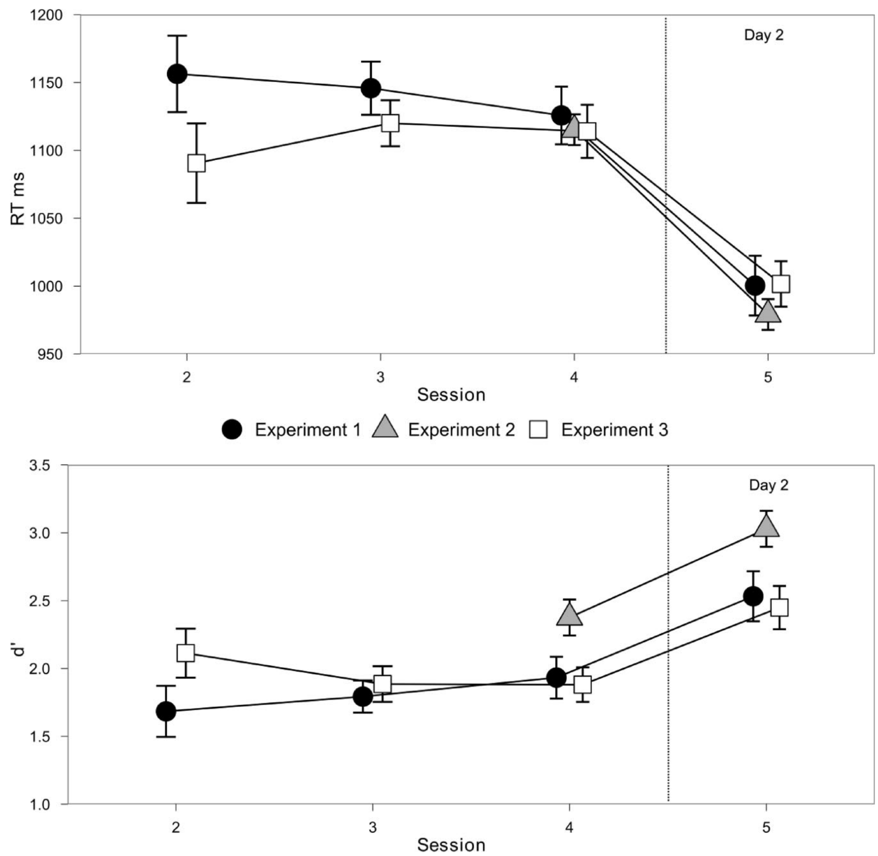

Figure 2. Mean RT (top) and d' (bottom) across experiments for familiarity decision. Error bars show within-participant 95% confidence intervals. RT = response time.

Experiment 2

Method

Participants. Thirty-five participants from the University of Sussex were tested. All were native English speakers without visual or auditory impairments. Participants received £15 for participation. One participant dropped out from the last two sessions, and one was excluded due to a data recording error. The data from these participants were excluded.

Design and procedure. The experiment was identical to Experiment 1 except for the schedule of tasks used in each testing session, illustrated in Table 4. As in Experiment 1, participants had three phoneme monitoring and stem completion tasks spaced out in three sessions during the first day. These sessions were expected to lead to very robust representations of the novel words. However, unlike in Experiment 1, the lexical decision and familiarity decision tasks were withheld until the final session in the late afternoon, as these were primarily intended as tests of learning rather than opportunities to learn. Approximately 24 hours later, participants completed another iteration of the lexical decision and familiarity decision tasks and a further repetition of the phoneme monitoring and stem completion tasks.

Results

Stem completion. Analysis of stem completion was the same as in Experiment 1, with identical exclusion criteria (this was also true for the other tasks). One participant's data were excluded from analysis due to a data recording error in one of the sessions. As in Experiment 1, an ANOVA revealed that correct vocalizations significantly increased over sessions, $F1(1, 33) = 75.7$, $p < .001$, $MSE = 1{,}608$, $\eta^2_G = .46$; $F2(1, 38) = 284.8$, $p < .001$, $MSE = 20{,}104$, $\eta^2_G = .61$. The drop in errors was significant between S1 and S2, $t1(31) = 5.1$, $p < .001$, $d = 0.90$; $t2(39) = 9.7$, $p < .001$, $d = 1.5$, but was nonsignificant between S2 and S3, $t1(31) = 1.19$, $p = .24$, $d = 0.21$; $t2(39) = 1.9$, $p = .06$, $d = 0.3$. The following day there was a large reduction between S3 and S5, $t1(31) = 8.68$, $p < .001$, $d = 1.53$; $t2(39) = 17$, $p < .001$, $d = 2.8$. The improvements in correct vocalizations are shown in Table 3 and Figure 1.

As in Experiment 1, we report only descriptive statistics for RT data due to the unbalanced design. Production latencies (shown in Table 2 and Figure 1) speeded up a small amount between the first and second sessions ($d = 0.21$) and then slowed slightly between S2 and S3 ($d = -0.21$). There was a substantial decrease in latency of vocalization the next day (S3–S5; $d = 2.28$).

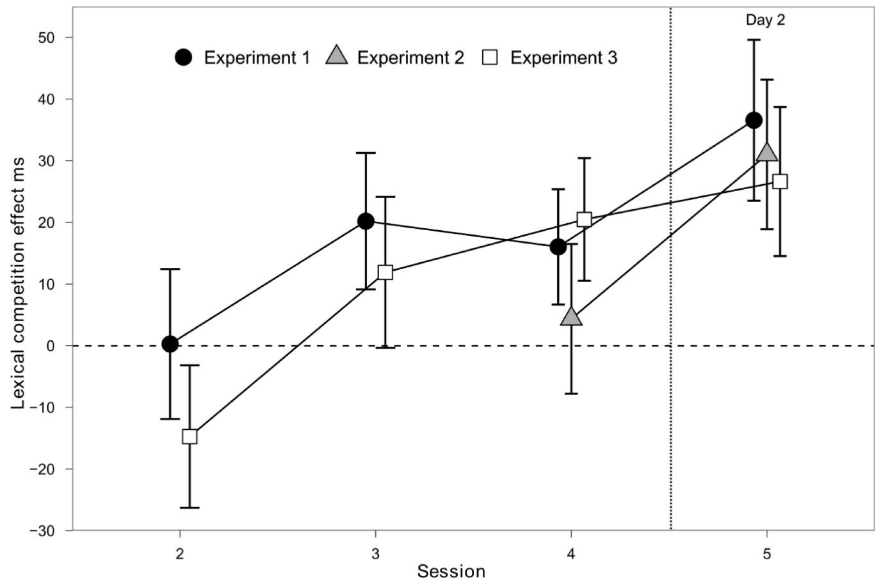

Figure 3. Mean lexical competition difference scores across experiments. Error bars show within-participant 95% confidence intervals.

Familiarity decision. In Experiment 2 we measured familiarity decision responses in only two sessions (S4 and S5), and as expected, discrimination scores (d') improved significantly between these sessions, $t1(32) = 5.01$, $p < .01$, $d = 0.87$; $t2(39) = 5.5$, $p < .001$, $d = 0.62$ (see Table 3 and Figure 2). There was also a large and significant drop in RTs for correct responses between sessions, $t1(32) = 12.2$, $p < .01$, $d = 2.12$; $t2(39) = 15$, $p < .001$, $d = 2.3$ (see Table 2 and Figure 2).

Lexical decision. As with familiarity decision, lexical decision performance was measured in two sessions: S4 at the end of the first day and S5 a day later. There was a significant Session × Word Type interaction, $F1(1, 31) = 5.25$, $p = .029$, $MSE = 11{,}630$, $\eta^2_G = .01$; $F2(1, 39) = 5.29$, $p = .027$, $MSE = 161{,}145$, $\eta^2_G = .01$. In the final session of Day 1, there was no significant difference between control and test base words, ($M = 4$ ms, $SD = 50$ ms), $t1(32) < 1$, $d = 0.08$; $t2(39) < 1$, $d = 0.04$. However, the next day a clear lexical competition effect emerged ($M = 31$ ms, $SD = 59$ ms), $t1(32) = 3.03$, $p < .01$, $d = 0.52$; $t2(39) = 2.77$, $p < .01$, $d = 0.43$. The change between these last two sessions was significant, $t1(35) = 2.2$, $p = .023$, $d = 0.38$, $t2(39) = 2.6$, $p < .02$, $d = 0.41$. The size of the lexical competition effect at each session is illustrated in Figure 3. As can be seen from Table 2, RTs for the control items sped up a day later, but no speedup was seen for the base words for which a novel competitor was learned. Unlike in Experiment 1, we failed to find any evidence for lexical competition on the first day.

Analyses of the main effects showed that responses were marginally faster overall on the second day, $F1(1, 31) = 3.04$, $p = .09$, $MSE = 15{,}069$, $\eta^2_G = .01$, $d = 0.28$; $F2(1, 39) = 6.72$, $p = .013$, $MSE = 10{,}162$, $\eta^2_G = .003$, $d = 0.36$, and that base words were slower overall, $F(1, 31) = 5.04$, $p = .03$, $MSE = 16{,}048$, $\eta^2_G = .01$, $d = 0.42$; $F2(1, 39) = 4.3$, $p = .043$, $MSE = 15{,}901$, $\eta^2_G = .01$, $d = 0.38$. Analysis of errors showed no significant effects (Fs < 1; see Table 3).

Discussion

Unlike Experiment 1, in Experiment 2 we did not find any evidence for the influence of novel items on recognition of their phonologically neighboring words on the day of learning. Despite the memory advantages from spaced learning and testing of the novel words, the phonological representations developed over the course of the day were still not in a form sufficient to demonstrate lexical competition, suggesting the novel words were not fully integrated into the mental lexicon. Only after a 24-hour period that included a night's sleep did lexical competition effects occur, without any further exposure to the novel words during that interval.

Table 4
Example Schedule for a 09:00 Start in Experiment 2

Session	Time	Day	Tasks (in test order)			
S1	09:00	1	Phoneme monitoring	Stem completion		
S2	11:30	1	Phoneme monitoring	Stem completion		
S3	14:00	1	Phoneme monitoring	Stem completion		
S4	16:30	1	Lexical decision	Familiarity decision		
S5	16:30	2	Lexical decision	Familiarity decision	Phoneme monitoring	Stem completion

Looking at the performance across the tasks and experiments, it does not appear that the lack of a lexical competition effect in Experiment 2 was due to the novel words having weaker representations. In the recognition memory test by the time of the fourth session in Experiment 1, participants had done the familiarity decision task three times. In Experiment 2 this task was done for the first time in Session 4, but performance was roughly the same (with equivalent RT and in fact fewer errors). Percentage of correct vocalizations in the stem completion task was also very similar across sessions on the first day in Experiment 1 and 2. This suggests that the spacing regime of Experiment 2 was sufficient to generate strong representations by the final session of the first day, at least as indexed by recognition memory and cued recall. Despite this, lexical competition effects were not found in Experiment 2. The implication from this result in combination with the results of the same test in Experiment 1 is that the exposure to the existing words was important in Experiment 1 in the establishment of a lexical integration effect by the end of the first day. This fits with the notion that explicit interleaving of novel and existing words facilitates the generation of the inhibitory links between the two that are required for swift lexical competition.

Turning to the finding that lexical competition effects were obtained in the final session on Day 2, there are multiple explanations to consider. The emergence of this effect is unlikely to be due to the minimal additional exposure caused by administering the familiarity decision test in Session 4. Gaskell and Dumay (2003) compared lexical decision followed by a 2AFC recognition task (similar to our familiarity decision task) with lexical decision without a subsequent recognition task and found the additional exposure and testing involved from the recognition task did not have any effect the emergence of lexical competition. Two possible explanations are more likely. One is that the additional exposure to the existing words in the final session of Day 1 was crucial, just as it was in the earlier tests during the day in Experiment 1. The second possibility is that the extended consolidation time, which included nocturnal sleep, provided an opportunity for offline interleaving, as has been found in previous experiments. It is important to note that explicit prior presentation of the base words is not required for effects of lexical competition after sleep. Davis et al. (2009) found that lexical competition effects can still be demonstrated using only a single lexical competition test but only when 24 hours has elapsed between training and testing. Their method contrasted two sets of novel words learned at different times. One set was learned on the first day. On the second day, another set of words was learned in the morning. Both sets of words were then tested in a single session later in the second day. They found only lexical competition effects for the set of words learned the previous day.

Although repeated tests of lexical competition are not necessary to bring about lexical integration (Davis et al., 2009), it appears the most significant difference between Experiment 1 and Experiment 2 was the testing using the lexical decision task spaced throughout the day in Experiment 1. This task provides exposure to the base words to which the novel competitors were derived from, and it suggests the interleaving (whether explicit or not) may be a crucial factor in bringing out pre-sleep lexical integration. In order to further test this possibility, we spaced out our lexical competition tests through the day, along with the familiarity decision task, in Experiment 3 as was done in Experiment 1. But in Experiment 3, instead of spacing out exposure and testing on the novel words throughout the day using phoneme monitoring and stem completion, we gave participants the equivalent amount of exposure and testing in the phoneme monitoring and stem completion tasks concentrated in a single massed session at the beginning of the day. If the crucial difference between Experiment 1 and 2 was the repeated exposure to the existing base words in the lexical decision task in Experiment 1, then we should expect to find lexical competition from the novel words on Day 1 in Experiment 3. Note that traditionally, massed learning refers to a within-list manipulation, where spaced exposure is contrasted with massed exposure, with massed items exposed once and spaced items exposed for the same amount of total time but distributed over intervals with exposure to other items. On that definition, in our "massed" single session we still have spaced exposure to items over time, but the time interval is over the order of minutes rather than the 2.5-hour intervals of Experiments 1 and 2.

Experiment 3

Method

Participants. Thirty-eight participants were recruited from the University of Sussex. All native English speakers without visual or auditory impairments. They received £15 for participation. During the course of the study two participants failed to complete all sessions, leaving data for 36 participants.

Design and procedure. The experiment was identical to Experiment 1, except that this time phoneme monitoring and stem completion tasks were not spaced out across the day (see Table 5). Instead, we used a massed training regime in the first session, where participants completed three sessions' worth of phoneme monitoring and stem completion tasks, with each stem completion phase (where a phase is equivalent to one session in previous experiments) following a phoneme monitoring phase. We then tested participants using lexical decision and familiarity decision

Table 5
Example Schedule for a 09:00 Start in Experiment 3

Session	Time	Day	Tasks (in test order)			
S1	09:00	1	Phoneme monitoring × 3	Stem completion × 3		
S2	11:30	1	Lexical decision	Familiarity decision		
S3	14:00	1	Lexical decision	Familiarity decision		
S4	16:30	1	Lexical decision	Familiarity decision		
S5	16:30	2	Lexical decision	Familiarity decision	Phoneme monitoring	Stem completion

in three spaced sessions without further training. A final session on Day 2 used all four tasks.

Results

Stem completion. As found in the previous experiments, an ANOVA showed a significant reduction in vocalization errors across task phases (shown in Table 3 and Figure 1), $F1(1, 31) = 126$, $p < .001$, $MSE = 22{,}112$, $\eta_G^2 = .60$; $F2(1, 38) = 284.8$, $p < .001$, $MSE = 20{,}104$, $\eta_G^2 = .62$. The drop in errors in the first session between the first and second phase was significant, $t1(35) = 10$, $p < .001$, $d = 1.7$; $t2(39) = 17$, $p < .001$, $d = 2.62$, as was the reduction between the second and third phase, $t1(35) = 6.13$, $p < .001$, $d = 1$; $t2(39) = 8.7$, $p < .001$, $d = 1.38$. There was another large reduction between the third phase of the first session and the final session on the second day, $t1(35) = 5.25$, $p < .001$, $d = 0.88$; $t2(39) = 7.4$, $p < .001$, $d = 1.16$.

Speed of productions (see Table 2 and Figure 1) dropped between the first and second phase within the first session ($d = 0.53$) and then stayed roughly the same for the third phase ($d = 0.08$). In the final session (S5), latencies substantially reduced compared with Phase 3 of Session 1 ($d = 1.1$).

Familiarity decision. An ANOVA on d' scores revealed that discrimination between learned words and foils changed across sessions, $F1(3, 35) = 14$, $p < .001$, $MSE = .061$, $\eta_G^2 = .07$; $F2(3, 39) = 4.9$, $p < .01$, $MSE = .02367$, $\eta_G^2 = .05$. There was a marginal by-participants improvement in mean d' between S2 and S3, $t1(35) = 1.98$, $p = .054$, $d = 0.32$; $t2(39) = 1.1$, $p = .28$, $d = 0.12$. Although there was no significant improvement between S3 and S4, $t1(35) < 1$, $d = 0.01$; $t2(39) < 1$, $d = 0.04$, there was a significant increase in d' between the last two sessions, S4 and S5, $t1(35) = 4.8$, $p < .01$, $d = 0.94$; $t2(39) = 3.5$, $p < .01$, $d = 0.37$. The accuracy scores and RT shown in Table 3 (and seen in Figure 2) reveal that participants had better performance in the first test (S2) compared with that in Experiment 1, which would be expected given the additional training in S1 of Experiment 3.

RT for correct decisions differed significantly across sessions (shown in Table 2 and Figure 2), $F1(3, 34) = 20.5$, $p < .001$, $MSE = 108{,}157$, $\eta_G^2 = .14$; $F2(3, 38) = 90.5$ $p < .001$, $MSE = 133{,}738$, $\eta_G^2 = .31$. There was no significant change between the first and second repetitions of the task (S2–S3), $t1(35) = 1.43$, $p = .16$, $d = 0.23$; $t2(39) = -1.9$, $p = .07$, $d = -0.29$. Between S3 and S4 there was a small reduction in RT, which was significant by items but not by participants, $t1(35) < 1$, $d = 0.07$; $t2(39) = 2.2$, $p = .037$, $d = 0.34$, and there was a significant reduction in RT between the last two sessions (S4–S5), $t1(35) = 8.23$, $p < .001$, $d = 1.37$; $t2(39) = 9.4$, $p < .001$, $d = 1.49$.

Lexical decision. There was a significant Session × Word Type interaction, $F1(3, 99) = 7.23$, $p < .001$, $MSE = 11{,}630$, $\eta_G^2 = .01$; $F2(3, 117) = 9.25$, $p < .001$, $MSE = 17{,}459$, $\eta_G^2 = .01$. As shown in Figure 3 and Table 2, no lexical competition effect was present at the first test point; in fact, a reverse competition effect was found that was significant by items but not by participants (S2; $M = -15$ ms, $SD = 56$ ms), $t1(35) = -1.5$, $p = .12$, $d = -0.26$; $t2(39) = 2.2$, $p = .032$, $d = -0.35$. At the second test point, a lexical competition effect was significant by items but not by participants, (S3; $M = 12$ ms, $SD = 53$ ms), $t1(35) = 1.35$, $p = .18$, $d = 0.22$; $t2(39) = 2.5$, $p < .02$, $d = 0.39$. By the final session of the first day, a significant lexical competition effect emerged both by participants and by items (S4; $M = 20$ ms, $SD = 43$ ms), $t1(35) = 2.83$, $p < .01$, $d = 0.47$; $t2(39) = 3.4$, $p < .01$, $d = 0.53$. The next day this effect was still present ($M = 27$ ms, $SD = 47$ ms), $t1(35) = 3.40$, $p < .01$, $d = 0.56$; $t2(39) = 4.2$, $p < .001$, $d = 0.65$. Unlike in the previous two experiments, the overnight shift was not significant, $t1(35) < 1$, $d = 0.11$; $t2(39) < 1$, $d = 0.08$.

RTs were significantly different across sessions, $F1(1, 33) = 8.54$, $p < .001$, $MSE = 54{,}164$, $\eta_G^2 = .05$; $F2(1, 39) = 23.8$, $p < .001$, $MSE = 7{,}643$, $\eta_G^2 = .04$, reflecting the general speeding up across sessions. Base words were also slower overall, $F1(1, 33) = 1$, $p < .001$, $MSE = 136{,}862$, $\eta_G^2 = .01$; $F2(1, 39) = 12.3$, $p < .01$, $MSE = 22{,}896$, $\eta_G^2 = .01$. The only significant effect involving errors (see Table 3) was a main effect of more errors overall for base words compared with controls, $F1(1, 33) = 5.65$, $p = .023$, $MSE = .02060$, $\eta_G^2 = .01$; $F2(1, 33) = 6.38$, $p = .015$, $MSE = .0217$, $\eta_G^2 = .01$.

The results from the lexical competition test in Experiment 3 are broadly similar to those from Experiment 1. Again, there is clear evidence that lexical competition effects can emerge within a single day, given the right conditions of exposure. This effect took longer to emerge compared with Experiment 1, occurring only in S4, where it was present at the previous test point in Experiment 1, and the overnight shift was not as large as the previous two experiments. Performance in the other tasks was comparable with that in the other two experiments.

General Discussion

Across three experiments we presented participants with different training and testing regimes to determine the circumstances in which novel words become integrated into our mental lexicons. We took as our measure of lexical integration the extent to which novel words could participate in lexical competition with preexisting neighbors, demonstrated by a slowing down of lexical decisions to the existing words. Previous studies have shown that lexical competition effects reliably occur only following a period of offline consolidation that includes sleep (e.g., Davis et al., 2009; Dumay & Gaskell, 2007; cf. Tamminen et al., 2010), indicating sleep to be a sufficient condition for lexical integration. In two of the three experiments here we demonstrated that sleep is not a necessary condition for lexical integration. In Experiment 1, we found lexical competition effects on Day 1 with a training regime that combined exposure and retrieval practice for novel words with repeated testing on the neighboring existing words. In Experiment 2, we spaced out novel word exposure and retrieval across the day but tested for lexical competition only at the end of the day, and we found no evidence for lexical competition on the first day. During Experiment 3, participants were trained on novel words in a single intensive session in the morning, and following spaced testing on the existing words, a lexical competition effect occurred at the end of the first day. In all experiments lexical competition effects were present the following day. We now describe the theoretical interpretation of our results.

Cross-Experiment Comparison

The most important question for understanding our pattern of results is why pre-sleep lexical competition effects were found in Experiments 1 and 3 but not in Experiment 2 or in other studies (e.g., Dumay & Gaskell, 2007). Several previous studies showing an absence of pre-sleep lexical competition effects have used a single (massed) exposure session (Davis et al., 2009; Dumay & Gaskell, 2007, 2012; Dumay et al., 2004; Tamminen & Gaskell, 2008). One of the reasons for using spaced learning and testing of the novel words was to help develop stronger novel word representations than might be achieved with just single-session massed learning. Perhaps engagement in lexical competition prior to sleep simply requires a sufficiently well-learned lexical representation. The results of Experiment 1 fit with this explanation, in that enhanced learning of the novel words via spaced exposure and spaced testing with retrieval practice led to lexical competition effects in two sessions on the same day. However, this hypothesis is clearly contradicted by the data from Experiments 2 and 3. The same training regime was retained in Experiment 2, but the within-day competition effect went away, whereas Experiment 3 reverted to massed training, and the within-day effect returned. Furthermore, looking across task performance across experiments (seen in Figures 1–3) it does not appear that the novel words were better learned in the experiments that showed within-day competition effects. In particular, looking at the results of Experiment 2 in comparison to the other two experiments, by the time of the lexical competition test in the final session of the first day, lexical representations would appear to have been well formed. Despite this, lexical competition was not found until the next day. In other studies, we have used a 2AFC task to assess explicit knowledge of word forms, and despite near ceiling performance on this task, we previously did not find good evidence for pre-sleep competition (e.g., Dumay & Gaskell, 2007; Gaskell & Dumay, 2003). On the basis of these considerations, the possibility that the strength of novel form representations modulates lexical competition effects with novel words must be rejected.[2] Although spaced learning may benefit memory for the novel items themselves in the long term, it appears that spaced learning of novel words in the absence of exposure to the related existing words does not facilitate their integration with lexical knowledge prior to sleep.

Interleaving and Novel Word Learning

Our initial prediction of pre-sleep competition effects in Experiment 1 was inspired by the complementary learning systems approach of McClelland et al. (1995). In particular, we pursued the idea that interleaved learning and testing, with repeated exposure to both the novel words and their existing phonological competitors, may provide an online alternative to the offline process of consolidation thought to occur during sleep and hence allow for pre-sleep lexical integration. Our results provide tentative support for this explanation. The within-day integration effect in Experiment 1 can be explained in terms of a general process of interleaving of novel words with nonlaboratory discourse or, more specifically, as a consequence of interleaving novel words with their immediate competitors. Experiment 2 ruled out the more general interleaving account by showing that withholding the interleaved presentation of the existing word neighbors eliminated the integration effect at the end of the day. It appears that the repeated activation of the existing base words in conjunction with exposure to the phonologically related novel words may be crucial to determining the time course of lexical competition found in this study. However this account is somewhat at odds with the results of Experiment 3, in which interleaving between existing and novel words was limited in comparison with Experiment 1 but a lexical competition effect still emerged, although this time only in the final session of Day 1. Nonetheless, despite the relatively limited interleaving of novel words in Experiment 3, some interleaving still occurred. Participants were exposed to the novel items in the familiarity task that followed each lexical decision task, and the lexical decisions to existing words may still have led to retrieval of the related novel words, which were well learned in the first session (though their retrieval may not have been sufficiently swift or automatic early in the day to bring about slowing down of lexical decisions).

Consolidation and Reconsolidation

A potential mechanism for the effect of exposure to existing representations on lexical competition involves the notion of reconsolidation (Finn & Roediger, 2011; Nader, Schafe, & Le Doux, 2000; Walker, Brakefield, Hobson, & Stickgold, 2003). By this proposal reconsolidation provides a mechanism whereby existing memories are updated with new information (Alberini, 2005; Hupbach, Gomez, Hardt, & Nadel, 2007), though evidence for reconsolidation in humans is not without controversy (Sederberg, Gershman, Polyn, & Norman, 2011). A speculative interpretation of our results is that exposure to the existing items induced plasticity that promoted reorganization of phonological lexical memory, allowing new phonologically related representations to be incorporated. This view fits particularly well with distributed theories of lexical representation and processing, such as the distributed cohort model (Gaskell & Marslen-Wilson, 1997). On such a model, learning a new phonological form can be seen to involve the modification of existing lexical forms rather than the creation of a new and distinct lexical entry. In understanding the results of our Experiment 2, the lack of exposure to the existing forms could have meant modification was more difficult and had to await sleep-dependent consolidation. If spaced testing is able to provide an alternative to sleep for lexical integration, it remains to be explained what differs between online and offline integration. One way of conceptualizing what may occur during offline consolidation is that sleep initiates a form of spaced learning, with the hippocampus training the neocortex by replaying newly acquired memories while the neocortex is offline. Theories of sleep-associated memory consolidation have posited an important role for this kind of replay, based on evidence for reactivation in animals (Skaggs & McNaughton, 1996) and in humans (Rudoy, Voss, Westerberg, & Paller, 2009). There also accumulating evidence that sleep results in a more generalizable or structured form of knowledge (Ellenbogen, Hu, Payne, Titone, & Walker, 2007; Payne et al., 2009). In contrast, the specific interleaving that we have used in our experiments might be better thought of as a more focused or impoverished type of integration that alters the representations of specific lexical items rather than a restructuring of the lexicon as a whole.

---

[2] As an additional confirmation of this point, we carried out a correlation analysis between individuals' performance in the familiarity decision for each session and the size of their lexical competition effect for that session. No correlations were found ($ps > .25$).

The Possible Role of Awareness

Another factor that may be relevant in explaining the patterns of lexical competition across experiments is the level of participants' awareness that the novel words were related to existing words. As participants were exposed to novel words derived from existing words and then exposed to those existing words, it is unsurprising that in debriefing many participants reported explicit awareness of the relationship for at least some of the items, despite our use of fillers in the lexical decision task. We also found in some cases participants produced the base word when given the stem in the stem completion task. Because there were only two sessions containing the lexical decision task in Experiment 2 (with the first one at the end of the day's testing), it is possible that participants were less aware of the relationship between base and novel words in that experiment, compared with the greater exposure to the base words in Experiments 1 and 3. It is unclear to what extent participants' awareness of this relationship impacted upon on our results, but relevant to understanding this issue is the study of Dumay and Gaskell (2012). They taught novel words in which an existing word was embedded; for example, *muck* could be embedded within *lirmucktoze*. Using pause detection to index lexical activation, they found that delayed responses to existing words embedded in novel words were found only after sleep (24 hours after initial training). Results using pause detection tasks (see also Dumay & Gaskell, 2007) have been similar to those involving lexical decision as a measure of lexical activation, but compared with the lexical decision task, pause detection has the advantage that as a lexical status judgment is not made, responses should not be influenced by the lexical status of competing forms. Crucially though, participants in this paradigm are much less likely to detect a relationship between novel words and their base words, which would suggest that explicit awareness of a link is not the key factor that distinguishes the presence or absence of pre-sleep lexical competition in our experiments. Further testament to this claim comes from the study of Davis et al. (2009), in which lexical competition effects were demonstrated using just a single test session (providing a period of time that included sleep occurred between initial exposure and test). This indicated that competition effects can emerge in circumstances where awareness would have been limited. Thus, it is clear that lexical integration via sleep can occur in the absence of awareness of the relationship between novel word and existing neighbor. Furthermore, if awareness of the similarity between novel and existing word (e.g., *cathedruke* and *cathedral*) is influential in lexical decision, perhaps by encouraging a more cautious response strategy to existing words just in case they turn out to be novel words, then apparent competition effects should be found soon after initial exposure to the novel words (while they are fresh in memory). Other studies have tested for inhibitory effects of novel words on lexical decision to base words immediately after learning and have found none (e.g., Davis et al., 2009; Dumay et al., 2004; Gaskell & Dumay, 2003; Tamminen et al., 2010; cf. Tamminen and Gaskell, 2008, for a possible exception). Thus we can be confident that the effects we see here are not solely due to explicit awareness of similarities in form between novel and existing words. Nonetheless, awareness may be a contributory factor (e.g., by promoting reconsolidation of the representation of the existing word). An important goal for future studies (potentially using pause detection instead of lexical decision) will be to determine to what extent inhibitory linking of novel and existing words during wake relies on attentional processing.

Strengthening of Representations and the Time Course of Lexical Consolidation

In other studies of word learning, the emerging picture is that of a dissociation in the time course of word learning, with some lexical behaviors emerging only after a period of offline consolidation and some behaviors being available immediately. Behaviors that rely upon rapid access to lexical representations seem to benefit from a period of offline consolidation, such as lexical competition (Bowers, Davis, & Hanley, 2005; Dumay & Gaskell, 2007), lexical facilitation in response in categorizing ambiguous phonemes (Lindsay, Sedin, & Gaskell, 2012), faster detection of assimilated segments in connected speech (Snoeren, Gaskell, & Di Betta, 2009), and semantic interference in naming (Clay, Bowers, Davis, & Hanley, 2007), In contrast, some behaviors are available immediately, such as recognition (as demonstrated here), lexical bias in categorization of ambiguous phonemes (Lindsay et al., 2012), and compensation for coarticulation (Snoeren et al., 2009). In the complementary systems model of word learning (Davis & Gaskell, 2009; Lindsay & Gaskell, 2010), we argue that immediately available lexical behaviors are mediated predominantly by hippocampal representations, and faster access depends more on neocortically stored memories. Although we test only after a day in the present work, in other work we have shown that lexical competition effects with novel words are still present after several months (Tamminen & Gaskell, 2008).

Interestingly, the regimes that we employed to engender lexical competition effects in the phonological neighbors of the novel words did not appear to influence the availability of the novel words themselves. This dissociation was perhaps most striking in the results of the familiarity decision task (recognition memory). As shown in Figure 2, changes in RTs and error rates during the course of the first day were negligible in this task. In contrast, all experiments showed a substantial facilitation in both RTs and error rates between Session 4 at the end of Day 1 and Session 5 at the same time the following day. This pattern of results for familiarity decision indicates that it is not simply repeating the task that led to improvements, nor was the interleaving of neighboring existing words relevant. Instead, extended time and/or sleep were crucial to enhanced speed and accuracy of recognition ability. Although the experiments reported here were not designed to explicitly test the role of sleep, the results are nonetheless consistent with sleep-associated memory consolidation leading to improvements in memory for the novel items. Tamminen et al. (2010) provided clearer evidence that improvements in novel item recognition memory (in their case just on RTs) were found after nocturnal sleep but not after an equivalent time awake during the day. The overnight change in this measure was also associated with slow-wave sleep duration, unlike the lexical competition measure, which was associated with degree of sleep spindle activity.

Conclusions

In summary, although results from other studies show that time and sleep are important for memory consolidation and the learning of new words, the present results show reliably for the first time that lexical integration as measured via engagement in lexical competition can occur within a single day. These experiments further demonstrate the benefits of spaced learning and testing for memory enhancement and shed valuable light on the processes involved in lexical integration. Previous research on word learning has demonstrated that sleep may be important for integration of new knowledge with old knowledge. The current study shows that interleaved exposure to novel words and their existing neighbors while awake may also be important in facilitating the full integration of new vocabulary items in the mental lexicon.

References

Alberini, C. M. (2005). Mechanisms of memory stabilization: Are consolidation and reconsolidation similar or distinct processes? *Trends in Neurosciences, 28,* 51–56. doi:10.1016/j.tins.2004.11.001

Ans, B., Rousset, S., French, R. M., & Musca, S. (2004). Self-refreshing memory in artificial neural networks: Learning temporal sequences without catastrophic forgetting. *Connection Science, 16,* 71–99. doi:10.1080/09540090412331271199

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61,* 213–238.

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition, 97*(3), B45–B54. doi:10.1016/j.cognition.2005.02.002

Carpenter, G. A., & Grossberg, S. (1991). *Pattern recognition by self-organizing neural networks.* Cambridge, MA: MIT Press.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20,* 633–642. doi:10.3758/BF03202713

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132,* 354–380.

Clay, F., Bowers, J., Davis, C., & Hanley, D. (2007). Teaching adults new words: The role of practice and consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 970–976. doi:10.1037/0278-7393.33.5.970

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. doi:10.1162/jocn.2009.21059

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society: Series B: Biological Sciences, 364,* 3773–3800. doi:10.1098/rstb.2009.0111

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79,* 162–170. doi:10.1037/0022-0663.79.2.162

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18,* 35–39. doi:10.1111/j.1467-9280.2007.01845.x

Dumay, N., & Gaskell, M. G. (2012). Overnight lexical consolidation revealed by speech segmentation. *Cognition, 123,* 119–132. doi:10.1016/j.cognition.2011.12.009

Dumay, N., Gaskell, M. G., & Feng, X. (2004). A day in the life of a spoken word. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 339–344). Mahwah, NJ: Erlbaum.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170–177. doi:10.1037/1082-989X.1.2.170

Ellenbogen, J. M., Hu, P. T., Payne, J. D., Titone, D., & Walker, M. P. (2007). Human relational memory requires time and sleep. *Proceedings of the National Academy of Sciences, USA, 104,* 7723–7728. doi:10.1073/pnas.0700094104

Finn, B., & Roediger, H. L. (2011). Enhancing retention through reconsolidation. *Psychological Science, 22,* 781–786. doi:10.1177/0956797611407932

Fischer, S., Hallschmid, M., Elsner, A. L., & Born, J. (2002). Sleep forms memory for finger skills. *Proceedings of the National Academy of Sciences, USA, 99,* 11987–11991. doi:10.1073/pnas.182178199

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35,* 116–124. doi:10.3758/BF03195503

Gais, S., & Born, J. (2004). Declarative memory consolidation: Mechanisms acting during human sleep. *Learning & Memory, 11,* 679–685. doi:10.1101/lm.80504

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89,* 105–132. doi:10.1016/S0010-0277(03)00070-2

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12,* 613–656. doi:10.1080/016909697386646

Henderson, L. M., Weighall, A., & Gaskell, M. G. (2012). *Science word learning in children: The impact of semantics on consolidation.* Manuscript in preparation.

Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning & Memory, 14,* 47–53. doi:10.1101/lm.365707

Karpicke, J. D., & Blunt, J. R. (2011, January 20). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331,* 772–775. doi:10.1126/science.1199327

Karpicke, J. D., & Roediger, H. L. (2008, February 15). The critical importance of retrieval for learning. *Science, 319,* 966–968. doi:10.1126/science.1152408

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. *Practical Aspects of Memory, 1,* 625–632.

Lindsay, S., & Gaskell, M. G. (2009). Spaced learning and the lexical integration of novel words. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2517–2522). Austin, TX: Cognitive Science Society.

Lindsay, S., & Gaskell, M. G. (2010). A complementary systems account of word learning in L1 and L2. *Language Learning, 60,* 45–63. doi:10.1111/j.1467-9922.2010.00600.x

Lindsay, S., Sedin, L., & Gaskell, M. G. (2012). Acquiring novel words and their past tenses: Evidence from lexical effects on phonetic categorisation. *Journal of Memory and Language, 66,* 210–225. doi:10.1016/j.jml.2011.07.005

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476–490. doi:10.3758/BF03210951

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide.* New York, NY: Cambridge University Press.

Marslen-Wilson, W. (1993). Issues in process and representation in lexical access. In G. Altmann & R. Shillcock (Eds.), *Cognitive models of language processes: Second Sperlonga meeting.* Hove, England: Erlbaum.

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. doi:10.1037/0033-295X.102.3.419

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105–125. doi:10.1037/1082-989X.7.1.105

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000, August 17). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature, 406,* 722–726. doi:10.1038/35021052

Payne, J. D., Schacter, D. L., Propper, R. E., Huang, L. W., Wamsley, E. J., Tucker, M. A., . . . Strickgold, R. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory, 92,* 327–334. doi:10.1016/j.nlm.2009.03.007

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 785–794. doi:10.1037/0278-7393.21.3.785

Robins, A. (1996). Consolidation in neural networks and in the sleeping brain. *Connection Science, 8,* 259–275. doi:10.1080/095400996116910

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20–27. doi:10.1016/j.tics.2010.09.003

Rudoy, J. D., Voss, J. L., Westerberg, C. E., & Paller, K. A. (2009, November 20). Strengthening individual memories by reactivating them during sleep. *Science, 326,* 1079. doi:10.1126/science.1179013

Schwartz, R. G., & Terrell, B. Y. (1983). The role of input frequency in lexical acquisition. *Journal of Child Language, 10,* 57–64. doi:10.1017/S0305000900005134

Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011). Human memory reconsolidation can be explained using the temporal context model. *Psychonomic Bulletin & Review, 18,* 455–468. doi:10.3758/s13423-011-0086-9

Shea, C. H., Lai, Q., Black, C., & Park, J. H. (2000). Spacing practice sessions across days benefits the learning of motor skills. *Human Movement Science, 19,* 737–760. doi:10.1016/S0167-9457(00)00021-X

Skaggs, W. E., & McNaughton, B. L. (1996, March 29). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science, 271,* 1870–1873. doi:10.1126/science.271.5257.1870

Snoeren, N. D., Gaskell, M. G., & Di Betta, A. M. (2009). The perception of assimilation in newly learned novel words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 542–549. doi:10.1037/a0014509

Tamminen, J. (2010). *Learning new words: Effects of meaning, memory consolidation, and sleep* (Unpublished doctoral dissertation). University of York, York, England.

Tamminen, J., & Gaskell, M. G. (2008). Newly learned spoken words show long-term lexical competition effects. *Quarterly Journal of Experimental Psychology, 61,* 361–371. doi:10.1080/17470210701634545

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with integration of new memories and existing knowledge. *Journal of Neuroscience, 30,* 14356–14360. doi:10.1523/JNEUROSCI.3028-10.2010

Walker, M. P., Brakefield, T., Hobson, J. A., & Stickgold, R. (2003, October 9). Dissociable stages of human memory consolidation and reconsolidation. *Nature, 425,* 616–620. doi:10.1038/nature01930

Walker, M. P., & Stickgold, R. (2006). Sleep, memory, and plasticity. *Annual Review of Psychology, 57,* 139–166. doi:10.1146/annurev.psych.56.091103.070307

Walker, M. P., & Stickgold, R. (2010). Overnight alchemy: Sleep-dependent memory evolution. *Nature Reviews Neuroscience, 11,* 218. doi:10.1038/nrn2762–c1

Appendix

List of Novel Words and Associated Base Words Used in Experiments

List A		List B	
Base word	Novel word	Base word	Novel word
alcohol	*alcohin*	anecdote	*anecdel*
apricot	*aprickel*	assassin	*assassool*
badminton	*badmintel*	bayonet	*bayoniss*
biscuit	*biscal*	blossom	*blossail*
bramble	*brambooce*	caravan	*caravoth*
cardigan	*cardigite*	clarinet	*clarinern*
consensus	*consensom*	crocodile	*crocodiss*
daffodil	*daffadat*	decibel	*decibit*
dolphin	*dolpheg*	dungeon	*dungeill*
grimace	*grimin*	hormone	*hormike*
hurricane	*hurricarb*	lantern	*lantobe*
mandarin	*mandarook*	methanol	*mathanack*
mistress	*mistrool*	molecule	*molekyen*
ornament	*ornameast*	parachute	*parasheff*
parsnip	*parsneg*	pelican	*pelikiyve*
profile	*profon*	pulpit	*pulpen*
pyramid	*pyramon*	skeleton	*skeletobe*
slogan	*slowgiss*	spasm	*spaset*
specimen	*specimal*	squirrel	*squirrome*
tulip	*tulode*	utensil	*utensont*

Received June 20, 2011
Revision received April 16, 2012
Accepted May 1, 2012 ■