Original Articles

Contextual priming of word meanings is stabilized over sleep

M. Gareth Gaskell[a,*], Scott A. Cairney[a], Jennifer M. Rodd[b]

[a] University of York, UK
[b] University College London, UK

A B S T R A C T

Evidence is growing for the involvement of consolidation processes in the learning and retention of language, largely based on instances of new linguistic components (e.g., new words). Here, we assessed whether consolidation effects extend to the semantic processing of highly familiar words. The experiments were based on the word-meaning priming paradigm in which a homophone is encountered in a context that biases interpretation towards the subordinate meaning. The homophone is subsequently used in a word-association test to determine whether the priming encounter facilitates the retrieval of the primed meaning. In Experiment 1 (N = 74), we tested the resilience of priming over periods of 2 and 12 h that were spent awake or asleep, and found that sleep periods were associated with stronger subsequent priming effects. In Experiment 2 (N = 55) we tested whether the sleep benefit could be explained in terms of a lack of retroactive interference by testing participants 24 h after priming. Participants who had the priming encounter in the evening showed stronger priming effects after 24 h than participants primed in the morning, suggesting that sleep makes priming resistant to interference during the following day awake. The results suggest that consolidation effects can be found even for highly familiar linguistic materials. We interpret these findings in terms of a contextual binding account in which all language perception provides a learning opportunity, with sleep and consolidation contributing to the updating of our expectations, ready for the next day.

1. Introduction

Over the last 20 years, a substantial body of psycholinguistic research has uncovered remarkable plasticity in the adult system. Whereas previously language development might have been characterised as a steady progression towards a fairly stable state, it is now clear that such a stable state is never achieved. Instead, we retain substantial plasticity as adults, allowing us to adapt our perception of phonemes when exposed to unfamiliar accents (Norris, McQueen, & Cutler, 2003), tailor our production system to reflect the statistical structure of our environment (Dell, Reed, Adams, & Meyer, 2000) and acquire and retain new forms (Gaskell & Dumay, 2003), meanings (Fang & Perfetti, 2017; Rodd et al., 2012) and syntactic constructions (Kaschak & Glenberg, 2004; Ryskin, Qi, Duff, & Brown-Schmidt, 2017). Along with these observations of plasticity, there has also been an enhanced recognition of the applicability of detailed theories of memory function to the domain of psycholinguistics (Davis & Gaskell, 2009; Gagnepain, Henson, & Davis, 2012; Szmalec, Page, & Duyck, 2012).

One key example of this increased synergy between memory and language has involved our understanding of the importance of consolidation processes in language learning. Studies of infants (Friedrich, Wilhelm, Mölle, Born, & Friederici, 2017; Gomez, Bootzin, & Nadel, 2006; Horváth, Myers, Foster, & Plunkett, 2015), children (Friedrich et al., 2017; Henderson, Weighall, Brown, & Gaskell, 2012; James, Gaskell, Weighall, & Henderson, 2017; Sandoval, Leclerc, & Gómez, 2017; Williams & Horst, 2014) and adults (Bakker-Marshall et al., 2018; Bakker, Takashima, van Hell, Janzen, & McQueen, 2014; Dumay & Gaskell, 2007; Kurdziel, Mantua, & Spencer, 2017) have shown that retention and integration of new linguistic knowledge can benefit from a consolidation period, and sometimes specifically from a sleep period (Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010). For example, interference from learning a new word (e.g., "cathedruke") on the recognition of its existing neighbour (e.g., *cathedral*) tends not to be observed immediately (although cf. McMurray, Kapnoula, & Gaskell, 2016), but instead emerges after a period of sleep (Dumay & Gaskell, 2007) and is associated with the prevalence of spindle activity (brief ~12–15 Hz bursts of activity in non-REM sleep) during the intervening night (Tamminen et al., 2010). These observations can be explained by systems consolidation models (e.g., Rasch & Born, 2013) applied to language learning (Davis & Gaskell, 2009) in which sleep provides an opportunity for new hippocampally mediated memories to be replayed (Ji & Wilson, 2007; Nadel, Hupbach, Gomez, & Newman-Smith, 2012)

and strengthened in cortical networks.

Although it is fairly clear that consolidation is a component process in the retention of language knowledge, there is substantial variability in the extent to which consolidation effects are found (McMurray et al., 2016), which likely reflects the nature of the material to be learned. The studies that have revealed consolidation effects in language learning have tended to focus on examples of new material (e.g., novel words or grammars), and it is possible that stimulus novelty is the main factor that determines the level of reliance on consolidation. This would fit with a complementary systems account of language learning (McClelland, 2013) in which the hippocampus steps in to facilitate learning in cases where adjustment of cortical weights would interfere with existing knowledge (Mirković & Gaskell, 2016).

An exception to this rule is the study of Gaskell et al. (2014), who examined the role of sleep in the acquisition of phonotactic constraints in speech production. The research exploited the work of Dell and colleagues (Dell et al., 2000; Warker & Dell, 2006; Warker, Dell, Whalen, & Gereg, 2008; Warker, Xu, Dell, & Fisher, 2009) who had shown that speakers can acquire new phonotactic constraints (e.g., absence of a /g/ at syllable offset) over brief periods of time, as evidenced by the structure of their speech errors. Gaskell et al. (2014) extended the work of Warker (2013) and showed that when the constraint is a more complex "second-order constraint" (e.g., absence of a /g/ after /ae/) the integration of this constraint into speech errors is facilitated by a period of sleep (specifically, slow-wave sleep). A recent study demonstrated that consolidation also benefits the acquisition of second-order constraints when the material to be learned is non-linguistic (Anderson & Dell, 2018). The phonotactic consolidation effect can in some ways be thought of as learning of new material (i.e. the "gaps" in the repertoire of allowable sequences are new), but in other ways the change can be thought of as a revision of existing knowledge about the co-occurrence probabilities of various phonemes. Therefore, it remains to be seen whether sleep and consolidation are important for the retention of new evidence that acts to revise a well-established body of existing linguistic knowledge.

In the current study, we examine the potential for sleep and/or consolidation to influence the process of selection between the various familiar meanings of lexically ambiguous words. This domain exemplifies the building up of a body of knowledge over the course of a lifetime relating to the likelihood of different meanings, and yet has been shown to be susceptible to priming effects in the short term, suggestive of plasticity (Rodd, Cutrin, Kirsch, Millar, & Davis, 2013). For example, the word "pen" has multiple meanings, and in the absence of biasing contextual cues participants tend to retrieve the most frequent meaning of the word (in this case, relating to the writing implement meaning; Twilley, Dixon, Taylor, & Clark, 1994). This frequency bias is likely to reflect some kind of learning mechanism that amasses frequency counts from experience of the usage of the ambiguous word over a long period of time.

Rodd et al. (2013) examined whether a recent experience with a particular meaning of an ambiguous word could alter the likelihood of retrieval of the different meanings. According to a kind of "knowledge crystallization" account, such an effect of recent experience would be unlikely, because frequency biases accumulated over many experiences across decades should not alter by a discernible amount on the basis of a single new encounter. However, an explanation that favours recent experience or maintains strong plasticity would suggest that meaning frequency biases should be more flexible. Rodd et al. tested these different accounts using a *word-meaning priming* paradigm.[1] Participants

first encountered a set of ambiguous words embedded in spoken sentences that biased the subordinate (i.e., less favoured) meaning of the word (e.g., "A pen was used by the farmer to enclose the stock before he moved them to the market"). Participants were then tested on their comprehension of the primed ambiguous words—compared with a baseline unprimed condition—by presenting the ambiguous words in isolation as cues and asking participants to generate an associated word. Rodd et al. found that the proportion of associate responses consistent with the primed meaning rose from about 0.17 in the unprimed condition to 0.24 in the primed condition. A second experiment showed that the priming effect could not be explained simply in terms of standard semantic priming, which was more short-lived.

This word-meaning priming effect is relatively abstract, in that it applies regardless of whether the same or a different speaker is used for the priming sentence and the isolated cue word (Rodd et al., 2013), and transfers across spoken and written modalities (Gilbert, Davis, Gaskell, & Rodd, 2018). Although the delay used between exposure and testing was relatively long compared with many semantic- or form-priming studies (about 20 mins on average), this latency does not really provide much information about whether long-term lexical representations are being altered. Rodd et al. (2016) went further in mapping out the time-course of word-meaning priming effects. They compared (in Experiment 2) exposure-test latencies of 1, 20 and 40 min with an unprimed baseline, finding that all three latencies showed some priming of associate responses, but with the 1-min condition stronger than the two longer latency conditions. This was taken as evidence of a relatively fast-fading component of the priming. Three further experiments examined longer latencies in a more naturalistic design, with good evidence that priming effects showed gradual decay across a day, and that beyond a day these effects weakened and were no longer significant (Experiment 1, Experiment 4). Intriguingly, though, participants with specific repeated experience of particular meanings of words (rowers with esoteric meanings of words like "feather") showed an influence of that experience on the likelihood of retrieval of the esoteric meaning several hours after that experience (e.g., rowing early morning and testing in the afternoon).

Rodd et al. (2016) outlined a working model that might explain this complex pattern of priming effects. They argued that distributed connectionist models of ambiguous word representation and processing (Joordens & Besner, 1994; Kawamoto, Farrar, & Kello, 1994; Rodd, Gaskell, & Marslen-Wilson, 2004) provide a natural account of how meaning biases could be updated as a consequence of a recent experience via adjustment of the long-term weights between form and meaning units. This would make the primed meaning a little easier to access and the unprimed meaning(s) a little harder to access. This model, then, can quite easily explain the enhanced likelihood of accessing a primed meaning of an ambiguous word at a later timepoint. But is it also possible to explain the apparent decay in this effect that is seen across the course of the remainder of the day? Rodd et al. suggested that this could be a consequence of further learning and updating of weights in response to intervening unrelated language exposure, given the highly interconnected nature of representations in a distributed model of meaning. However, they also pointed out that the specific decay function observed in their studies, with strong decay initially and weaker decay later on, might be difficult to accommodate in such a model, speculating that multiple mechanisms might reasonably be involved.

Borrowing again from the memory literature, the apparent decay observed by Rodd et al. (2016) might indeed be a product of a second system involved in the priming of ambiguous word meanings. Several models of memory and forgetting have argued that the hippocampus incorporates a prodigious ability to encode new associations through pattern separation of sparse representations (Sadeh, Ozubko, Winocur, & Moscovitch, 2014; Yassa & Stark, 2011). This makes these representations resistant to interference, given that they have little overlap with other representations, but at the same time they are

---

[1] We use the term word-meaning priming for consistency with the prior literature on this paradigm. The term "priming" is used in its simplest sense, as a description of the facilitation of access to a particular meaning as a consequence of a prior stimulus presentation, rather than as a description of a particular mechanism.

susceptible to decay. As Hardt, Nader, and Nadel (2013; p. 112) put it "decay-driven forgetting is a direct consequence of a memory system that engages in promiscuous encoding". Therefore, a second potential explanation of the time-course of word-meaning priming might make use of the hippocampus to facilitate the encoding or binding of new associations between the words in a sentence during the exposure session. This new representation could form a second source of information alongside more permanent lexical semantic knowledge when participants are asked at a later point to generate associates of a given word. Furthermore, hippocampal trace decay would provide an explanation of why word-meaning priming effects tend to weaken as the delay between exposure and test increases within the course of a day.

As discussed above, there is a growing body of evidence that the hippocampus has an important role to play in the encoding of novel words. However, no new words are learned in word-meaning priming sentences; they are simply semantically coherent sentences containing familiar words. So the notion that the hippocampus is involved in learning simply during the comprehension of these sentences may seem unlikely. Nonetheless this suggestion is not new. Evidence from amnesic participants suggests that the hippocampus is involved in a range of online linguistic tasks that contribute to normal everyday conversation beyond simply word learning (Duff & Brown-Schmidt, 2012, 2017). These include tasks as diverse as the maintenance of common ground (Duff, Gupta, Hengst, Tranel, & Cohen, 2011), the use of co-speech hand gesture (Hilverman, Cook, & Duff, 2016), the flexible use of language (Duff, Hengst, Tranel, & Cohen, 2006) and, potentially, the updating of verb biases in syntactic ambiguity resolution (Ryskin, Qi, Covington, Duff, & Brown-Schmidt, 2018). Of particular relevance to the current study is the work of Klooster and Duff (2015). They questioned the received wisdom that remote semantic memory does not require a functioning hippocampus by testing a group of patients with hippocampal amnesia on the richness and depth of semantic knowledge for a range of different word types. They found that the amnesic participants performed worse than matched controls in both productive and expressive measures, suggesting that the hippocampus is involved in the maintenance of semantic representations well beyond initial acquisition. With specific reference to lexical ambiguity, they tested their patients on a set of ambiguous words and asked them to list as many senses as they could. Compared with both healthy comparison participants and patients with brain damage that spared the medial temporal lobe, the amnesic patients retrieved significantly fewer senses. Similar results were observed in a semantic features task and a word association task. These results suggest that a functioning hippocampus is needed for the maintenance and updating of rich semantic representations of lexical knowledge, including lexically ambiguous words. Further, they strengthen the viability of our alternative model of word-meaning priming that incorporates hippocampal involvement in the perception and retention of the sentences used as context for exposure to ambiguous words.

Existing evidence is therefore largely consistent with two quite different explanations of how word-meaning priming operates. The preferred explanation up to now is one in which each new experience with an ambiguous word immediately alters the long-term cortical lexical connections between form and meaning in favour of the contextually appropriate meaning (e.g., Gilbert et al., 2018). The alternative explanation is that new experiences with ambiguous words recruit hippocampal resources to temporarily bind the word with its context to provide an additional source of knowledge alongside permanent lexical knowledge. These two explanations can be differentiated in terms of their predictions relating to systems consolidation of word-meaning priming. For the *immediate alteration* account, there is little reason to suspect that systems consolidation might be relevant. The learning has already taken place in the cortex, and so the knowledge has already been made permanent. On the other hand, the *contextual binding* account suggests that systems consolidation via hippocampal replay during sleep will gradually integrate this new piece of

information to ensure that long term lexical semantic representations remain rich and up-to-date (Duff & Brown-Schmidt, 2017).

We can test these predictions by looking further into the time-course of word-meaning priming, this time comparing periods spent awake and asleep. If sleep facilitates hippocampal replay of memories then we should see a benefit of sleep over wake in terms of the strength of word-meaning priming after the sleep/wake period. More specifically, slow-wave sleep and/or sleep spindles should be beneficial for strengthening or maintenance of word-meaning priming, given that slow-wake sleep is understood to be the main stage in which hippocampal replay occurs, marked by spindle activity (Rasch & Born, 2013). On the other hand, if word-meaning priming occurs via direct rewiring of cortical lexical connections, then sleep should have no particular active role to play in preserving word-meaning priming effects. That said, it still remains possible that sleep might have some passive role to play in maintaining word-meaning priming by providing passive protection against interference or decay (Ellenbogen, Hulbert, Stickgold, Dinges, & Thompson-Schill, 2006). We return to this point in the light of the evidence provided by Experiment 1.
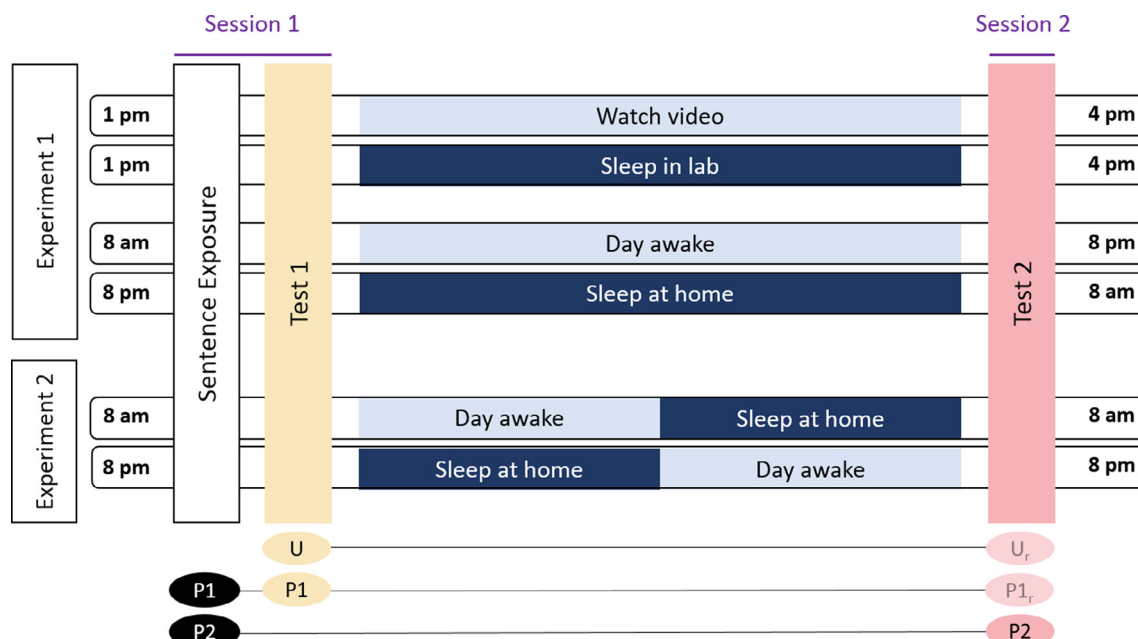
2. Experiment 1

Testing in Experiment 1 involved two sessions separated by a delay that contained either sleep or wake (see Fig. 1). In Session 1 we exposed participants to homophones (e.g., *pen*) in spoken sentence contexts. As in previous studies (e.g., Rodd et al., 2013, 2016) the sentences always biased comprehension of the ambiguous word towards the weaker, subordinate meaning (e.g., "A pen was used by the farmer to enclose the stock before he moved them to the market."). The subordinate direction of bias was chosen for comparability with previous studies and because we thought that this was the circumstance that was most likely to reveal strong effects. There were three counterbalanced sets of items, and participants heard sentences for two of the three sets, with the third set forming the control unprimed items (see Fig. 1). After a short filler task, we then tested participants' interpretations of a subset of the homophones using a word-association test (the unprimed set, plus one of the primed sets). The crucial question was whether the proportion of associate responses related to the subordinate meaning (e.g., *pen-pig*) was higher than for a set of matched homophones that had not been primed during exposure. A second word-association test was conducted using the second of the primed sets of ambiguous words in Session 2 after a delay period that included sleep or wake. This allowed us to determine whether the priming varied as a consequence content of the delay. For a subsidiary analysis, the Session 2 word-association test also included the other two sets of previously tested words in order to determine whether repeated testing influenced the pattern of performance after a delay. We implemented two protocols for the delay. In one, the delay was ~12 h outside the lab and began either in the morning (wake) or in the evening (sleep). This kind of design is common in sleep and memory studies because it can assess the potential benefit of a full night of nocturnal sleep. This design is also susceptible to circadian confounds in term of the mental state of the participants at training and test (Doyon et al., 2009). The second protocol employed a shorter delay period taken in the lab and matched sleep and wake groups in terms of time of day of exposure and testing. This is also a commonly used design because it eliminates any circadian confound but with the potential cost of a less substantial difference between groups in terms of the time spent in sleep. For the second protocol the sleep group had a polysomnographically recorded afternoon nap whereas the wake group watched a silent film.

2.1. Methods

2.1.1. Participants

Participants were undergraduates at the University of York with English as their first language. In total, 74 participants were tested in

Fig. 1. Illustrative timelines for procedures in Experiments 1 and 2. A participant would be assigned to one of the four timelines in Experiment 1, or one of the two timelines in Experiment 2. Timings are approximate. The lower section illustrates the three stimulus sets that are involved in each exposure/test session (U = unprimed; P1 = primed, tested in Session 1; P2 = primed, tested in Session 2; r subscripts represent repeated tests).

Experiment 1, evenly split between sleep and wake groups. To provide adequate power for the individual differences analyses that were conducted on the participants who had their afternoon nap polysomnographically recorded, 44 participants were allocated to the 2-h protocol, with the remaining 30 allocated to the 12-h protocol. The four groups were well-matched for mean ± SD age in years (2-h sleep 20.8 ± 2.3; 2-h wake 20.4 ± 0.9; 12-h sleep 20.5 ± 0.6; 12-h wake 20.6 ± 0.9) and sex (each group contained 4 males). Participants gave informed consent for the experiment, which was approved by the Research Ethics Committee of the Department of Psychology, University of York. As assessed by self-report, participants were non-smokers, not taking any psycho-active medication, had no prior history of drug or alcohol abuse, no neurological, psychiatric or sleep disorder, and had a sleep-wake pattern where they typically rose by at least 9 am each morning, after a duration of at least 6 h of sleep per night. For the 24 h before the experiment and during the experiment, participants were asked to abstain from alcohol, caffeine and smoking, and to awaken by 8 am on the day of the experiment.

2.1.2. Materials

The key materials (see Appendix A) were 87 homophonic spoken words and sentence contexts previously used in word-meaning priming studies (e.g., spade; Rodd et al., 2016, Experiment 2; Betts, Gilbert, Cai, Okedara, & Rodd, 2018; e.g., spade; Rodd et al., 2016, Experiment 2). Prior dominance ratings had established for each word a dominant meaning (e.g., *digging implement*; mean dominance = 0.70; SD = 0.14) and a subordinate meaning (e.g., *suit in card game*; mean dominance = 0.23; SD = 0.13). For each word, a sentence context provided a prior biasing context that clearly indicated the subordinate meaning (e.g., "The gambler knew that his opponent wanted a spade"). Each sentence was also matched with a semantically unrelated probe word (e.g., *hair*) that was used in a relatedness matching task during exposure to ensure that participants attended to meaning. The test sentences were divided into three matched sets for the purposes of counterbalancing.

Thirty spoken filler sentences were included in the exposure phase in Session 1 alongside the test materials. These were similar in construction to the test sentences, but had probe words that were

semantically related to the sentences (e.g., "Soup had spilled all over the counter", broth). The sentences and the isolated ambiguous words had been recorded by a female native British English speaker (Rodd et al., 2016).

The Stanford Sleepiness Scale (SSS; Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973) was used to identify rated sleepiness at key points in the experiment, and a simple RT test provided a second source of evidence relevant to sleepiness. Participants also filled out general sleep quality and language background questionnaires.

Two videos were used in the experiment. Between exposure and Test 1, an 8-minute episode of "Shaun the Sheep" (http://shaunthesheep.com/) occupied the participants. For the 2-h delay period in the wake condition participants watched an animated film (*L'Illusioniste*; Chomet, 2010). Both these videos were selected because they had minimal linguistic content.

2.1.3. Polysomnography

Polysomnographic monitoring was conducted on participants in the two-hour sleep condition using an Embla N7000 system with Remlogic 3.0 software in our sleep lab. Electrodes were attached using EC2 electrode gel after the scalp was prepared using NuPrep exfoliator. Scalp electrodes were attached at O1, O2, C3, C4, F3 and F4 according to the international 10–20 system, each referenced to the contralateral mastoid (A1, A2). Left and right electrooculograms (EOG) electrodes were attached, as were electromyography electrodes at the mentalis and submentalis bilaterally, and a ground electrode was attached to the forehead. EEG electrodes had a connection impedance of < 5 kΩ and all signals were digitally sampled at 200 Hz. Sleep data were bandpass filtered (0.3–40 Hz) and scored in 30 s epochs using RemLogic 3.0 in accordance with the standardised sleep scoring criteria of the American Academy of Sleep Medicine (Iber, 2007). All scoring was carried out by a single trained researcher (the second author) based on the frontal channels, allowing determination of the time spent in each sleep stage (N1, N2, N3 and REM).

For epochs scored as either N2 or N3 (i.e., those that could contain spindles), artefacts were eliminated by visual inspection before a 12–15 Hz linear finite impulse response filter was applied to frontal and central channels. An automated detection algorithm (Ferrarelli et al.,

2007) counted discrete spindle events as amplitude modulations that exceeded a threshold of 8 × mean filtered amplitude for that channel.

As stated in the introduction, if word meaning priming is supported by hippocampal replay during sleep then one would predict that duration of N3 (slow-wave sleep) should be a predictor of change in priming over sleep, as should the density of spindle activity (Mednick et al., 2013; Rasch & Born, 2013; Tamminen et al., 2010). Nonetheless, some have argued that improvements in priming may be driven by REM sleep (Cai, Mednick, Harrison, Kanady, & Mednick, 2009; Plihal & Born, 1999; Stickgold, Scott, Rittenhouse, & Hobson, 1999), so we also tested the subsidiary prediction that there might be an association between duration of REM sleep and changes in word-meaning priming over sleep.

In the 12-h sleep condition, participants were asked to wear a commercial sleep recording device (Zeo; Zeo Inc.). However, equipment failure, noncompliance and problems with the headband slipping off led to absent or incomplete sleep data for the majority of participants in this condition and the remaining data were not analysed further.

2.1.4. Design

The main dependent variable was the proportion of word associates generated in response to the test material that matched the sententially primed subordinate meaning of the word as opposed to the dominant meaning. Two between-participants independent variables manipulated the presence or absence of sleep in the interval between testing points (sleep vs. wake) as well as the interval duration (2 h vs 12 h). Within participants, the main analysis manipulated three levels of the variable priming: control performance without priming during exposure was assessed in Session 1 (unprimed), whereas half the primed items were tested in Session 1 (primed-early) and the remainder in Session 2 (primed-delayed). The three matched subsets of the 87 test items were rotated across these three conditions in a counterbalanced fashion so that all items were used in all conditions across the whole experiment. The use of three separate sets of items ensured that any comparison of priming between test points did not involve the repeated presentation of homophones in the word-association test. However, we thought it prudent to include all sets of items in the second word-association test as a secondary measure of any effect of priming. Therefore we were also able to look at priming specifically in the repeated items in Session 2 as a separate analysis.

2.1.5. Procedure

Participants signed up for either the 2-h or the 12-h version of the experiment, but were randomly allocated to the wake or sleep group within these two protocols. Both sessions took place in individual study rooms or bespoke testing booths at the University of York. For the 12-r protocol one session took place between 7 and 11 am and the other between 7 and 11 pm. The wake group had the morning session first, followed that evening by the second. The sleep group had the reverse order, with the sessions separated by a night's sleep. For the 2-h protocol, Session 1 began at 1 pm, with Session 2 beginning roughly 2 h after the end of Session 1.

2.1.5.1. Session 1. All participants were first asked to complete a language background questionnaire to assess if they were monolingual British English speakers and to determine their level of experience with other languages. Participants then completed the SSS to assess their alertness at the start of the experiment. Other experimental tasks were completed on PC or laptop computers with DMDX software (Forster & Forster, 2003) using headphones for the spoken stimuli. Responses were recorded via a computer keyboard. Participants completed the simple RT test, followed by the exposure task (7 min). They then watched the 8-min filler video before taking the word-association test for the first time.

2.1.5.2. Delay period. In the 12-h conditions participants left the lab and went about their normal activities before returning to the lab 12 h later. In the 2-h condition, the sleep participants had a 90-min nap opportunity with polysomnographic monitoring followed by a 10-min rest period to minimise the influence of sleep inertia. In the wake condition, participants watched the film L'Illusioniste for 90 min and then had a 10-min rest.

2.1.5.3. Session 2. Participants completed the SSS and the simple RT test followed by the second, longer word-association test that included items from all three counterbalancing sets.

2.1.5.4. The tasks.

1. *Simple RT.* Participants were presented with 18 trials of a simple decision task to assess alertness (Reid, 2013). On each trial they saw a fixation cross for 500 ms before seeing two digits ordered as "1 0" or "0 1"). Their task was to press one of two keys depending on whether the 1 was on the left or the right of the pair. Nine trials of each type were used in a random order. Participants were asked to respond as quickly and accurately as possible.
2. *Semantic relatedness exposure task.* Participants completed a practice block of five trials before completing all 88 experimental trials (58 test trials, 30 fillers), taking a short break at the half way point. Each trial began with a fixation cross for 1000 ms, followed by the presentation of a sentence over headphones. Following the completion of the sentence there was a 1000 ms delay and then the probe word was presented visually. Participants were asked to use the shift buttons on the keyboard to indicate whether they thought the probe was semantically related to the sentence (related = right, unrelated = left). They were asked to respond as quickly and accurately as possible. Three counterbalanced versions of the test were used, each containing two of the three 29-item sets, leaving the third item set as unprimed.
3. *Word-association task.* After 5 practice items, each trial began with a fixation cross for 1000 ms, and then a homophone item (e.g., "pen") was presented as a cue via headphones. A response cue followed 200 ms after the word offset and participants were asked to type in the cue word (to check whether it was perceived correctly), and then were instructed to type "the first word that comes to mind that is related to the word you heard". Participants were given the example of "tennis-Wimbledon". In Session 1, three counterbalanced versions of the experiment were used, each containing the unprimed item set plus one of the two primed item sets (58 trials in total). In Session 2, all three item sets were used (87 trials in total). In each case, the order or trials was randomised afresh.

2.2. Results

2.2.1. Sleep questionnaire and Stanford Sleepiness Scale

According to the sleep questionnaire wake participants had a mean bedtime in the last month of 23:44 (SD = 62 mins) and sleep participants had a mean bedtime of 23:56 (SD = 72 min; $t(72) = -0.78$, $p = 0.44$). The groups were also well matched on estimated typical sleep duration (mean ± SD: wake = 7.78 ± 1.01 h, sleep: 7.97 ± 1.03 h; $t(71) = -0.79$, $p = 0.43$). The vast majority of participants reported that they had not taken any prescription medicine to help them sleep in the last month (wake: 36/37; sleep: 36/37). An ANOVA on participant SSS ratings with the within-participants independent variable Session (1 vs. 2) and between participants variables Interval Duration (2 h/12 h) and Group (sleep/wake) revealed a main effect of Session ($F[1, 68] = 8.10$, $p = 0.006$). Participants rated their sleepiness to be higher on the scale in Session 2 (2.97) than Session 1 (2.49). No other effects or interactions were statistically significant, although the interactions between Session and Interval Duration ($F[1, 68] = 3.42$, $p = 0.069$) and Session and Group ($F[1, 68] = 3.01$, $p = 0.069$) approached significance. As a further precaution, we tested

whether sleepiness ratings could explain priming levels by running correlations between rated sleepiness and level of priming in each Session (see below). Out of the four tests (wake-Session 1, sleep-Session 1, wake-Session 2, sleep-Session 2), only the correlation for wake participants in Session 1 was significant at $p < 0.05$ uncorrected for multiple comparisons ($r_\tau = 0.26$, $p = 0.043$) providing a weak hint that priming effects might be stronger when participants are sleepier. Crucially, the main pattern of priming change across session could not be explained in this way (in fact it runs contrary to the levels of sleepiness in Session 2).

2.2.2. Simple RT

The RT test was a secondary measure of alertness. An ANOVA on participant mean RTs with same independent variable as above revealed a significant interaction between Session and Interval Duration ($F[1, 64] = 4.80$, $p = 0.032$). For participants in the nap study, RTs varied very little across session (mean ± SEM; Session 1: 393 ± 10 ms, Session 2: 397 ± 10 ms), whereas there was a little more variation across session for the 12 h protocol (Session 1: 395 ± 11 ms, Session 2: 380 ± 11 ms). The Session × Group interaction was also significant ($F[1, 64] = 5.65$, $p = 0.020$) with sleep participants improving their RTs slightly across the delay (400 ± 10 vs 384 ± 10 ms) and wake participants getting a little slower (388 ± 10 vs 393 ± 10 ms). These effects are indicative of small differences in alertness in the different groups and sessions, and so to determine whether they could explain the main priming effects for the word association test, we ran correlations between every participant's RT speed and level of priming in the two sessions. None of these approached significant level. In sum, both the SSS and RT analyses suggested that the pattern of priming found in the word association test could not be explained in terms of any confound of alertness or sleepiness.

2.2.3. Word association responses

Obvious spelling mistakes were corrected, and word association responses were coded by two experimenters who were blinded as to the experimental condition. The coding categories along with the number and percentage of responses that fell into the categories were as follows.

(1) Dominant meaning (7294, 68%).
(2) Subordinate (primed) meaning (2907, 27%).
(3) Other meaning (301, 3%).
(4) Unclear as to which of the meanings the response is related to 121 (1%).
(5) Homophone cue incorrectly identified (63, 1%).
(6) No response (3, 0%).
(7) Other response (42, 0%).

The analyses focused on the 95% of responses from the first two categories, with the proportion of these trials that were consistent with the primed meaning being the dependent variable (see Table 1). Mixed effects logistic analyses used the glmer software from the lme4 (Bates, Maechler, Bolker, & Walker, 2014) package of R (R Core Team, 2013). The binary between-participant independent variables Group (wake vs sleep), and Interval Duration (2 h vs 12 h) were effect coded. Within participants, the independent variable Priming had three levels (Unprimed, Primed-Session 1, Primed-Session 2), which represented the presence or absence of priming, and the session of testing. This variable was coded using orthogonal Helmert contrasts that compared: (i) the unprimed condition vs the two combined primed conditions (Priming1) and (ii) the two primed conditions with each other (Priming2). Our analysis method followed Jaeger (2011; https://hlplab.wordpress.com/2011/06/25/more-on-random-slopes/) by building the maximal model including random items and participants effect structures that was justified by the sample (Barr, Levy, Scheepers, & Tily, 2013) and then enhancing the power of the analysis (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017) by reducing the random effects structure of the

Table 1
Experiment 1 means and standard deviations for the proportional consistency of the word association test responses with the prime sentential context as a function of Interval Duration, Group and Priming.

| | Priming | | | | | | | | | |
| | Unprimed | | Primed-S1 | | Primed-S2 | | Unprimed-R | | Primed-R | |
Group	M	SD	M	SD	M	SD	M	SD	M	SD
2-h interval duration										
Wake	0.28	0.07	0.32	0.08	0.29	0.08	0.27	0.07	0.30	0.07
Sleep	0.26	0.09	0.28	0.09	0.30	0.07	0.25	0.07	0.26	0.08
12-h interval duration										
Wake	0.26	0.07	0.34	0.08	0.28	0.08	0.28	0.07	0.31	0.08
Sleep	0.26	0.07	0.31	0.11	0.34	0.09	0.25	0.09	0.29	0.10

Note. M and SD represent mean and standard deviation, respectively. S1 = Session 1, S2 = Session 2, R = repeated.

model to the simplest structure that was not substantially different in terms of its explanatory value (using anova model comparison and a threshold of chisq < 3). Likelihood of convergence was enhanced by using the "bobyqa" optimiser and the "maxfun = 2e + 05" setting. If necessary to help interpret interactions, pairwise comparisons were conducted on the mixed effects model using the R package "emmeans" with Holm correction for multiple comparisons. In these cases, uncorrected p-values are reported with those that remain significant at alpha < 0.05 marked with an asterisk.

This analysis, reported in detail in Table 2, revealed an effect of Priming1, showing that, overall, there were more responses consistent with the prime sentences for the primed items than the unprimed items. More importantly, there was an interaction between Group and Priming2. This showed that the effect of Session for the primed items varied significantly depending on Group (see Fig. 2), with the percentage of prime-sentence consistent responses dropping from 32.9% to 28.8% across a wake interval ($Z = 2.07$, $p = 0.039$) but rising from 29.1% to 31.7% across an interval including sleep ($Z = 1.00$, $p = 0.32$). Aside from an intercept effect, no other effects were significant.

A subsidiary analysis examined just the two sets of items in Session 2 that had already been tested in Session 1 (i.e., participants had already generated an associate response to these items in Session 1). This analysis followed along the same lines as the first analysis, except that

Table 2
Summary of the mixed effects analyses for Experiment 1.

Fixed effect	Estimate	Std error	z value	p(> |z|)
Main analysis				
(Intercept)	−1.150	0.135	−8.521	< 0.001***
Priming1	0.087	0.028	3.153	0.002**
Priming2	−0.031	0.043	−0.734	0.463
Group	−0.036	0.042	−0.869	0.385
Interval	0.029	0.038	0.758	0.449
Priming1:Group	0.011	0.023	0.458	0.647
Priming2:Group	0.093	0.038	2.433	0.015*
Priming1:Interval	0.028	0.023	1.213	0.225
Priming2:Interval	−0.009	0.038	−0.248	0.804
Group:Interval	0.043	0.038	1.125	0.261
Priming1:Group:Interval	0.007	0.023	0.290	0.772
Priming2:Group:Interval	0.034	0.038	0.888	0.375
Repeated Items				
(Intercept)	−1.244	0.137	−9.072	< 0.001***
Group	−0.089	0.048	−1.842	0.065
Priming	0.139	0.049	2.863	0.004**
Interval	0.042	0.050	0.852	0.394
Group:Priming	−0.003	0.039	−0.066	0.948
Group:Interval	0.007	0.044	0.155	0.876
Priming:Interval	0.011	0.040	0.266	0.790
Group:Priming:Interval	0.028	0.039	0.724	0.469
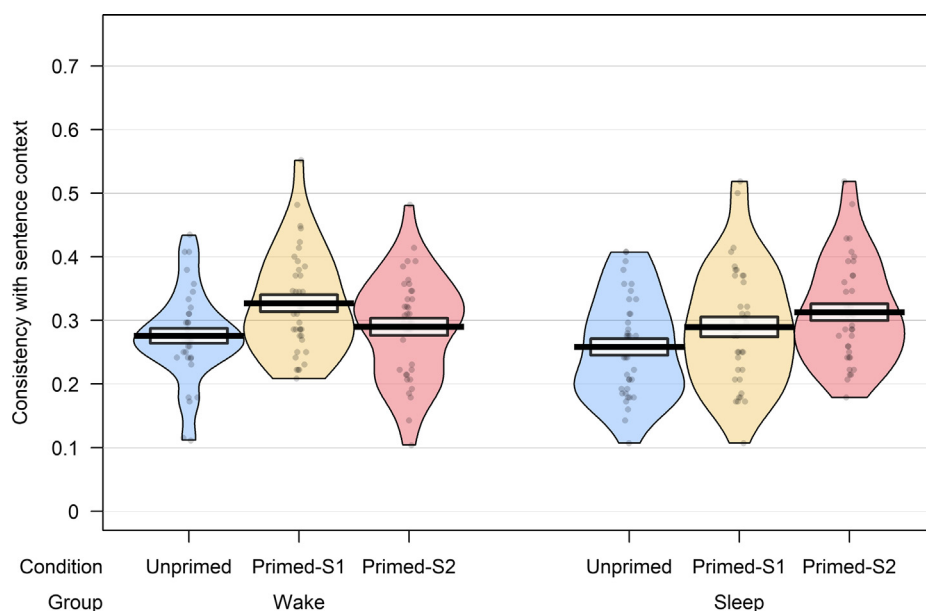
Fig. 2. Experiment 1 priming effects for wake and sleep groups based on participant means in the consistency of word association responses with the primed meaning. The horizontal solid bar represents the mean, with the band marking standard error. Points represent individual participant means with the coloured areas representing smoothed density.
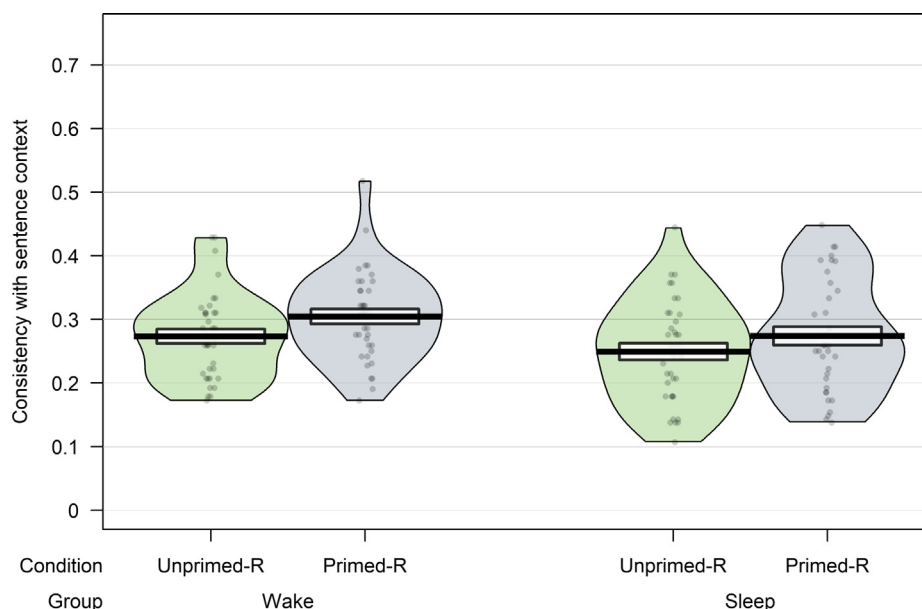
Fig. 3. Experiment 1 priming effects for repeated items in Session 2. Horizontal bars represent the mean, with bands marking standard error. Points represent individual participant means, with the coloured areas representing smoothed density.

there were only two levels of the Priming variable (unprimed vs primed) and so this variable was effect-coded. Note that if these repeated testing items showed the same pattern of responses as the unrepeated items above then one would expect stronger priming for the sleep group than the wake group. In fact, there was no evidence for such a difference (see Tables 1 and 2). Instead there was an overall effect of Priming, with similar levels of priming in the sleep and wake groups (see Fig. 3). No other effects reached significance level apart from the intercept.

2.2.4. Sleep data

Total sleep time for the participants who napped in the sleep lab was 66.9 ± 3.0 min (mean ± SEM). Of this 13.9 ± 1.5 min was spent in stage N1, 26.5 ± 2.6 min in N2 and 21.0 ± 2.9 min in N3 (slow-wave sleep). All but 4 participants reached stage N3. The duration of REM sleep was 5.5 ± 1.5 min, although only 11 of 22 participants reached this stage of sleep in their nap. We tested whether time in minutes in N3

or REM sleep (after controlling for total sleep time) was correlated with the change in a participant's level of prime-sentence consistency for primed items from Session 1 to Session 2 (i.e., Primed-S2 – Primed-S1), but neither partial correlation approached significance level (N3: $r = 0.20$, $p = 0.38$; REM: $r = 0.05$, $p = 0.83$).[2] We also analysed the sleep data in terms of the prevalence of sleep spindles (spindle density per minute) during N2 and N3 stages of the nap (Ferrarelli et al., 2007). In the absence of any prior expectation of one particular electrode location being influential, we opted to use the average density across frontal and central electrodes as our predictor, but this was not associated with the behavioural change measure ($r = 0.067$, $p = 0.767$).

---

[2] The partial correlation with REM sleep used all participants, including those that had 0 min of REM sleep.

2.3. Discussion

Experiment 1 was the first attempt to examine the robustness of word-meaning priming over relatively extended periods of time (2 and 12 h) that varied in terms of whether the intervening period included sleep or not. We found that the change in priming across the interval depended on whether the participant slept or remained awake during the interval. Participants who stayed awake throughout the interval showed reduced priming at the final test (−4.0%) whereas in the conditions for which the interval included sleep showed stronger priming (+2.5%). The two protocols we used varied substantially in terms of the duration of the interval (2 h vs. 12 h) and the way in which the intervening period was filled (short sleep in the lab vs. long sleep at home; watching silent video vs. conducting normal daytime activities). As one might expect, the sleep effects were numerically a little stronger for the 12-h manipulation than the 2-h manipulation, but there was no significant effect of interval duration, suggesting that the sleep benefits in terms of word-meaning priming applied across a range of durations.

This behavioural effect is consistent with the notion that the sleep interval facilitated the consolidation of word-meaning priming. This was a prediction of the contextual binding account, whereby ambiguous words are associated with their sentential context during comprehension. This new association is then consolidated during sleep in much the same way that memory for pairs of associated words are consolidated during sleep (e.g., Plihal & Born, 1997). If we had observed a significant correlation between the changes in performance and slow-wave sleep duration or spindle activity for the participants napping in the lab, this would have strengthened the case for the effect being driven by consolidation during sleep. Nonetheless, the absence of a significant correlation is not strong evidence against sleep having a causal role in the preservation of word-meaning priming. As mentioned, although the key effects we observed did not interact with interval duration, the nap participants still showed numerically weaker priming after sleep (4%) as compared with the overnight participants (8%). Therefore it is possible that the behavioural effects after a short sleep were too weak to reveal correlations with sleep parameters in our sample. This point is particularly pertinent to our subsidiary test of whether REM sleep might be associated with changes in the strength of word-meaning priming. Because the nap was an afternoon nap it was not optimised for variation in REM duration: only half the participants actually reached REM sleep. Although it should be noted that nap studies have in the past been useful for understanding the involvement of REM sleep in memory consolidation (e.g., Batterink, Westerberg, & Paller, 2017), further research (e.g., using overnight sleep recording) would be valuable to identify what aspects of sleep if any are influential.

The main results were derived comparing sets of materials that were tested for the first time either in Session 1 or Session 2. A secondary result of interest was the pattern of priming for words that were tested in both sessions. In contrast to the results for unrepeated items, the wake and sleep groups showed similar performance across sessions for the repeated items. That is, the drop-off in priming for the wake participants using unrepeated stimuli was not seen for the repeated items. This suggests that the act of generating an associate of a word can in itself represent a learning experience, but this effect is not differentially affected by sleep relative to wakefulness. We return to this point in the discussion of Experiment 2.

Although the key results of Experiment 1 are consistent with the contextual binding account, it is worth considering whether they are strong evidence against the immediate alteration account that has been previously put forward as an explanation of word-meaning priming. This account would not necessarily predict a consolidation effect, but it might predict that interference from new episodes of learning reduces the strength of word-meaning priming effects over time. As Rodd et al. (2016, p. 35) suggested, "One possibility is that the decay function could arise purely due to interference from intervening encounters with other unrelated words: each such encounter would result in weight

changes, which could potentially influence even apparently unrelated words because these may share some connections within the highly interconnected distributed network." Therefore the benefit of sleep over wake might simply be one of passive protection from interference by reducing the opportunity for encoding of new experiences. This debate over active versus passive accounts of sleep benefits in memory is a familiar one (e.g., Ecker, Brown, & Lewandowsky, 2015; Ellenbogen, Payne, & Stickgold, 2006; Walker & Stickgold, 2010), and studies exploiting methods for enhancing sleeping brain oscillations (Ngo, Martinetz, Born, & Mölle, 2013) and reactivating targeted memories during sleep (cf. Paller, 2017) have largely favoured an active account for memories of object location or word pairs. Nonetheless we cannot automatically assume that word-meaning priming is supported by those same active mechanisms.

In Experiment 2 we attempted to tease apart active and passive accounts of the benefit of sleep for word-meaning priming by extending the delay between training and testing. As in the 12-h delay protocol of Experiment 1, we gave two groups of participants an exposure session and initial word-association test either in the evening or the morning and then left them to go about their normal activities outside the lab. However, instead of recalling the participants after 12 h, instead we recalled them after 24 h. These two groups should therefore have equivalent amounts of sleep and equivalent amounts of time spent awake, with the only difference being whether wake comes before or after sleep.

We know from Experiment 1 that word-meaning priming largely dissipates after 12 h spent awake. Therefore, for a passive account of the sleep benefit in Experiment 1 to hold, both groups in Experiment 2 should show equally dissipated word-meaning priming effects after 24 h, because both groups have spent a whole day awake encoding experiences that should interfere with the word-meaning priming effect. On the other hand, if sleep actively stabilizes word-meaning priming then the participants who sleep soon after exposure should show stronger effects after 24 h than those who have had a whole day for word-meaning priming to decay prior to sleep. In other words, sleep should largely crystallize the benefit found in Experiment 1 after 12 h. For the participants encoding in the morning, we know that there is very little priming left to crystallize overnight, whereas those encoding in the evening still show a strong word-meaning priming effect prior to sleep, which should then be preserved over the next 24 h (with perhaps some forgetting).

3. Experiment 2

3.1. Methods

3.1.1. Participants

Participants were undergraduates at the University of York with English as a first language. In total, 55 participants were tested, with participants randomly assigned to an exposure session either in the evening (PM group; N = 28) or in the morning (AM group; N = 27). The groups were well-matched for mean ± SD age in years (PM 19.7 ± 1.7; AM 19.5 ± 1.5) and sex (PM 7 males, AM 4 males). Participant instructions, self-report characteristics and ethical review were as in Experiment 1.

3.1.2. Design, materials and procedure

The design matched Experiment 1 with a small number of exceptions. Rather than a between-participants manipulation of interval duration, all participants had a 24-h interval between testing points (see Fig. 1). Furthermore, the key difference between the two participant groups was no longer sleep versus wake. Instead, the key difference was the time of day of both tests, either in the morning (AM group) or evening (PM group). Within participants, the manipulation of priming was the same. The materials were identical to Experiment 1, and the procedure matched the 12-h interval condition of Experiment 1, except

of course for the duration of the interval, which was always 24 h.

3.2. Results

3.2.1. Sleep questionnaire and Stanford Sleepiness Scale

According to the sleep questionnaire AM participants had a mean ± SD estimated bedtime in the last month of 23:42 ± 55 min and PM participants had a typical bedtime of 23:48 ± 50 min (t (53) = −0.39, p = 0.70). The groups were also well matched on mean ± SD estimated typical sleep duration (AM: 7.81 ± 0.91 h, PM: 7.96 ± 0.78 h; t(53) = −0.70, p = 0.49). All participants apart from one reported that they had not taken any prescription medicine to help them sleep in the last month (AM: 27/27; PM: 27/28). An ANOVA on participant SSS ratings with the within-participants independent variable session and between participants variable group revealed only a significant interaction between Session and Group (F[1, 53] = 4.66, p = 0.035). The groups were well matched on mean ± SD SSS score in Session 1 (AM: 2.37 ± 0.69, PM 2.36 ± 0.62), but for the AM group sleepiness ratings rose in Session 2 (2.70 ± 0.87), whereas they showed a slight fall for the PM group (2.21 ± 0.96). It is unclear why this shift may have taken place, but as in Experiment 1, we tested whether sleepiness ratings could explain priming levels in the word association test by running correlations between every participant's rated sleepiness and level of priming. None of the four tests revealed any association that approached significance level uncorrected for multiple comparisons (all ps > 0.11).

3.2.2. Simple RT

An ANOVA on participant mean RTs with same independent variables as above revealed only a main effect of Session (F[1, 49] = 33.58, p < 0.001), with both groups responding more quickly in Session 2 (mean ± SEM; 385 ± 5 ms) than Session 1 (424 ± 9 ms). As with the sleepiness ratings, we ran correlations between every participant's mean RT and level of word association priming in the two sessions. None of these approached significance level. In sum, both the SSS and RT analyses suggested that the pattern of priming found in the word association test could not be explained in terms of any confound of alertness or sleepiness.

3.2.3. Word association responses

Word association responses were again coded blind as to the experimental condition. We used the same excluded categories, with the proportion of the remaining trials that were consistent with the primed meaning being the dependent variable (see Table 3 and Fig. 4). Mixed effects logistic analyses (see Table 4) followed the same analysis protocol, except that there was only one effect-coded between-participants independent variable: Group (AM vs PM). Within participants, the independent variable Priming again had three levels (Unprimed, Primed-Session 1, Primed-Session 2), coded using orthogonal Helmert contrasts that compared: (i) the Unprimed condition vs. the two Primed conditions (Priming1) and (ii) the two primed conditions with each other

Table 3
Experiment 2 means and standard deviations for the proportional consistency of the word association test responses with the prime sentential context as a function of Group and Priming.

	Priming									
	Unprimed		Primed-S1		Primed-S2		Unprimed-R		Primed-R	
Group	M	SD	M	SD	M	SD	M	SD	M	SD
AM	0.23	0.09	0.33	0.11	0.26	0.07	0.24	0.09	0.32	0.10
PM	0.22	0.08	0.31	0.08	0.32	0.08	0.25	0.08	0.29	0.09

Note. M and *SD* represent mean and standard deviation, respectively. S1 = Session 1, S2 = Session 2, R = repeated.

(Priming2).

As in Experiment 1, there was an overall effect of Priming1, showing that there were more responses consistent with the prime sentences for the primed items than the unprimed items. Here, there was also an effect of Priming2, with stronger priming in Session 1 than Session 2. However, the interaction between Group and Priming2 showed that this reduction in priming across session was not uniform. The group trained in the morning showed a reduction in sentence-consistent responses over 24 h from 32.7% to 26.5% (Z = 3.03, p = 0.002*), whereas the sentence-consistent response percentage for participants trained in the evening was if anything slightly higher after 24 h (32.0%) than at initial test (30.9%) (Z = 0.03, p = 0.97). The intercept effect was significant, and the remaining effects were non-significant.

As before, a subsidiary analysis examined just the two sets of items in Session 2 that had already been tested in Session 1 (see Tables 3 and 4). There was an overall effect of Priming, with similar levels of priming in the AM and PM groups (see Fig. 5). Aside from the intercept, no other effects reached significance level.

3.3. Discussion

Experiment 2 was intended to discriminate between active and passive accounts of the sleep benefit found in Experiment 1 for sentence priming of ambiguous word meanings. If this effect was due to sleep passively protecting the priming effect by reducing the potential for retroactive interference from language then 12 h of interference awake post-sleep should have the same interfering effect as 12 h of wake interference prior to sleep. This means that both groups of participants would be predicted to show the same reduction in priming 24 h after sleep. However, this pattern of results was not found. Participants who were trained in the morning initially showed a 9.3% priming effect when tested shortly afterwards, but this dropped to 3.1% after 24 h. Remarkably, the priming effect for participants trained in the evening showed no such reduction: in fact, the 8.5% initial priming was numerically higher (9.6%) after 24 h. This pattern of results suggests that the passive account of the priming effects in Experiment 1 is wrong. Instead, as we flesh out in the General Discussion, these results fit with an explanation in which sleep alters the memory of the recent experience with the ambiguous word to make the new learning more robust to interference or decay during wake.

As in Experiment 1, we also looked at the effect of a 24-h period with sleep or wake first on the sentential priming effects for words that were tested twice. once in each session. Again, this repeated testing showed no significant effect of group, unlike the main analyses. On a methodological level this suggests that for word-meaning priming repeated testing does not provide a clear test, given that the act of testing changes the later outcome. But on a theoretical level this result is interesting as well. Why might the priming effect for AM participants survive the 24-h period (or in Experiment 1 a 12-h period awake) when words are tested shortly after training but not when they are left untested until the second session? It seems likely that the explanation for this difference links in with the well-attested retrieval practice benefit (e.g., Karpicke & Blunt, 2011). When participants are asked to generate an associate of a word, potentially they remember their response, so forming a second memory that can retained or forgotten over time. Participants tended to produce the same associate in response when tested the second time (overall, just over 50% of repeated words also had repeated associate responses), which diminishes the potential for group differences to emerge. Furthermore, retrieval practice has been posited as a form of consolidation (Antony, Ferreira, Norman, & Wimber, 2017), which might further reduce the potential for sleep/ wake differences to be found.

4. General discussion

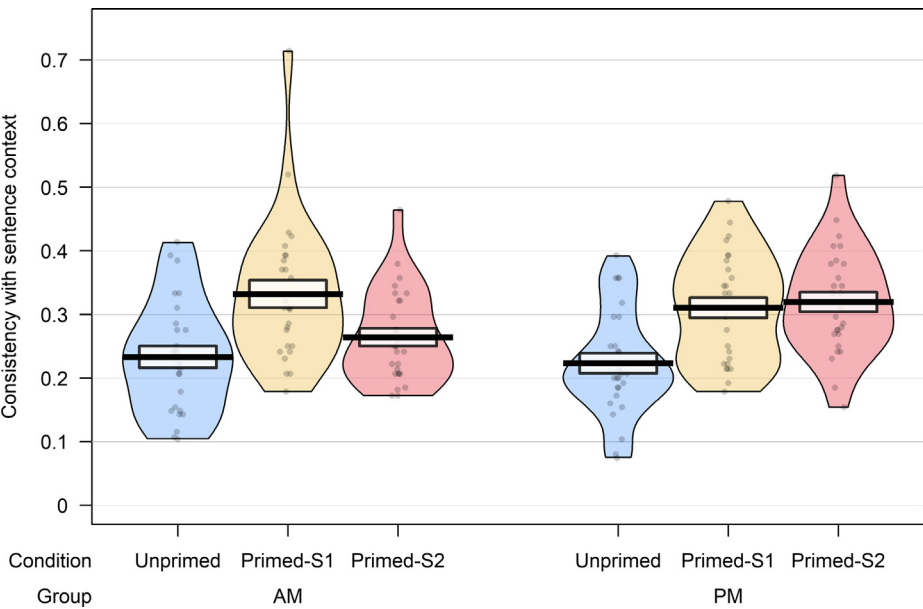The experiments presented here were intended to shed further light

Fig. 4. Experiment 2 priming effects for wake and sleep groups based on participant means in the consistency of word association responses with the primed meaning. The horizontal filled bar represents the mean, with the band marking standard error. Points represent individual participant means with the coloured areas representing smoothed density.

Table 4
Summary of the mixed effects analyses for Experiment 2.

Fixed effect	Estimate	Std error	z value	p(> |z|)
Main analysis				
(Intercept)	−1.264	0.150	−8.452	< 0.001***
Group	0.026	0.039	0.661	0.509
Priming1	0.181	0.039	4.701	< 0.001***
Priming2	−0.106	0.052	−2.052	0.040*
Group:Priming1	0.033	0.032	1.030	0.303
Group:Priming2	0.108	0.046	2.339	0.019*
Repeated items				
(Intercept)	−1.313	0.154	−8.515	< 0.001***
Group	−0.028	0.048	−0.587	0.557
Priming	0.182	0.065	2.815	0.005**
Group:Priming	−0.039	0.053	−0.737	0.461

on the mechanism that provides plasticity in our understanding of word meanings. The data from two experiments reveal a clear and coherent pattern. Like several other studies (Betts et al., 2018; Gilbert et al., 2018; Rodd et al., 2013, 2016), we found evidence that encountering a lexically ambiguous word in a disambiguating sentential context alters the word's representation in a way that has consequences for its future usage. When presented 20 min later, participants tended to generate associates that were consistent with this sententially primed meaning. Consistent with Rodd et al. (2016), this priming effect tended to fade or decay over time spent awake. However, in Experiment 1, we found that when sleep (either a nap or overnight) followed reasonably swiftly after the contextual exposure a priming effect was observed with no evidence of decay over the period of sleep. Furthermore, in Experiment 2 we found that this period of sleep stabilised the primed words' representations such that no decay in priming was seen across the subsequent day spent awake.

As described in the introduction, Rodd et al. (2016) argued that

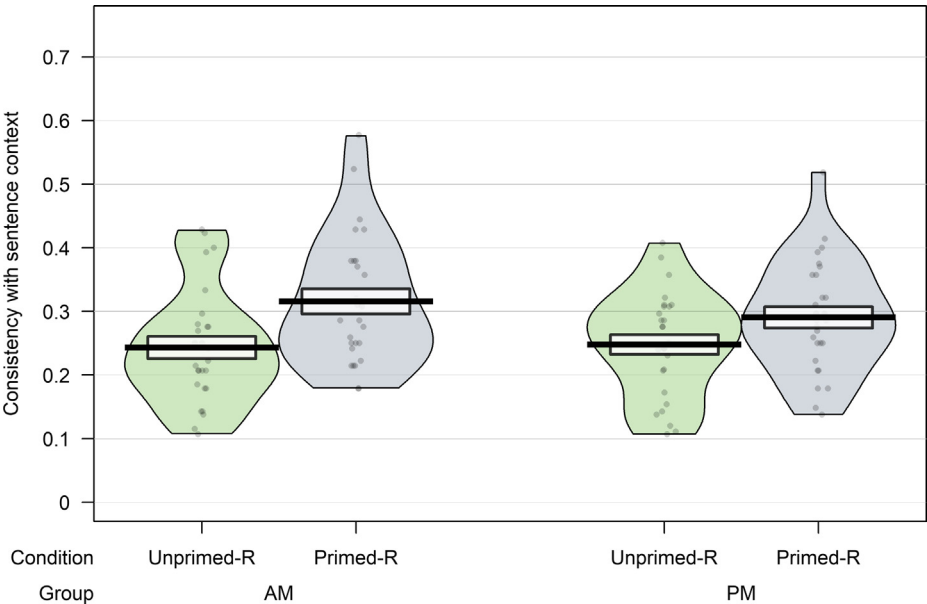
Fig. 5. Experiment 2 priming effects for repeated items in Session 2. Horizontal bars represent the mean, with bands marking standard error. Points represent individual participant means, with the coloured areas representing smoothed density.

word-meaning priming effects of this type can be explained with reference to a connectionist network model of word recognition that maps from a representation of word form to a distributed representation of meaning. In the case of lexical ambiguity, the network learns through linguistic experience to map from one word-form to multiple meanings. Once the multiple meanings have been learned, recognition of an ambiguous word will briefly activate a weighted "blend" of the meanings, with meaning frequency and context of the word on each occasion helping the network to settle into the appropriate meaning's attractor state. Word-meaning priming in such a model is an extension of the process that acquires the mapping in the first place. That is, the exposure phase of the experiment causes a small immediate alteration in the connections in the network from form to meaning, making the contextually appropriate meaning a little easier to access and the inappropriate meaning a little harder to access. This weight adjustment process then influences the later word association test, meaning that words associated with the primed meaning are more likely to be retrieved.

This immediate alteration model provides a clear mechanism to explain the basic word-meaning priming effect, and can perhaps also explain the decay in the priming effect over time awake in terms of further weight changes in response to subsequent linguistic input. However, the current results are more challenging. In particular, the fact that sleep appears to make word-meaning priming more robust and resilient to further interference/decay is not predicted. In order to explain this sleep benefit, the immediate alteration model might appeal to the synaptic homeostasis hypothesis (SHY; Tononi & Cirelli, 2014), which explains consolidation not at a systems level but at a more local level. By this account, the learning that takes place as a consequence of encountering an ambiguous word in its subordinate context involves a strengthening of the cortical synapses relating to that meaning. SHY then predicts that synaptic downscaling during sleep enhances plasticity, while at the same time effectively silencing synapses that may be peripheral or irrelevant to the representation of that meaning. This has the consequence of improving the signal-to-noise ratio of the representation of the recent learning experience, hence boosting retention. This would allow an immediate alteration explanation to accommodate a benefit of sleep over wake. While we should not rule out such an explanation at this point, it is worth noting that the recruitment of SHY to explain sleep benefits also makes it harder to explain why word-meaning priming should show quite strong decay during wake, given that the cortical strengthening during learning reduces plasticity and hence susceptibility to interference. It is hard to see how a single network model can accommodate at the same time both the benefits of sleep to subsequent word-meaning priming and the decay of priming that is seen prior to sleep.

We think that the current results are more amenable to the contextual binding account described in the introduction, in which there is a division of labour in a complementary systems model between the main long-term repository for established lexical knowledge (a cortical network along the lines of the model described above) and the network that provides shorter-term plasticity to enable learning. By this account, the exposure phase of the experiment does not necessarily change connection weights in the long-term cortical network to any substantial degree. Instead, on encountering a sentence the hippocampus is recruited to bind together essentially a new memory of that sentence or utterance, combining the various components of the sentence in a similar way to the formation of a new association between words (Eichenbaum & Bunsey, 1995). This new memory in the short term presumably allows listeners to keep track of, and act on, conversations or other forms of dialogue. It also provides a source of information that can be used when participants are asked to generate an associate of the word encountered in the exposure phase. In these circumstances, an unprimed word can only make use of the cortical long-term lexical network, whereas a primed word can make use of both the cortical network and also the hippocampal representation of the recent

experience with that word.

Importantly, like many other hippocampal memories, the new hippocampal memory should be susceptible to decay over time (Hardt et al., 2013) but can be consolidated via offline replay chiefly during sleep (Bendor & Wilson, 2012), integrating the new material with existing knowledge of the meanings of the constituent words. The decay explains why long periods awake prior to sleep lead to a diminishing of the word priming effect. In turn, the involvement of consolidation explains why word priming remains after sleep, as long as the sleep begins before too much wake decay has occurred. It is important to note that we do not need to assume that consolidation is complete after a single period of sleep. Instead it may be that consolidation is partial, with the hippocampal memory still present perhaps still susceptible to interference, or perhaps more robust as a consequence of consolidation within the hippocampus (cf. Winocur, Moscovitch, & Bontempi, 2010). That said, there is good reason to think that consolidation processes should be completed relatively quickly, given that the long-term consequence of learning in this case is an adjustment of pre-existing weights. This is similar to other examples in which prior schema knowledge can support the rapid consolidation of new information into the cortex (Lewis & Durrant, 2011; McClelland, 2013; Tse et al., 2007).

Given that new learning is simply a matter of connection weight adjustment, one might ask whether a complementary systems model should need to rely in hippocampal mediation in the first place. One of the key computational arguments for the necessity of a hippocampal network in a complementary systems model has been the potential for catastrophic interference (McClelland, McNaughton, & O'Reilly, 1995). This occurs when a distributed connectionist network is required to learn a new mapping that is unlike existing mappings. However, in this case there is no new mapping, just an existing mapping that requires updating. Mirković and Gaskell (2016) presented data that suggested that in the acquisition of a new artificial language, arbitrary mappings (high risk of catastrophic interference) showed sleep benefits whereas systematic mappings (low risk of catastrophic interference) did not. Why then, does the current paradigm show such clear benefits of sleep? Mirković and Gaskell suggested that the prioritization is an important aspect of sleep's role in consolidation (Stickgold & Walker, 2013). In the Mirković and Gaskell study, the whole artificial language was new, and so the hippocampal system (and consequently sleep) may have intervened in the component of the language for which it was most needed, namely the arbitrary aspects. In the case of word-meaning priming, no elements are new, and so the hippocampal system can intervene for mappings that are less in need of rescue from catastrophic interference.

It is important to note that previous results from the word priming paradigm have been taken as evidence against an episodic model of word-meaning priming. In the light of the current data we should review this evidence. First, Rodd et al. (2013) showed that word-meaning priming does not vary depending on whether the speaker of the priming context sentence matches the speaker of the ambiguous word used as the cue in the word association test. In fact, they do not even need to match in terms of the modality of presentation. Gilbert et al. (2018) showed that there was no reliable difference in priming levels for between-modality priming (e.g., auditory prime, visual test) as opposed to within-modality priming. These data point very clearly towards a relatively high-level model of priming that does not depend on a detailed perceptual match between prime sentence and test cue.

Surface detail and modality matter to a great extent for paradigms such as repetition priming (e.g., Goldinger, 1998; Luce & Lyons, 1999), but equally there is much evidence to suggest that many aspects of priming for spoken words is insensitive to such details. Furthermore priming based on similarity at a surface level may have a slower, more effortful mechanism than priming that relies on match at an abstract level (e.g., McLennan & Luce, 2005). Our proposal based on the current evidence is that word-meaning priming is supported the same medial temporal lobe circuitry that underpins paired-associate learning, which

at least in part may have a more abstract element (Cairney, Sobczak, Lindsay, & Gaskell, 2017), linking the words at a lexical or conceptual level rather than a more superficial level.

A second argument was based on the incidence of words from the priming sentence context occurring as responses in the word associate test. If word-meaning priming relies on learning direct associations between the ambiguous word and the words in its context then one might expect these context words to crop up as associate responses reasonably often. Rodd et al. (2013) found that this type of response was relatively rare (e.g., 3% of all responses in Experiment 1), with a modest increase in prevalence for primed (4%) vs unprimed (2%) conditions. The word-meaning priming effect remained when these responses were excluded. We found similar figures in the current experiments: 3% of responses in untrained conditions were words from the context sentences, and this rose to 4% in the conditions for which we saw priming. Clearly, then, the vast majority of associate responses were not simply taken from the context. Interestingly, though, the incidence of these repetitions of context were not even across sentences. For most items (49/87) context words were never reused as associate responses at test, probably because these sentences provided a bias without containing any direct lexical associates of the ambiguous words (e.g., for the word *fence*, the context sentence was "He wanted to learn how to fence"). However, when there was an obvious associate of the ambiguous word in the sentence, then this associate tended to be used a little more often in the primed conditions (e.g., for *pipe*, participants might respond with *smoke* for "The grandfather picked up his pipe to smoke". On closer analysis, then, we think such findings are not incompatible with a complementary systems account of word-meaning priming in which the new association between the ambiguous word and its sentential context provides a source of information to guide the subsequent word association response one of two ways:

(1) By directly boosting the likelihood of responding with a word that is both contained in the sentence context and is an established associate of the ambiguous word.
(2) By biasing the retrieval of an associate of the ambiguous word that has some semantic relation with the theme of the context sentence.

It remains possible, then, that all the previous findings using the word meaning paradigm can be reinterpreted in terms of a contextual binding account by which the hippocampus has a crucial role in relating ambiguous word meanings to their sentential context. However, although the current results provide evidence for this contextual binding mechanism, it is not necessarily the case that the immediate alteration account has no role to play at all. Rather, it is possible that for the ambiguous words in sentence context, in fact *both* cortical and hippocampal systems operate in parallel, perhaps explaining why in the current results the AM participants showed some suggestion of priming even 24 h after initial exposure. A goal in future work would be to determine whether there are ways of further teasing these accounts apart, perhaps by testing the effects of sleep and wake delays on different dependent measures.

As discussed in the introduction, we see our model as compatible with the work of Duff and Brown-Schmidt (2017), who have provided a substantial body of evidence from amnesic participants relating to the involvement of the hippocampus and medial temporal lobe in the everyday use of language, including the updating and enriching of both lexical semantic representations (Klooster & Duff, 2015; MacKay, Stewart, & Burke, 1998) and syntactic biases (Ryskin et al., 2018). Converging evidence can also be found from fMRI evidence suggesting that the left hippocampus is sensitive to the comprehensibility of spoken sentences (Davis & Johnsrude, 2003). Studies using direct EEG recording in the medial temporal lobe also implicate the hippocampus in normal language comprehension. For example, Meyer et al. (2005) recorded event-related potentials (ERPs) of epilepsy patients in rhinal cortex and hippocampus during the perception of different types of sentence violation. They found that semantic violations generated an N400 response (Kutas & Hillyard, 1980) in rhinal cortex in the anterior medial temporal lobe, whereas syntactic violations elicited a later negativity (500–800 ms) in the hippocampus. Both these effects point to the involvement of the medial temporal lobe in the process of sentence comprehension. With particular relevance to the current study, Piai et al. (2016) recorded oscillatory activity directly in the hippocampus for sentences that varied in the contextual constraints that predicted a final word that had to be named, given a picture (e.g., "She walked in here with the…" vs. "She locked the door with the…" followed by a picture of a key). They found that the level of theta activity during the course of the lead-in sentence was dependent on the constraint strength, with the strong-constraint condition showing greater activity. Indeed, within the strong constraint condition, theta activity appeared to track the semantic predictivity of the sentence (as measured by latent semantic analysis) on a word-by-word basis. Hippocampal theta power is strongly implicated in memory processes of recollection of previously studied material (e.g., Guderian & Düzel, 2005) and episodic memory retrieval (e.g., Lega, Jacobs, & Kahana, 2012), although it is important to note that this oscillation also has associations with other aspects of cognitive function such as spatial and time coding (Korotkova et al., 2018). Nonetheless, perhaps surprisingly given earlier evidence (e.g., Kensinger, Ullman, & Corkin, 2001), the emerging understanding from recent studies is of a role for the hippocampus not just in learning language, but also in language comprehension, production and maintenance (see also Cross, Kohler, Schlesewsky, Gaskell, & Bornkessel-Schlesewsky, 2018, for a proposal on the involvement of hippocampal theta in incremental sentence comprehension).

The current results represent the first evidence for a contextual binding account implicating sleep and consolidation in comprehension of lexically ambiguous words. Unsurprisingly, several important questions remain. One is the generality of this mechanism. Is it focused on cases of lexical ambiguity, or does it operate for all utterances and all words irrespective of ambiguity? Indeed it is feasible that contextual binding plays a prominent role only when the subordinate meaning of an ambiguous is retrieved. We can only speculate as to the answer at this point, but a parsimonious working hypothesis would be that the contextual binding we see here as benefiting from sleep is a general component of comprehension, but perhaps (consistent with Piai et al., 2016) sensitive to the degree of predictability of each word meaning. This would link in well with models of memory that see prediction error as driving encoding (Henson & Gagnepain, 2010) and might predict that consolidation effects are stronger when the subordinate (unlikely) meaning of a word is comprehended as compared with the dominant (likely) meaning.

A second area of uncertainty is at what level of representation this contextual binding occurs. Thus far we have been describing the binding as relating to a collection of words, suggesting quite a superficial level. However, the observation here and elsewhere that participants rarely actually reproduce context words in their word-association test responses provides some evidence that the contextual binding occurs at a higher, more gist-like, level. One possibility is that the process we identify here is related to the generation of a situation model of the comprehended utterance, fitting with a construction-integration account of discourse comprehension (Kintsch, 1988). This would be consistent with the common finding that memories for sentences that endure over several days (perhaps through consolidation) tend to be gist-like rather than verbatim representations of the original material (Kintsch, Welsch, Schmalhofer, & Zimny, 1990).

A third outstanding question is to what extent the priming that we see here is related to explicit and/or intentional retrieval of the context sentences either in their verbatim form or (more likely) a more gist-like form. Priming is often thought of as implicit and automatic, but this is not necessarily true in all cases. Here, the proportion of word-association trials that related to the sentences heard in the exposure phase is quite high (50% in Session 1, 67% in Session 2), and the participants

were not under strong time pressure. Furthermore, the exposure and test phases were all in the same physical environment, providing a contextual cue at test to the exposure session. To some extent, these issues have been addressed in previous studies of word-meaning priming. For example, Experiment 1 of Rodd et al. (2016) involved exposure during the course of a radio show and then a later online test. In those circumstances environmental cues were likely to be more variable and yet word-meaning priming was still found. Experiments 3 and 4 were even more naturalistic in that they tested individuals with particular experience of specific meanings of ambiguous words (rowers tested on words such as *feather* that had rowing-related meanings). Likewise, word-meaning priming effects have been found using more time-critical tests such as speeded semantic relatedness judgements (e.g., Cai et al., 2017; Gilbert et al., 2018), allowing less time for intentional retrieval of the context sentences to occur. Nonetheless, as mentioned earlier it is not necessarily the case that all effects that are described as word-meaning priming are must be attributable to the same single mechanism, This means that future studies that assess the contribution of consolidation in the absence of environmental matching cues or with speeded tasks would be valuable to complete the picture in terms of the mechanism or mechanisms that contribute to word-meaning priming. Nonetheless, we find it striking that the word meaning effects that we have observed here can be found even 12 or 24 h after the exposure session despite there being no instruction to memorise the sentences, and no way in which the key ambiguous words were flagged up as important or special. It would be surprising if participants worked hard to intentionally encode or retain the sentences, as there was absolutely nothing for them to gain in doing so.

Part of the motivation for this work was to address the question of the extent of involvement of consolidation in language plasticity. We have shown that consolidation is not only important for learning of completely new material. Instead it seems to play a role in the refinement and updating of lexical semantic representations for highly familiar words. When we come across an ambiguous word we need to make use of the context of that word to determine the intended meaning and update our expectations of the word's meaning ready for the next encounter. We have argued that this involves the learning of an association at some level of abstraction between the words or concepts in the sentence so that updating can occur via consolidation, associated with sleep. A potential implication of this result is that effectively *all* language comprehension constitutes a new hippocampal learning experience that may lead to sleep-associated consolidation. The use of lexically ambiguous words merely provides a clear example of how the learning experience employs consolidation to improve our predictive model of language.

Appendix A. Sentences used in semantic relatedness task, with ambiguous word and the subordinate sentential context used in priming

Item	Homophone	Prime sentence	Dominant meaning	Subordinate (primed) meaning	Probe
1	Fence	He wanted to learn how to fence	post	sword sport	butter
2	Button	Although he was told not to, Fred pushed the button	clothes fastener	pressure-operated switch	grass
3	Sun/Son	The mother made a packed lunch for her son	solar body	male offspring	bee
4	Bed	There were weeds growing in the bed	piece of furniture for sleeping	part of garden where plants grow	theatre
5	Blue/Blew	It was so windy the tent nearly blew away	colour	created air current with mouth	notebook
6	Cricket	Only the male cricket can produce a sound	type of sport	type of insect	tiger
7	Jam	The roadworks caused a jam all through the town	fruit preserve	obstruction/tight spot	wallet
8	Bowl	The sportsmen's bowl won the game	dish	to throw ball in a sports game	chocolate
9	Gear	The new employees were told to put on the work gear	machine/engine part	equipment/clothing for specific purpose	particle
10	Spade	The gambler knew that his opponent wanted a spade	digging implement	card suit	hair
11	Temple	One of the pressure points of the body is the temple	place of worship	side of the forehead	taxi
12	Cap	The purchasers were subject to a spending cap	type of hat	limit/restriction	soup
13	Cold	All the employees had caught the same cold that week	low temperature	viral illness	earplug
14	Landing	There was very little space on the first floor landing	returning to the ground from the air/water	platform between flights of stairs	mouth
15	Pupil	Lizzie was the best pupil in the class	part of the eye	student	garage

16	Gum	The boy could feel his new tooth coming through his gum	chewing sweet	part of mouth	violin
17	Organ	The young woman wanted to learn how to play the organ	body part	musical instrument	slipper
18	Night/ Knight	The armour was shiny after the knight polished it	opposite of day	medieval soldier	parachute
19	Park	With his new car, he struggled to park	public recreational area	to position a car	bench
20	Stitch	The athlete was in a lot of pain because of the stitch	portion of thread sewn through material - object/action	cramping pain in the side with exercise	motorbike
21	Bonnet	Petals decorated the brim of the girl's bonnet	car engine cover	close-fitting headgear	napkin
22	Case	The crime was suspected to be a case of mistaken identity	container for something	legal or police matter	frog
23	Pipe	The grandfather picked up his pipe to smoke	hollow tube	smoking implement	nurse
24	Mould	He wanted to use tools to help him mould the statue	fungus	make a malleable substance into a specific shape - action/ equipment used	lettuce
25	Panel	There was a temporary wooden panel separating the two rooms	group of people	thin section of solid material	eagle
26	Watch	Rebecca wanted to watch the puppy all day	device for keeping time	observe	castle
27	Appendix	The author put his memos in the appendix of the book	small redundant part of the intestines	supplement to a text	house
28	Bulb	The wire connecting the bulb was broken	underground part of plant	light component	tomato
29	Bark	The branches and the bark had been damaged by the storm	dog noise	outer covering of tree	elephant
30	Hand	The clock had a broken hand, so it didn't give the time	body part	indicator on clock	camera
31	Sea/See	He sat up straight but still couldn't see	expanse of water	perceive with eyes	igloo
32	Hair/Hare	The fox tried to chase the hare	fine filaments that grow from the skin	rabbit-like animal	rainbow
33	Flower/ Flour	The bread was made using wholemeal flour	part of a plant	finely ground grain	twig
34	Sign	After giving her name and the date, she had to sign the contract	gives information to travellers	to write one's name	mushroom
35	Skip	The home owners were advised to hire a skip	jumping game with rope	used to hold rubbish	question
36	Ball	There were many professional dancers at the ball this year	spherical object	dancing party	spider
37	Toast	The host was asked to make a toast	grilled bread	speech before drinking	shadow
38	Strike	A policy amendment led to strike across the profession	aim/deliver an attack/blow/stroke/offensive manoeuvre - military/sporting -literal/ figurative - act/action	official work stoppage - act/ action	biscuit
39	Palm	The start of the private beach was indicated with a palm	inner side of hand	type of tree	gadget
40	Record	Today was the hottest day on record	grooved disc for playing music	account preserving knowledge of events/information	plastic
41	Match	The derby was certainly going to be an exciting match	fire-lighter	sports game	blanket
42	Note	The lower note suited the singer's voice better than the higher one	brief written communication	musical tone	coin
43	Break	The vase was wrapped so that it wouldn't break	period of rest - occurrence/act of taking	damage	jumper
44	Band	Everyone in the group wore a band	group of musicians	strip of material	sellotape

45	Spring	To make it jump, the toy had a spring inside it	season	elastic coil - object/action of/action resembling	drawer
46	Pear/Pair	Since the heel was broken, she bought a new pair of shoes	type of fruit	group of two things	crystal
47	Cross	His mum was extremely cross with him	symbol with two intersecting lines	annoyed	heart
48	Plug	The plumber had forgotten to put the plug in	connects appliance to electricity	blocks a drain	lion
49	Iron	A simple school science experiment involves iron filings	smoothing device	type of metal	potato
50	Step	She daren't take one more step in the area	part of a staircase/doorway	leg movement	cushion
51	Bow	The girl tied a bow around her ponytail	weapon for shooting arrows	looped knot	lemon
52	Change	The cashier had given the customer the wrong change	alteration	money	street
53	Racket	Adam turned the TV up because of the racket	stringed bat	noisy disturbance	hairbrush
54	Trailer	The new film trailer was released yesterday	cargo vehicle that is towed	short promotional film	calendar
55	Chest	The large wooden chest was covered in dust	part of torso	box	leaf
56	Bat	The fruit bat is a flying mammal	wooden sport stick	night animal	cupboard
57	Card	She didn't have a Christmas card to give him	piece of stiff paper used in games	piece of stiff paper sent as a greeting	moustache
58	Speaker	Dan connected his iPhone to the speaker	person talking	device for playing sound	window
59	Figure	All of the bankers knew what the figure would be	body shape	numerical symbol/amount	truck
60	Key	The musician altered the song's key several times	device that operates a lock	button on computer keyboard	trolley
61	Pen	The pig pen was muddier than ever	writing instrument	animal enclosure	wood
62	Glasses	She poured the champagne into the glasses	lenses to correct eyesight	glass vessels for drinks	cotton
63	Joint	The police searched the suspected drug dealer and found a joint	where two bones meet	drugs rolled to smoke	bottle
64	Cup	The plaque on the cup was engraved	vessel for hot drinks	trophy	curtain
65	Mouse	The Apple iMac boasts a new mouse or track-pad device	small rodent	computer pointing device	tongue
66	Punch	The guests made the most of the free punch	hit with a fist	type of fruit-based drink	tower
67	Ring	She picked up the phone to give her son a ring	piece of jewellery	telephone call/noise	floor
68	Coach	Lee was the most respected coach in the business	vehicle	sports trainer	computer
69	Bar	Access was prevented with a long wooden bar	pub	length of solid material	screen
70	Box	The fighter had to box better than he had before	rectangular container	to play in fighting sport	song
71	Trunk	Pat lifted the lid of the trunk to check what was inside	elephant's nose	large storage case	cereal
72	Staff	When walking up-hill, the hiker used a staff to help	workers employed	tall strong stick	door
73	Issue	Alex had edited the most recent issue of the publication	topical problem	instance of publication	pillow
74	Interest	The bank charges more interest than others	enthusiasm/attention/curiosity/enjoyment for something - state/action or resulting object/target/pastime	fee paid for borrowing/saving money	road

75	Mark	The lecturer hadn't had time to give the essay a mark	visible trace on a surface/indicator point	grade	tree
76	Letter	The little girl sounded the word out one letter at a time	item of mail	alphabetic character	fork
77	Straw	The foal was born on the straw in the barn	thin tube for sucking up liquid	dry stalks of cereal plants	bin
78	Sink	There was a soap pump on the side of the sink	to become submerged	water basin	lottery
79	Deck	The magician asked the volunteer to pick from the deck	flooring/open platform	pack of cards	lava
80	Cabinet	The ministry advised the cabinet on policy alterations	piece of furniture	group of senior politicians	clothes
81	Wave	He couldn't hear what she said but he saw her wave	disturbance in a body of water	moving one's hand as a greeting	carrot
82	Draw/Drawer	She kept her gloves and scarves in the drawer	to produce a picture	storage compartment	cat
83	Lace	The teacher stopped Lily to tie up her lace	net-like ornamental fabric	cord for fastening shoes	money
84	Nail	She was upset that she had broken her nail	sharp metal fastener	plate on end of finger or toe	rabbit
85	China	She bought a new set of china for the tea party	proper name - country	porcelain	clown
86	Fan	Some celebrities don't interact with members of their fan base	ventilation device	admirer/follower	bird
87	Calf	The muscle in his leg had weakened, particularly the calf	baby cow	lower leg	oven

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.cognition.2018.09.007 and https://osf.io/bft7a/.

References

Anderson, N. D., & Dell, G. S. (2018). The role of consolidation in learning context-dependent phonotactic patterns in speech and digital sequence production. *Proceedings of the National Academy of Sciences.*

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in Cognitive Sciences, 21*(8), 573–576.

Bakker-Marshall, I., Takashima, A., Schoffelen, J.-M., van Hell, J. G., Janzen, G., & McQueen, J. M. (2018). Theta-band oscillations in the middle temporal gyrus reflect novel word consolidation. *Journal of Cognitive Neuroscience (Early Access),* 1–13.

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language, 73,* 116–130.

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*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). Lme4: Linear mixed-effects models using eigen and s4. *R Package Version, 1*(7), 1–23.

Batterink, L. J., Westerberg, C. E., & Paller, K. A. (2017). Vocabulary learning benefits from rem after slow-wave sleep. *Neurobiology of Learning and Memory, 144,* 102–113.

Bendor, D., & Wilson, M. A. (2012). Biasing the content of hippocampal replay during sleep. *Nature Neuroscience.*

Betts, H., Gilbert, R., Cai, Z., Okedara, Z., & Rodd, J. (2018). Retuning of lexical-semantic representations: Repetition and spacing effects in word-meaning priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* https://doi.org/10.1037/xlm0000507.

Cai, D. J., Mednick, S. A., Harrison, E. M., Kanady, J. C., & Mednick, S. C. (2009). Rem, not incubation, improves creativity by priming associative networks. *Proceedings of the National Academy of Sciences of the United States of America, 106*(25), 10130–10134. https://doi.org/10.1073/pnas.0900271106.

Cai, Z. G., Gilbert, R. A., Davis, M. H., Gaskell, M. G., Farrar, L., Adler, S., & Rodd, J. M. (2017). Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology, 98,* 73–101. https://doi.org/10.1016/j.cogpsych.2017.08.003.

Cairney, S. A., Sobczak, J. M., Lindsay, S., & Gaskell, M. G. (2017). Mechanisms of memory retrieval in slow-wave sleep. *Sleep, 40*(9).

Chomet, S. (Writer). (2010). L'illusioniste. In S. Chomet & B. Last (Producer). France.

Warner Bros.

Cross, Z. R., Kohler, M. J., Schlesewsky, M., Gaskell, M. G., & Bornkessel-Schlesewsky, I. (2018). Sleep-dependent memory consolidation and incremental sentence comprehension: Computational dependencies during language learning as revealed by neuronal oscillations. *Frontiers in Human Neuroscience, 12,* 18.

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 364*(1536), 3773–3800. https://doi.org/10.1098/rstb.2009.0111.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *The Journal of Neuroscience, 23*(8), 3423.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1355.

Doyon, J., Korman, M., Morin, A., Dostie, V., Tahar, A. H., Benali, H., ... Carrier, J. (2009). Contribution of night and day sleep vs. Simple passage of time to the consolidation of motor sequence and visuomotor adaptation learning. *Experimental Brain Research, 195*(1), 15–26. https://doi.org/10.1007/s00221-009-1748-y.

Duff, M. C., & Brown-Schmidt, S. (2012). The hippocampus and the flexible use and processing of language. *Frontiers in Human Neuroscience, 6,* 69.

Duff, M. C., & Brown-Schmidt, S. (2017). Hippocampal contributions to language use and processing. *The hippocampus from cells to systems* (pp. 503–536). Springer.

Duff, M. C., Gupta, R., Hengst, J. A., Tranel, D., & Cohen, N. J. (2011). The use of definite references signals declarative memory: Evidence from patients with hippocampal amnesia. *Psychological Science, 22*(5), 666–673.

Duff, M. C., Hengst, J., Tranel, D., & Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. *Nature Neuroscience, 9*(1), 140.

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

Ecker, U. K., Brown, G. D., & Lewandowsky, S. (2015). Memory without consolidation: Temporal distinctiveness explains retroactive interference. *Cognitive Science, 39*(7), 1570–1593.

Eichenbaum, H., & Bunsey, M. (1995). On the binding of associations in memory: Clues from studies on the role of the hippocampal region in paired-associate learning. *Current Directions in Psychological Science, 4*(1), 19–23.

Ellenbogen, J. M., Hulbert, J. C., Stickgold, R., Dinges, D. F., & Thompson-Schill, S. L.

(2006). Interfering with theories of sleep and memory: Sleep, declarative memory, and associative interference. *Current Biology, 16*(13), 1290–1294.

Ellenbogen, J. M., Payne, J. D., & Stickgold, R. (2006). The role of sleep in declarative memory consolidation: Passive, permissive, active or none? *Current Opinion in Neurobiology, 16*(6), 716–722.

Fang, X., & Perfetti, C. A. (2017). Perturbation of old knowledge precedes integration of new knowledge. *Neuropsychologia, 99*, 270–278. https://doi.org/10.1016/j.neuropsychologia.2017.03.015.

Ferrarelli, F., Huber, R., Peterson, M., Massimini, M., Murphy, M., Riedner, B., ... Tononi, G. (2007). Reduced sleep spindle activity in schizophrenia patients. *American Journal of Psychiatry, 164*(3), 483–492.

Forster, K. I., & Forster, J. C. (2003). A windows display program with millisecond accuracy. *Behavior Research Methods Instruments & Computers, 35*(1), 116–124.

Friedrich, M., Wilhelm, I., Mölle, M., Born, J., & Friederici, A. D. (2017). The sleeping infant brain anticipates development. *Current Biology, 27*(15), 2374–2380 e2373.

Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology, 22*(7), 615–621.

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

Gaskell, M. G., Warker, J., Lindsay, S., Frost, R., Guest, J., Snowdon, R., & Stackhouse, A. (2014). Sleep underpins the plasticity of language production. *Psychological Science, 25*(7), 1457–1465. https://doi.org/10.1177/0956797614535937.

Gilbert, R. A., Davis, M. H., Gaskell, M. G., & Rodd, J. M. (2018). Listeners and readers generalize their experience with word meanings across modalities. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/xlm0000532.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279.

Gomez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychological Science, 17*(8), 670–674.

Guderian, S., & Düzel, E. (2005). Induced theta oscillations mediate large-scale synchrony with mediotemporal areas during recollection in humans. *Hippocampus, 15*(7), 901–912.

Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: The role of active forgetting in memory. *Trends in Cognitive Sciences, 17*(3), 111–120.

Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, M. G. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science, 15*(5), 674–687.

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus, 20*(11), 1315–1326. https://doi.org/10.1002/hipo.20857.

Hilverman, C., Cook, S. W., & Duff, M. C. (2016). Hippocampal declarative memory supports gesture production: Evidence from amnesia. *Cortex, 85*, 25–36.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. (1973). Quantification of sleepiness: A new approach. *Psychophysiology, 10*(4), 431–436.

Horváth, K., Myers, K., Foster, R., & Plunkett, K. (2015). Napping facilitates word learning in early lexical development. *Journal of Sleep Research, 5*(24), 503–509.

Iber, C. (2007). *The aasm manual for the scoring of sleep and associated events: Rules, terminology and technical specifications.* American Academy of Sleep Medicine.

James, E., Gaskell, M. G., Weighall, A., & Henderson, L. (2017). Consolidation of vocabulary during sleep: The rich get richer? *Neuroscience & Biobehavioral Reviews, 77*, 1–13.

Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience, 10*(1), 100.

Joordens, S., & Besner, D. (1994). When banking on meaning is not (yet) money in the bank: Explorations in connectionist modelling. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1051–1062.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*(6018), 772–775.

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General, 133*(3), 450.

Kawamoto, A. H., Farrar, W. T., & Kello, C. (1994). When two meanings are better than one: Modeling the ambiguity advantage using a recurrent distributed network. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1233–1247.

Kensinger, E. A., Ullman, M. T., & Corkin, S. (2001). Bilateral medial temporal lobe damage does not affect lexical or grammatical processing: Evidence from amnesic patient hm. *Hippocampus, 11*(4), 347–360.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29*(2), 133–159.

Klooster, N. B., & Duff, M. C. (2015). Remote semantic memory is impoverished in hippocampal amnesia. *Neuropsychologia, 79*, 42–52.

Korotkova, T., Ponomarenko, A., Monaghan, C. K., Poulter, S. L., Cacucci, F., Wills, T., ... Lever, C. (2018). Reconciling the different faces of hippocampal theta: The role of theta oscillations in cognitive, emotional and innate behaviors. *Neuroscience & Biobehavioral Reviews, 85*, 65–80. https://doi.org/10.1016/j.neubiorev.2017.09.004.

Kurdziel, L. B., Mantua, J., & Spencer, R. M. (2017). Novel word learning in older adults: A role for sleep? *Brain and Language, 167*, 106–113.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203–205.

Lega, B. C., Jacobs, J., & Kahana, M. (2012). Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus, 22*(4), 748–761.

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences, 15*(8), 343–351.

Luce, P. A., & Lyons, E. A. (1999). Processing lexically embedded spoken words. *Journal of*

Experimental Psychology: Human Perception and Performance, 25(1), 174–183.

MacKay, D. G., Stewart, R., & Burke, D. M. (1998). Hm revisited: Relations between language comprehension, memory, and the hippocampal system. *Journal of Cognitive Neuroscience, 10*(3), 377–394.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type i error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315.

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General, 142*(4), 1190–1210.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning-systems in the hippocampus and neocortex - insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*(3), 419–457.

McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 306.

McMurray, B., Kapnoula, E., & Gaskell, M. (2016). Learning and integration of new word-forms: Consolidation, pruning and the emergence of automaticity. *Speech Perception and Spoken Word Recognition*, 116–142.

Mednick, S. C., McDevitt, E. A., Walsh, J. K., Wamsley, E., Paulus, M., Kanady, J. C., & Drummond, S. P. A. (2013). The critical role of sleep spindles in hippocampal-dependent memory: A pharmacology study. *The Journal of Neuroscience, 33*(10), 4494–4504. https://doi.org/10.1523/jneurosci.3127-12.2013.

Meyer, P., Mecklinger, A., Grunwald, T., Fell, J., Elger, C. E., & Friederici, A. D. (2005). Language processing within the human medial temporal lobe. *Hippocampus, 15*(4), 451–459. https://doi.org/10.1002/hipo.20070.

Mirković, J., & Gaskell, M. G. (2016). Does sleep improve your grammar? Preferential consolidation of arbitrary components of new linguistic knowledge. *PLoS One, 11*(4), e0152489. https://doi.org/10.1371/journal.pone.0152489.

Nadel, L., Hupbach, A., Gomez, R., & Newman-Smith, K. (2012). Memory formation, consolidation and transformation. *Neuroscience & Biobehavioral Reviews, 36*(7), 1640–1645.

Ngo, H.-Viet V., Martinetz, T., Born, J., & Mölle, M. (2013). Auditory closed-loop stimulation of the sleep slow oscillation enhances memory. *Neuron, 78*(3), 545–553. https://doi.org/10.1016/j.neuron.2013.03.006.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*(2), 204–238.

Paller, K. A. (2017). Sleeping in a brave new world: Opportunities for improving learning and clinical outcomes through targeted memory reactivation. *Current Directions in Psychological Science, 26*(6), 532–537.

Piai, V., Anderson, K. L., Lin, J. J., Dewar, C., Parvizi, J., Dronkers, N. F., & Knight, R. T. (2016). Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. *Proceedings of the National Academy of Sciences, 113*(40), 11366–11371. https://doi.org/10.1073/pnas.1603312113.

Plihal, W., & Born, J. (1997). Effects of early and late nocturnal sleep on declarative and procedural memory. *Journal of Cognitive Neuroscience, 9*(4), 534–547.

Plihal, W., & Born, J. (1999). Effects of early and late nocturnal sleep on priming and spatial memory. *Psychophysiology, 36*(5), 571–582.

R Core Team (2013). R: A language and environment for statistical computing.

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews, 93*(2), 681–766.

Reid, A. M. (2013). *The influence of prior knowledge on memory consolidation* (Ph.D.) University of York.

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., & Davis, M. H. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Memory & Cognition, 40*(7), 1095–1108. https://doi.org/10.3758/s13421-012-0209-1.

Rodd, J. M., Cai, Z. G., Betts, H. N., Hanby, B., Hutchinson, C., & Adler, A. (2016). The impact of recent and long-term experience on access to word meanings: Evidence from large-scale internet-based experiments. *Journal of Memory and Language, 87*, 16–37.

Rodd, J. M., Cutrin, B. L., Kirsch, H., Millar, A., & Davis, M. H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language, 68*(2), 180–198.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28*(1), 89–104.

Ryskin, R. A., Qi, Z., Covington, N. V., Duff, M., & Brown-Schmidt, S. (2018). Knowledge and learning of verb biases in amnesia. *Brain and Language, 180–182*, 62–83. https://doi.org/10.1016/j.bandl.2018.04.003.

Ryskin, R. A., Qi, Z., Duff, M. C., & Brown-Schmidt, S. (2017). Verb biases are shaped through lifelong learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(5), 781.

Sadeh, T., Ozubko, J. D., Winocur, G., & Moscovitch, M. (2014). How we forget may depend on how we remember. *Trends in Cognitive Sciences, 18*(1), 26–36.

Sandoval, M., Leclerc, J., & Gómez, R. (2017). Words to sleep on: Naps facilitate verb generalization in habitually and nonhabitually napping preschoolers. *Child Development, 88*(5), 1615–1628.

Stickgold, R., Scott, L., Rittenhouse, C., & Hobson, J. A. (1999). Sleep-induced changes in associative memory. *Journal of Cognitive Neuroscience, 11*(2), 182–193.

Stickgold, R., & Walker, M. P. (2013). Sleep-dependent memory triage: Evolving generalization through selective processing. *Nature Neuroscience, 16*(2), 139–145.

Szmalec, A., Page, M. P., & Duyck, W. (2012). The development of long-term lexical representations through hebb repetition learning. *Journal of Memory and Language, 67*(3), 342–354.

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep

spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience, 30*(43), 14356–14360. https://doi.org/10.1523/jneurosci.3028-10.2010.

Tononi, G., & Cirelli, C. (2014). Sleep and the price of plasticity: From synaptic and cellular homeostasis to memory consolidation and integration. *Neuron, 81*(1), 12–34. https://doi.org/10.1016/j.neuron.2013.12.025.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., ... Morris, R. G. M. (2007). Schemas and memory consolidation. *Science, 316*(5821), 76–82.

Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). University of alberta norms of relative meaning frequency for 566 homographs. *Memory & Cognition, 22*(1), 111–126.

Walker, M. P., & Stickgold, R. (2010). Overnight alchemy: Sleep-dependent memory evolution. *Nature Reviews Neuroscience, 11*(3), 218.

Warker, J. A. (2013). Investigating the retention and time course of phonotactic constraint learning from production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(1), 96.

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology-Learning Memory and Cognition, 32*(2), 387–398. https://doi.org/10.1037/0278-7393.32.2.387.

Warker, J. A., Dell, G. S., Whalen, C. A., & Gereg, S. (2008). Limits on learning phonotactic constraints from recent production experience. *Journal of Experimental Psychology-Learning Memory and Cognition, 34*(5), 1289–1295. https://doi.org/10.1037/a0013033.

Warker, J. A., Xu, Y., Dell, G. S., & Fisher, C. (2009). Speech errors reflect the phonotactic constraints in recently spoken syllables, but not in recently heard syllables. *Cognition, 112*(1), 81–96. https://doi.org/10.1016/j.cognition.2009.03.009.

Williams, S. E., & Horst, J. S. (2014). Goodnight book: Sleep consolidation improves word learning via storybooks. *Frontiers in Psychology, 5*, 184.

Winocur, G., Moscovitch, M., & Bontempi, B. (2010). Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal–neocortical interactions. *Neuropsychologia, 48*(8), 2339–2356.

Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences, 34*(10), 515–525.

Further reading

Gaskell, M. G., Cairney, S. A., & Rodd, J. M. (2018). Contextual priming of word meanings is stabilized over sleep. *Open Science Framework*. https://osf.io/bft7a/.