

Memory consolidation is linked to spindle-mediated information processing during sleep

Cairney, S.A.¹, Guttesen, A.¹, El Marj, N.², Staresina, B.P.^{2,*}

¹*Department of Psychology, University of York, UK.*

²*School of Psychology, University of Birmingham, UK.*

Abstract

How are brief encounters transformed into lasting memories? Previous research has established the role of non-rapid eye movement (NREM) sleep, along with its electrophysiological signatures of slow oscillations (SOs) and spindles, for memory consolidation. More recently, experimental manipulations have demonstrated that NREM sleep provides a window of opportunity to selectively strengthen particular memory traces via the delivery of sensory cues. It has remained unclear, however, whether experimental memory cueing triggers the brain's endogenous consolidation mechanisms (linked to SOs and/or spindles) and whether those mechanisms in turn mediate effective processing of the cue information. Here we devised a novel paradigm in which associative memories (adjective-object and adjective-scene pairs) were selectively cued during a post-learning nap, successfully stabilising next-day retention relative to non-cued memories. First, we found that compared to novel control adjectives, memory cues were accompanied by an increase in fast spindles coupled to SO up states. Critically, EEG pattern decodability of the associated memory category (object vs. scene) was temporally linked to cue-induced spindles and predicted next-day retrieval performance across participants. These results provide highly controlled empirical evidence for an information processing role of sleep spindles in service of memory consolidation.

*to whom correspondence should be addressed:

b.staresina@bham.ac.uk

Introduction

Of the innumerable experiences we encounter throughout the day, only a fraction remain accessible to us in the future. Understanding the principles of consolidation, i.e., the mechanisms by which certain memory traces prevail across time, has remained a major challenge in memory research. Despite the boundaries of current understanding, there is general consensus that non-rapid eye movement (NREM) sleep plays an important role in memory consolidation, particularly for declarative memories that are initially reliant on the hippocampus. According to the *Active Systems* framework¹⁻⁴, consolidation is driven by coordinated interactions between canonical electrophysiological signatures of NREM sleep. In concert with the depolarising up states of cortical slow oscillations (SOs, ~.75 Hz), thalamocortical spindles (~10-16 Hz) are thought to promote the stabilisation of memory traces in neocortical target areas. Accordingly, SOs and spindles have been robustly linked to overnight memory retention in a wide variety of studies⁵⁻¹¹, signifying involvement of these oscillatory phenomena in sleep-dependent consolidation. Importantly, however, the functional role of SOs and spindles in memory processing during sleep has remained elusive.

Given the established role of sleep for memory, post-learning sleep periods might be used to strengthen memories even further by using experimental manipulations. Indeed, the recent advent of 'targeted memory reactivation' (TMR) paradigms has shown that NREM sleep provides a window of opportunity to bolster consolidation of specific memory traces¹²⁻¹⁴. In a typical TMR experiment, new memories are associated with auditory stimuli before a subset of the stimuli are replayed as memory cues during NREM sleep. For example, replaying environmental sounds associated with visuo-spatial memories improves the retrieval of cued relative to non-cued object-locations¹⁵⁻¹⁸. Likewise, replaying verbal foreign language cues facilitates the retrieval of cued (vs. non-cued) vocabulary translations^{19, 20}. These findings are supplemented by a range of behavioural^{21, 22}, neuropsychological²³ and neuroimaging²⁴ studies, indicating that TMR can be used to selectively strengthen declarative memories.

This begs the question of whether TMR in NREM sleep prompts selective memory benefits by exploiting the brain's endogenous offline consolidation mechanisms. Of the neural oscillations unique to NREM sleep, spindles appear to be the prime candidate to support consolidation and task-related information processing in a targeted manner. Spindles tend to be more localised than more global SOs^{25, 26}, and topographical differences in spindle activity have been observed following verbal vs. visuospatial learning^{27, 28}. Moreover, recent work has linked spontaneous EEG activity in the spindle frequency band to memory processing in NREM sleep²⁹, and BOLD increases in learning networks during NREM sleep are modulated by spindle parameters³⁰. These observations suggest

that spindles play a privileged role in mediating consolidation during sleep, and TMR may therefore bolster retention by generating finely-tuned windows of spindle-mediated information processing.

Here, we tested the prediction that experimental memory strengthening relies on deployment of sleep spindles and concomitant processing of task-related information. To this end, we devised a novel paradigm in which we controlled both the timing and content of memory consolidation. In particular, we presented sleeping participants with learning-related verbal cues and assessed (i) whether those cues triggered the emergence of SOs and/or spindles and (ii) whether those oscillatory events in turn mediate cue-specific information processing. Our findings critically revealed that valid memory cues – relative to novel control adjectives – led to an increase in spindle events modulated by SO up states. Moreover, during this time window of cue-induced spindle activity, the mnemonic association linked to the verbal cue could be reliably decoded, with the fidelity of this decoding predicting the behavioural consolidation benefits of TMR.

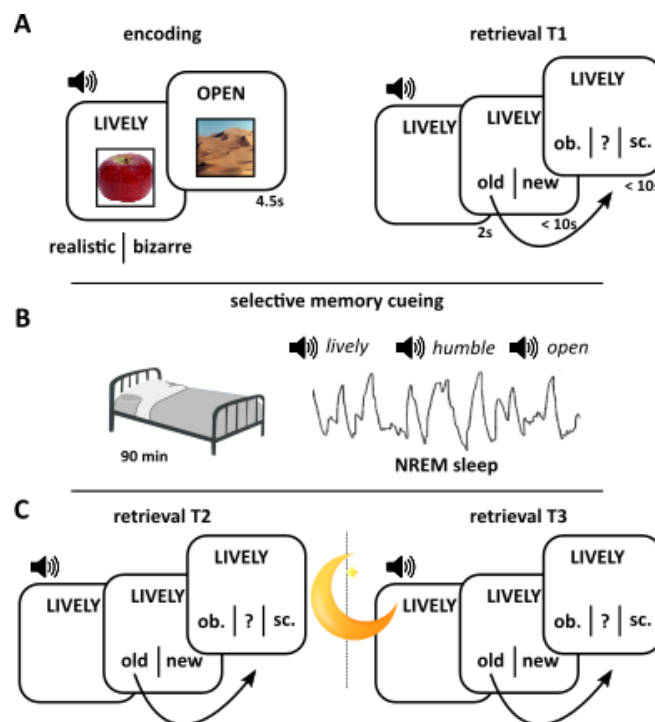


Figure 1: Paradigm. **A.** During encoding, participants were presented with 50 adjective-object and 50 adjective-scene combinations (randomly intermixed) and indicated whether they elicited a realistic or bizarre mental image. Prior to encoding, participants performed a familiarisation phase for both the adjectives and the images (see methods). Approximately 5 minutes after encoding, participants performed the first retrieval session (T1) in which all previously seen (old) adjectives were intermixed with 50 previously unseen (new) adjectives and participants indicated whether they thought the adjective was old or new. In case of an “old” response, they were asked whether they also remembered the associated image category (object or scene) or whether they did not remember the associated category (“?” response option). If they indicated “object” or “scene”, another screen appeared (not shown) in which participants could type in the exemplar if they remembered it or simply type in a “?” if they did not. Adjectives were presented visually and acoustically throughout. **B.** In the

NAP group, participants were set up for polysomnographic recordings and were given the opportunity to sleep for 90 minutes. Once they entered deep NREM sleep, (i) half of the adjectives for which the object category was remembered at T1, (ii) half of the adjectives for which the scene category was remembered at T1 and (iii) a matched number of novel adjectives (controls) were continuously played via external speakers (targeted memory reactivation, TMR). In a matched WAKE group, participants started with 30 minutes of playing the online game Bubble Shooter, followed by 30 minutes of performing a 1-back working memory task during which TMR was applied, followed again by 30 minutes of playing Bubble Shooter. **C.** After the offline period (T2), participants performed the same test as in T1, but with a new set of 50 lure adjectives. Finally, after a night of sleep, participants returned the next morning (T3) for another retrieval session, again with 50 new lure adjectives.

Results

As shown in Figure 1, participants ($n=46$) encoded pairwise associations (adjective-object and adjective-scene pairs) before an initial test phase (T1), in which adjective recognition judgements (old or new) were made. Critically, for recognised adjectives, recall of the associated image (object or scene) was assessed. Half of the correctly recalled adjective-object and adjective-scene pairs were randomly assigned to a cued condition, such that the adjectives would be replayed during the offline period (TMR). The other half of the correctly recalled pairs were assigned to a non-cued condition. The to-be-cued adjectives were also intermixed with a number of non-encoded control adjectives. Immediately after T1, participants either took a 90-min nap (nap group; $n=27$) or remained awake for the same period of time (wake group; $n=19$). In the nap group, TMR took place during NREM stages N2 and N3. In the wake group, TMR coincided with a working memory task to prevent participants from directly attending to the cues^{17, 20}. After the offline period, participants completed a second test (T2), before returning after an additional night of sleep to complete a final test phase (T3).

	<i>T1</i>	<i>T2*</i>		<i>T3**</i>	
		<i>Not Cued</i>	<i>Cued</i>	<i>Not Cued</i>	<i>Cued</i>
<i>Nap</i>	50.15 (± 2.97)	79.57 (± 2.14)	80.69 (± 2.45)	88.48 (± 1.58)	94.36 (± 0.84)
<i>Wake</i>	46.84 (± 2.89)	64.17 (± 4.10)	64.75 (± 3.66)	82.52 (± 3.61)	82.76 (± 3.14)

Table 1: Category recall performance (%) at each test phase. Data are shown as means ± SEM. *Refers to the proportion of T1-recalled categories that were also recalled at T2. **Refers to the proportion of T2-recalled categories that were also recalled at T3.

Behaviour

Category recall at T1 did not differ between the nap and wake groups ($t(44)=0.77$, $P=.45$; see Table 1, Figure 2A). Category retention at T2 (i.e. the proportion of T1-recalled categories that were also recalled at T2) was significantly better after sleep than wakefulness (Group main effect: $F(1,44)=17.10$; $P<.0001$), but unaffected by cueing (TMR main effect: $F(1,44)=0.19$, $P=.66$; TMR*Group interaction: $F(1,44)=0.02$, $P=.89$) (Figure 2C). However, category retention at T3 (i.e. the proportion of T2-recalled categories that were also recalled at T3) revealed both a memory benefit of sleep (Group main effect: $F(1,44)=9.34$; $P=.004$) and a recall advantage from cueing (TMR main effect: $F(1,44)=4.65$, $P=.04$; TMR*Group interaction: $F(1,44)=3.94$, $P=.05$) (Figure 2D). The memory enhancing effects of TMR were driven by the nap group, who exhibited superior retention of cued relative to non-cued categories ($t(26)=3.83$, $P=.001$), whereas no such difference was observed in the wake group ($t(18)=0.09$, $P=.93$). Taken together, these results suggest that TMR in post-learning sleep augmented overnight consolidation processes, improving retention the following day. No differences were observed between object and scene recall (see supplementary materials).

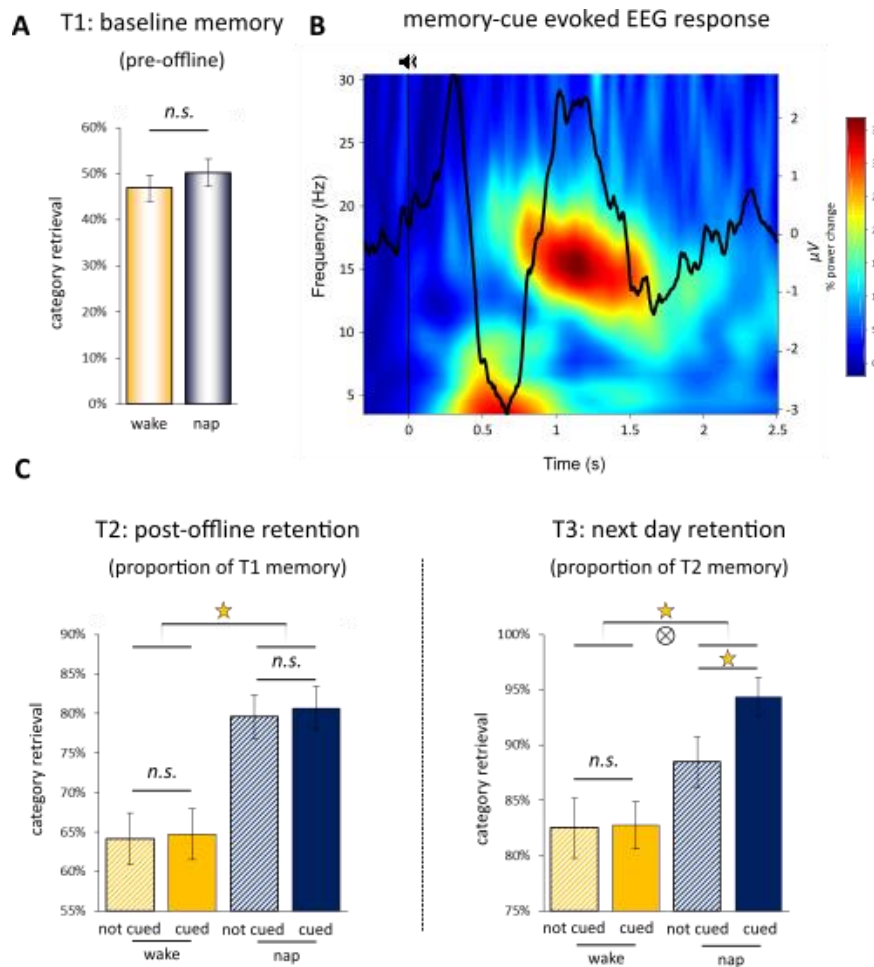


Figure 2: A. Behavioural results at T1 (pre-offline period). Bar graphs show mean (\pm SEM) accuracy for adjective-category retrieval for the NAP group (blue) and the WAKE group (orange). Note that 50% accuracy is not to be mistaken as chance performance given that participants had a “?” response option (see Figure 1A). **B.** Event-related potential (ERP) and time-frequency representation (TFR) evoked by the onset of memory cues. The figure depicts the unthresholded TFR along with the grand average ERP (both collapsed across all channels and then averaged across participants), revealing a strong increase of theta/slow spindle power in the evoked down state followed by an increase in fast spindle power in the ensuing up state. The statistically thresholded TFR map is shown in supplementary Figure S1. **C.** At T2 and T3, behavioural results are further separated into cued trials (solid fill) and not cued trials (hatched fill) and retrieval accuracy is expressed as proportions retained from the most recent memory assessment. Stars denote significant effects, \otimes denotes an interaction effect.

EEG

Sleep EEG data (i.e. time spent in each stage of sleep and spindle/SO event densities) are available in Table 2. Slow and fast spindle topography is illustrated in supplemental Figure S2.

Responses evoked by memory cues

As a first step, we explored the event-related potentials (ERPs) evoked by the auditory memory cues for previously presented (old) adjectives. As shown in Figure 2B and supplemental Figures S1 and S3, cues were followed by a pronounced ERP response resembling a SO/k-complex. Consistent with previous findings¹⁹, time-frequency representation (TFR) results showed that these cue-induced SOs harboured a theta/slow spindle burst in the down state which was followed by a fast spindle burst in the ensuing up state.

A	<i>N1</i>	<i>N2</i>	<i>N3</i>	<i>REM</i>	<i>TST</i>				
	16.56	35.78	14.85	20.81	90.89				
	(± 1.45)	(± 3.09)	(± 2.32)	(± 2.88)	(± 2.47)				
B	<i>Overall</i>	<i>F3</i>	<i>F4</i>	<i>C3</i>	<i>C4</i>	<i>P3</i>	<i>P4</i>	<i>O1</i>	<i>O2</i>
<i>Slow Spindle Density</i>	2.00	2.80	2.74	2.12	2.04	1.66	1.59	1.52	1.50
	(± 0.12)	(± 0.18)	(± 0.17)	(± 0.14)	(± 0.14)	(± 0.09)	(± 0.10)	(± 0.06)	(± 0.06)
<i>Fast Spindle density</i>	3.37	3.10	3.04	3.70	3.45	3.92	3.66	2.99	3.12
	(± 0.13)	(± 0.10)	(± 0.12)	(± 0.09)	(± 0.14)	(± 0.10)	(± 0.17)	(± 0.16)	(± 0.14)
<i>SO Density</i>	2.69	2.63	2.68	2.60	2.64	2.64	2.65	2.82	2.88
	(± 0.09)	(± 0.11)	(± 0.10)	(± 0.09)	(± 0.08)	(± 0.10)	(± 0.08)	(± 0.08)	(± 0.07)

Table 2: Sleep data. **A.** Time (min) spent in each stage of sleep and total sleep time (TST). **B.** Event density measures for slow spindles (10-13 Hz), fast spindles (13-16 Hz) and slow oscillations (SOs, 0.5-2 Hz) at each EEG channel. Density is calculated as the number of electrophysiological events divided by time spent in non-rapid eye movement (NREM) sleep stages N2 and N3. Data are shown as means ± SEM.

To more directly isolate the mechanisms engaged in the processing of mnemonic cues, we next asked whether evoked oscillatory responses might be able to distinguish between old cues and novel control adjectives. As shown in Figure 3A and 3B, the direct contrast revealed that old cues elicited a significant increase in oscillatory power in the fast spindle band (13-16 Hz) from ~1.7-2.3 sec after cue onset ($P < .05$, corrected for multiple comparisons across channels, time and frequency). Topographical representation of the significant spindle increase revealed that the effect stemmed from left hemisphere electrodes, with a maximum at centroparietal sites C3/P3 (Figure 3C).

To fully characterise the observed spindle increase for old cues, we algorithmically detected discrete spindle events in C3 and P3 from 1-2.5 sec after the onset of old cues (see methods). Figure 3D displays the resulting grand average spindle across participants, aligned to the maximum of the

detected spindle events. Not only does this result confirm that the oscillatory increase (Figure 3A and 3B) reflects discrete spindle events, but, as can be appreciated by the above-zero spindle centre, these spindles appeared to be modulated by the up states of concomitant SOs. To statistically quantify this observation, we derived the preferred phase of spindle-SO coupling for the detected events (see methods). Indeed, as shown in Figure 3D, the preferred phase of SO-spindle modulation clustered significantly around the SO up state (0°). This was confirmed when pooling all detected spindles across participants (Rayleigh test: $z=35.8$, $P<1.4e-16$; V test vs. 0: $v=131.6$, $P<1.1e-16$), as well as when deriving the preferred phases for each participant separately and then subjecting the resulting phase values to second-level statistics (mean preferred phase: 2° , Rayleigh test: $z=8.3$, $P<1.3e-4$; V test vs. 0: $v=14.4$, $P<2.3e-05$). In sum, these results show that old memory cues relative to new control adjectives elicit an increase in fast spindle activity which is localised to left hemisphere sites and modulated by the SO up state.

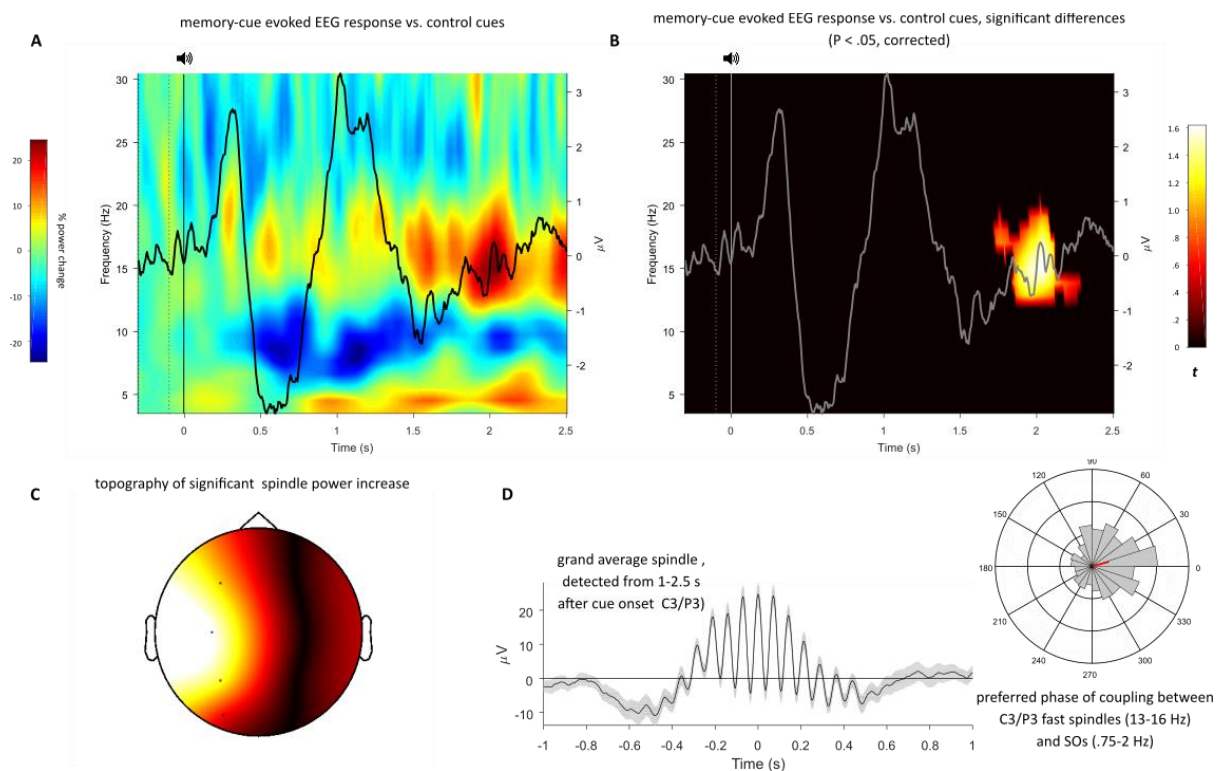


Figure 3: A. TFR difference map of responses elicited by old memory cues vs. new control adjectives, with the corresponding ERP for old cues superimposed. B. same as A, but after statistical thresholding ($P<.05$, corrected). Note the significant increase in fast spindle power (13-16 Hz) from ~ 1.7 -2.3 s post cue onset. C. Topography of the resulting cluster, revealing left-hemisphere specificity of the effect (see also supplemental Figure S4). D. *left*: Grand average (\pm SEM) of discrete spindle events detected from 1-2.5 s after onset of old memory cues at centroparietal sites. *right*: histogram of the corresponding SO-spindle modulation phases, pooled across all spindles at C3 and P3 across all participants ($n=517$, mean direction = 15° , shown in red).

Information processing linked to fast spindles

If the surge of fast spindles was indeed indicative of cue-induced information processing, we reasoned that we should be able to decode from evoked EEG responses the categorical features (object vs. scene, Figure 1) of the images paired with the adjectives during learning. We used a representational similarity analysis (RSA) approach to tackle this question³¹. First, we derived a feature vector of 8 channels x 41 time points (spanning 200 ms at our sampling rate of 200 Hz) centred at each sample from -.2 s to 2.5 s relative to cue onset. Next, using Spearman correlations, we quantified both the within-category similarity (how similar is the EEG pattern of a given object-related adjective to the EEG pattern of all other object-related adjectives and how similar is the pattern of a given scene-related adjective to the pattern of all other scene-related adjectives) and the between-category similarity (how similar is the pattern of a given object-related adjective to the pattern of all scene-related adjectives and vice versa) at each time point. Category-specific information processing would be expressed as an increase in within-category similarity (collapsed across object- and scene-related adjectives) compared to the between-category similarity. As shown in Figure 4A, within-category similarity tended to exceed between-category similarity throughout the post-cue period. Importantly, however, the strongest effect that also reached statistical significance ($P < .05$, corrected for multiple comparisons across time) was observed from 1.76-2.12 s after cue onset, which overlaps precisely with the time window in which we observed the fast spindle increase for old relative to new cues (1.7-2.3 s, Figure 3B). Note also that converging evidence was obtained when using linear discriminant analysis (LDA, using k-fold cross-validation with $k=10$) to classify object- vs. scene-related cues, employing the same features and time windows as for the RSA analysis: the integrated classifier performance from 1.7-2.3 s (the window of fast spindle increase) showed a significant increase from chance performance ($t(26)=2.3$, $P=.03$ (two-tailed)). However, the advantage of our RSA approach is that it provides, via the between-category similarity, a single measure that captures the pattern separability across objects and scenes (i.e. the smaller the between-category similarity, the more distinctive the object vs. scene patterns; see next step).

Lastly, we asked whether cue-specific information processing during the period of fast spindle increases bears any relevance for behavioural manifestations of consolidation. To this end, we assessed the correlation between (i) the relative benefit of cued vs. not cued items on T3 retention in each participant and (ii) the degree to which their evoked patterns of object- vs. scene-related adjectives are separable ($1 - \text{between-category similarity}$, averaged from 1.8-2.1 s after cue onset). As shown in Figure 4B, results revealed a significant positive relationship between the two variables

(Pearson $r=.45$, $P=.02$; Spearman $r=.41$, $P=.03$). Taken together, these findings point to processing of learning-related information during the cue-evoked fast spindle surge, with the distinctiveness or fidelity of that processing predicting the behavioural benefit of TMR.

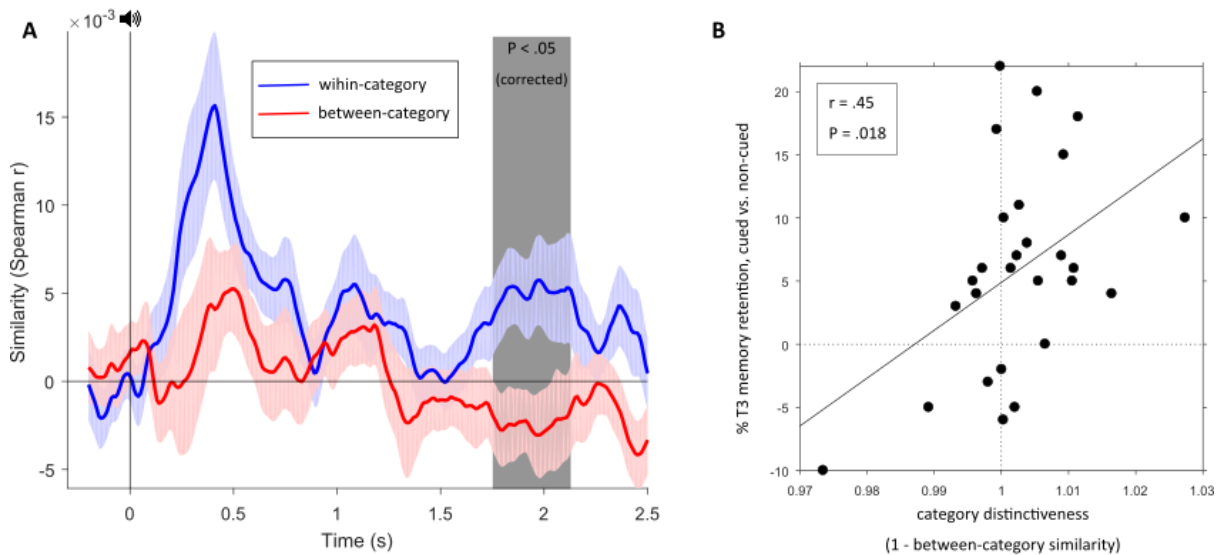


Figure 4: Information processing evoked by memory cues. **A.** Time course (mean \pm SEM) of within- vs. between category similarity in response to old memory cues. Shaded area from 1.8-2.1 s highlights significant increase ($P < .05$, corrected for multiple comparisons across time). **B.** Correlation of category distinctiveness (1-between-category similarity from 1.8-2.1 s) and behavioural benefit of cueing on overnight consolidation across participants.

Discussion

Our study reveals that spindle-mediated information processing is a central mechanism for offline consolidation, and that targeted memory reactivation (TMR) may exploit this mechanism to selectively strengthen cued memory traces. Specifically, we found that memory cues delivered in NREM sleep prompted a transient increase in spindle activity that was coupled with depolarising SO up states (Figure 3). During this surge in spindle activity, the categorical features of cued representations could be reliably decoded, with the extent of such decoding predicting the behavioural benefits of TMR (Figure 4).

The hierarchical nesting of spindles in SO up states is central to the *Active Systems* framework of sleep-dependent declarative memory consolidation¹⁻⁴. Synchronised interactions between these

oscillations are thought to be crucial for effective hippocampal-to-neocortical memory transfer. Interestingly, previous studies have indicated that fast spindles (13-16 Hz) in particular display strong temporal coupling with SO up states^{6, 7, 25}. Corroborating a unique role for these oscillations in declarative memory processing, encoding of hippocampus-dependent memories has been found to enhance fast but not slow spindle activity in NREM sleep³², and fast relative to slow spindles are associated with greater BOLD activation in the hippocampus³³. Here, auditory memory cues (but not novel control adjectives) elicited an increase in such fast spindle events, modulated by SO up states (Figure 3).

What is the functional significance of fast spindles for memory consolidation? Simultaneous EEG-fMRI recordings have shown that reactivation of learning networks is linked to spindle parameters during subsequent NREM sleep³⁰. Moreover, olfactory memory cueing has been shown to evoke sleep spindles in task-relevant brain regions³⁴. Although ceiling memory performance in these two studies precluded a direct link to consolidation, their findings suggest that a TMR-induced increase in fast spindle activity may reflect mnemonic processing in relevant hippocampal-neocortical networks. This view is substantiated by the left hemispheric specificity of the effect observed in our study (Figure 3C), which may reflect the verbal properties of the auditory cues.

More critically, during this transient, cue-induced increase in fast spindle activity, we were able to reliably decode the categorical features (i.e. object vs. scene) of the association linked to the verbal cue (Figure 4A), with the fidelity of this decoding predicting next-day retrieval performance across nap group participants (Figure 4B). One intriguing possibility is that spindles effectively gate activation towards category-specific cortical modules, leading to discriminable distributions of the spatiotemporal EEG patterns. Consistent with this notion, recent work applying EEG classifiers has shown that spectral power in the spindle range contributes to the ability to spontaneously decode previously learned materials in overnight recordings²⁹. This converging evidence provides a good complement to the experimentally-evoked windows of spindle-mediated mnemonic activity observed in the current study. Finally, other work has shown that a short period following TMR-induced spindles is highly susceptible to disruption from additional auditory inputs, capable of abolishing the benefits of TMR for consolidation^{19, 35}. Collectively, these findings point towards a crucial role of spindles in mediating information processing in service of sleep-dependent memory consolidation.

Mechanistically, modelling and empirical data suggest that spindle oscillations induce a massive Ca²⁺ influx into dendrites of pyramidal neurons, opening a molecular 'gate' for synaptic plasticity and, consequently, permanent network changes³⁶⁻³⁸. Finely-tuned windows of spindle activity,

triggered by TMR, may therefore prime or 'tag' relevant synapses for plastic changes during subsequent periods of sleep. This may explain why the behavioural benefits of TMR observed in this study did not emerge until the following day, once the tagged representations had undergone additional overnight processing. Because sleep spindles are central to the proposed mechanism of synaptic priming via TMR, this may further explain why a memory benefit of cueing was not observed in the wake group (Figure 2). It should be noted though that the vast majority of studies employing nap-based TMR have reported improved retention of cued (vs. non-cued) memories immediately after sleep^{15-17, 21}. However, our study is the first to integrate into a TMR paradigm the task of encoding arbitrary adjective-object or adjective-scene pairs. Thus, the spindle-mediated plastic processes induced by cueing in this context may have been insufficient to prompt a behavioural change without further overnight consolidation. Moreover, given recent views on the potential mechanistic overlap between online reactivation and offline consolidation³⁹, another possibility is that T2 retrieval modulated interactions between prior TMR and subsequent overnight memory processing. How the memory effects of cueing are influenced by mnemonic activity in online and offline periods will be a fruitful avenue for future research.

Owing to the limited spatial resolution of scalp EEG monitoring, the putative role of hippocampally-generated ripples (>80 Hz oscillations) in our paradigm remains open. This is an important consideration as neuronal reactivations are mostly observed in conjunction with ripple events⁴⁰⁻⁴², which are temporally nested within the oscillatory troughs of spindles⁴³⁻⁴⁵. The *Active Systems* framework postulates that these synchronised spindle-ripple interactions enable spindle oscillations to shuttle reactivated hippocampal representations to distributed neocortical sites during excitable SO up states¹⁻⁴. Unifying our experimental paradigm with methods for detecting hippocampal ripples in the human brain (e.g. intracranial EEG) would thus provide exceptional insights into mnemonic processing in the sleeping brain. Notably, spindle-ripple interactions may reveal even greater detail on the informational content of decoded associations than spindle oscillations alone.

In sum, our findings suggest that experimental memory cueing generates finely-tuned windows of spindle-mediated information processing, which underpins the selective strengthening of cued representations. These findings not only offer mechanistic insights into the mnemonic impacts of TMR, but provide unique and highly controlled experimental evidence for the critical role of spindles in offline memory consolidation.

Acknowledgements

This research was supported by a Wellcome Trust Institutional Strategic Support Fund (105624) through a University of York Centre for Chronic Diseases and Disorders (C2D2) Fellowship and a Medical Research Council (MRC) Career Development Award (MR/P020208/1) to SAC, and a Wellcome Trust/Royal Society Sir Henry Dale Fellowship to BPS (107672/Z/15/Z). The authors are grateful to Aidan Horner and Gareth Gaskell for fruitful discussions of the data.

References

1. Born, J., Rasch, B. & Gais, S. Sleep to Remember. *The Neuroscientist* **12**, 410-424 (2006).
2. Born, J. & Wilhelm, I. System consolidation of memory during sleep. *Psychological Research* **76**, 192-203 (2012).
3. Diekelmann, S. & Born, J. The memory function of sleep. *Nat Rev Neurosci* **11**, 114-126 (2010).
4. Rasch, B. & Born, J. About Sleep's Role in Memory. *Physiological Reviews* **93**, 681-766 (2013).
5. Mednick, S.C., *et al.* The critical role of sleep spindles in hippocampal-dependent memory: a pharmacology study. *The Journal of Neuroscience* **33**, 4494-4504 (2013).
6. Ngo, H.-Viet V., Martinetz, T., Born, J. & Mölle, M. Auditory Closed-Loop Stimulation of the Sleep Slow Oscillation Enhances Memory. *Neuron* **78**, 545-553 (2013).
7. Ngo, H.-V.V., *et al.* Driving Sleep Slow Oscillations by Auditory Closed-Loop Stimulation—A Self-Limiting Process. *The Journal of Neuroscience* **35**, 6630-6638 (2015).
8. Wilhelm, I., *et al.* Sleep Selectively Enhances Memory Expected to Be of Future Relevance. *The Journal of Neuroscience* **31**, 1563-1569 (2011).
9. Marshall, L., Helgadottir, H., Mollle, M. & Born, J. Boosting slow oscillations during sleep potentiates memory. *Nature* **444**, 610-613 (2006).
10. Cox, R., Hofman, W.F. & Talamini, L.M. Involvement of spindles in memory consolidation is slow wave sleep-specific. *Learning & Memory* **19**, 264-267 (2012).
11. Schabus, M., *et al.* Sleep spindles and their significance for declarative memory consolidation. *Sleep* **27**, 1479-1485 (2004).
12. Oudiette, D. & Paller, K.A. Upgrading the sleeping brain with targeted memory reactivation. *Trends in Cognitive Sciences* **17**, 142-149 (2013).
13. Schouten, D.I., Pereira, S.I.R., Tops, M. & Louzada, F.M. State of the art on targeted memory reactivation: Sleep your way to enhanced cognition. *Sleep Medicine Reviews* **32**, 123-131 (2017).

14. Rasch, B., Büchel, C., Gais, S. & Born, J. Odor Cues During Slow-Wave Sleep Prompt Declarative Memory Consolidation. *Science* **315**, 1426-1429 (2007).
15. Cairney, S.A., Lindsay, S., Sobczak, J.M., Paller, K.A. & Gaskell, M.G. The Benefits of Targeted Memory Reactivation for Consolidation in Sleep are Contingent on Memory Accuracy and Direct Cue-Memory Associations. *Sleep* **39**, 1139-1150 (2016).
16. Oudiette, D., Antony, J.W., Creery, J.D. & Paller, K.A. The Role of Memory Reactivation during Wakefulness and Sleep in Determining Which Memories Endure. *The Journal of Neuroscience* **33**, 6672-6678 (2013).
17. Rudoy, J.D., Voss, J.L., Westerberg, C.E. & Paller, K.A. Strengthening Individual Memories by Reactivating Them During Sleep. *Science* **326**, 1079-1079 (2009).
18. Batterink, L.J., Creery, J.D. & Paller, K.A. Phase of Spontaneous Slow Oscillations during Sleep Influences Memory-Related Processing of Auditory Cues. *The Journal of Neuroscience* **36**, 1401-1409 (2016).
19. Schreiner, T., Lehmann, M. & Rasch, B. Auditory feedback blocks memory benefits of cueing during sleep. *Nature Communications* **6**, 8729 (2015).
20. Schreiner, T. & Rasch, B. Boosting Vocabulary Learning by Verbal Cueing During Sleep. *Cerebral Cortex* **25**, 4169-4179 (2015).
21. Cairney, S.A., Durrant, S.J., Hulleman, J. & Lewis, P.A. Targeted Memory Reactivation During Slow Wave Sleep Facilitates Emotional Memory Consolidation. *Sleep* **37**, 701-707 (2014).
22. Cairney, S.A., Sobczak, J.M., Lindsay, S. & Gaskell, M.G. Mechanisms of Memory Retrieval in Slow-Wave Sleep. *Sleep* **40** (2017).
23. Fuentemilla, L., *et al.* Hippocampus-Dependent Strengthening of Targeted Memories via Reactivation during Sleep in Humans. *Current Biology* **23**, 1769-1775 (2013).
24. van Dongen, E.V., *et al.* Memory stabilization with targeted reactivation during human slow-wave sleep. *Proceedings of the National Academy of Sciences* **109**, 10575-10580 (2012).
25. Andrillon, T., *et al.* Sleep Spindles in Humans: Insights from Intracranial EEG and Unit Recordings. *The Journal of Neuroscience* **31**, 17821-17834 (2011).
26. Nir, Y., *et al.* Regional Slow Waves and Spindles in Human Sleep. *Neuron* **70**, 153-169 (2011).
27. Clemens, Z., Fabó, D. & Halász, P. Overnight verbal memory retention correlates with the number of sleep spindles. *Neuroscience* **132**, 529-535 (2005).
28. Clemens, Z., Fabó, D. & Halász, P. Twenty-four hours retention of visuospatial memory correlates with the number of parietal sleep spindles. *Neuroscience Letters* **403**, 52-56 (2006).
29. Schönauer, M., *et al.* Decoding material-specific memory reprocessing during sleep in humans. **8**, 15404 (2017).

30. Bergmann, T.O., Mölle, M., Diedrichs, J., Born, J. & Siebner, H.R. Sleep spindle-related reactivation of category-specific cortical regions after learning face-scene associations. *NeuroImage* **59**, 2733-2742 (2012).
31. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience* **2** (2008).
32. Mölle, M., Bergmann, T.O., Marshall, L. & Born, J. Fast and Slow Spindles during the Sleep Slow Oscillation: Disparate Coalescence and Engagement in Memory Processing. *Sleep* **34**, 1411-1421 (2011).
33. Schabus, M., *et al.* Hemodynamic cerebral correlates of sleep spindles during human non-rapid eye movement sleep. *Proceedings of the National Academy of Sciences* **104**, 13164-13169 (2007).
34. Cox, R., Hofman, W.F., de Boer, M. & Talamini, L.M. Local sleep spindle modulations in relation to specific memory cues. *NeuroImage* **99**, 103-110 (2014).
35. Farthouat, J., Gilson, M. & Peigneux, P. New evidence for the necessity of a silent plastic period during sleep for a memory benefit of targeted memory reactivation. *Sleep Spindles & Cortical Up States* **1**, 14-26 (2017).
36. Rosanova, M. & Ulrich, D. Pattern-Specific Associative Long-Term Potentiation Induced by a Sleep Spindle-Related Spike Train. *The Journal of Neuroscience* **25**, 9398-9405 (2005).
37. Sejnowski, T.J. & Destexhe, A. Why do we sleep? *Brain Research* **886**, 208-223 (2000).
38. Timofeev, I., *et al.* Short- and medium-term plasticity associated with augmenting responses in cortical slabs and spindles in intact cortex of cats in vivo. *The Journal of Physiology* **542**, 583-598 (2002).
39. Antony, J.W., Ferreira, C.S., Norman, K.A. & Wimber, M. Retrieval as a Fast Route to Memory Consolidation. *Trends in Cognitive Sciences* **21**, 573-576 (2017).
40. Buzsáki, G. Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus* **25**, 1073-1188 (2015).
41. Kudrimoti, H.S., Barnes, C.A. & McNaughton, B.L. Reactivation of Hippocampal Cell Assemblies: Effects of Behavioral State, Experience, and EEG Dynamics. *The Journal of Neuroscience* **19**, 4090-4101 (1999).
42. Nakashiba, T., Buhl, D.L., McHugh, T.J. & Tonegawa, S. Hippocampal CA3 Output Is Crucial for Ripple-Associated Reactivation and Consolidation of Memory. *Neuron* **62**, 781-787 (2009).
43. Clemens, Z., *et al.* Temporal coupling of parahippocampal ripples, sleep spindles and slow oscillations in humans. *Brain* **130**, 2868-2878 (2007).

44. Clemens, Z., *et al.* Fine-tuned coupling between human parahippocampal ripples and sleep spindles. *European Journal of Neuroscience* **33**, 511-520 (2011).
45. Staresina, B.P., *et al.* Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nat Neurosci* **18**, 1679-1686 (2015).

Methods

Participants

A total of 83 participants took part in this study. However, 15 participants were excluded because they did not meet the performance criterion in the pre-sleep test (T1; see below). One further participant withdrew having not understood the necessary time commitments of the study. Of those participants remaining who took part in the nap version of the experiment, a further 21 were excluded for the following reasons: insufficient sleep such that at least one full round of targeted memory reactivation (TMR) could not be attained (9), exhibiting an arousal or awakening during TMR and not returning to non-rapid eye movement sleep stage N2/N3 (10) and computer malfunction (2). The analyses reported in this paper were thus carried out on 46 participants, who were assigned to a nap group (n=27, 20 female, mean \pm SD age, 19.70 ± 1.51 years) or a wake group (n=19 females, mean \pm SD age, 19.26 ± 1.15 years). Pre-study screening questionnaires indicated that participants had no history of sleep, psychiatric or neurological disorders, were not using any psychologically active medications. Participants were informed that they were taking part in a memory study, but were unaware that targeted memory reactivation (TMR) would take place. Written informed consent was obtained from all participants in line with the Research Ethics Committees of the Department of Psychology, University of York and the School of Psychology, University of Birmingham.

Stimuli

Adjectives

250 adjectives were randomly selected from a longer list of 355¹ for each participant. Mean (\pm SD) adjective length was 6.85 ± 1.84 characters and number of syllables ranged from 1-4. All adjectives were recorded in a female voice. Mean (\pm SD) duration of all included adjectives was 704 ± 146 ms.

Images: Objects and Scenes

100 images (50 objects and 50 scenes) were randomly selected from a set of 200^{2, 3} for each participant. Objects were images of everyday items, animals or food presented on a plain white background (e.g. an apple). Scenes were images of landscapes or places (e.g. a bowling alley). Care was taken to ensure that scenes contained sufficient background detail to be easily distinguishable from objects. The distinction between objects and scenes was clearly explained to participants.

Procedure

Participants completed a short practice version (10 trials) of each experimental task to ensure that they fully understood the instructions. All responses were made via keyboard press on a PC. Experimental stimuli were always presented in random order.

Familiarisation

A familiarisation phase at the beginning of the experiment was designed to facilitate learning of the adjective-image pairs in the main encoding session. First, participants completed an adjective familiarisation task. On each trial, one of 100 adjectives (e.g. “exotic”) was presented aurally and displayed for 2.5 s on the computer screen. Participants indicated whether they considered the adjective to be emotionally positive or negative. Each trial was separated by an inter-stimulus interval (ISI) with a fixation cross for 1.5 s (± 100 ms random jitter). Next, participants completed an object/scene categorisation task. On each trial, one of 50 objects (e.g. butterfly) or 50 scenes (e.g. golf course) was displayed for 2.5 s. Participants indicated whether they considered the image to be an object or a scene (ISI=1.5s ± 100 ms).

Encoding

Participants learned randomised pair-wise associations between the adjectives and images presented in the familiarisation phase (100 adjective-image pairs). On each trial, a randomly selected adjective (e.g. “exotic”) was presented aurally and displayed above an object or scene (e.g. object: butterfly) for 4.5 s. To facilitate learning, participants were instructed to form a vivid mental image or story that closely linked the adjective and the object/scene, and then indicate whether the image they had formed was realistic or bizarre (ISI=1.5s ± 100 ms). For example, the mental image corresponding to the adjective “exotic” and the object butterfly would presumably be realistic as butterflies can be exotic creatures. Participants were informed that their memory for adjective-image pairs would be tested immediately afterwards.

Immediate Test (T1)

T1 included the 100 adjectives from encoding intermixed with 50 new adjectives that participants had not seen before (foils). On each trial of the test phase, an adjective (e.g. “exotic”) was presented aurally and displayed for 2 s. Afterwards, participants were asked to indicate whether the adjective was ‘old’ (i.e. they recognised it from the encoding phase) or ‘new’ (i.e. it was not seen at encoding) within 10 s. When participants provided a “new” response, they immediately moved on to the next trial (ISI=1.5 s ± 100 ms). When an “old” response was provided, participants were required to

indicate whether the associated image was an object, a scene, or unknown to them. In order to ensure that object or scene responses were not mere guesses, participants also provided a typed description of the image they had remembered. Across all T1 trials where the category was correctly recalled, participants were able to correctly describe the image on the majority of occasions (mean \pm SD: $80.95 \pm 14.59\%$), demonstrating that the category responses reflected veridical memory.

TMR Set Up

Of the adjective-image pairs that were correctly recalled at T1 (i.e. when the adjective was correctly recognised and the associated image category correctly recalled), half were randomly allocated to the cued condition (i.e. for TMR), whereas the other half were assigned to the non-cued condition. This ensured that baseline category recall performance was balanced between the cued and non-cued memories. For example, if a participant correctly recalled 40 pairs at T1, then 20 of these would be assigned to the cued condition and the other 20 assigned to the non-cued condition. On occasions where there were an odd number of correctly recalled pairs, the spare item was randomly allocated to the cued or non-cued condition. To ensure that a sufficient number of stimuli were available for TMR in sleep, participants were required to correctly recall at least 14 objects and 14 scenes at T1. Participants that did not meet this criterion were excluded ($n=15$). The adjectives from pairs assigned to the cued condition were replayed during the TMR phase.

Importantly, an additional set of control adjectives that participants had not encountered at either encoding or T1 were randomly intermixed with the TMR stimuli. The number of control adjectives was equal to half the number of stimuli in the cued condition. For example, if there were 40 adjectives associated with correctly recalled categories in the cued condition, then a further 20 control adjectives would be added to the TMR set (total=60). Inclusion of the control adjectives enabled a direct comparison of brain activity during cued memory reactivation and non-specific, matched auditory stimulation.

Offline Period (Nap or Wakefulness)

The offline period began at ~2pm and lasted 90 min. Participants in the nap group were left to sleep in a laboratory bedroom while their brain activity was monitored with polysomnography (set up before the study began). TMR was initiated when participants were in late NREM stage N2/early stage N3. The TMR set was presented in a randomised order ($ISI=4s \pm 200$ ms) at a sound intensity of ~40dB. After each full round of cueing, the adjectives were reshuffled and immediately presented again. Cueing continued for as long as participants were in sleep stage N2/N3, but immediately paused if they showed signs of micro-arousal or awakening, or moved into sleep stage N1 or rapid

eye movement (REM) sleep. The cues were continued if participants re-entered sleep stage N2/N3 after an arousal.

Participants in the wake group played an online game (Bubble Shooter) for the first 30 min of the offline period. For the next 30 min, the TMR cues were presented continuously while participants completed a 1-back working memory task. This approach reduced the probability that participants directly attended to the cues during TMR^{4, 5}. During the 1-back task, a series of random numbers between 0 and 10 were presented one after another in the centre of the screen. The task was to indicate whether the current number was the same as or different to the number one digit prior. After completing the 1-back task, participants played Bubble Shooter again for the remaining 30 min of the offline period. The mean (\pm SD) number of presentations per cue was 9.16 ± 3.67 in the wake group and 5.96 ± 5.85 in the sleep group.

Follow-Up Tests (T2 and T3)

Participants returned 6 hours later for a follow-up test (T2). This followed the same procedures as T1 with the single exception that new foil adjectives were used. The next morning (after a night of sleep), participants completed another test (T3). Again, this followed the same procedures as T1 and T2, but with a new set of foils.

Discrimination Task

After completing T3, participants were informed of the true purpose of the study and asked if they had been aware of the auditory cues in the offline period. To assess their explicit knowledge of the cues, participants were asked to complete an adjective discrimination task. On each trial, one of 100 adjectives from the encoding phase was presented aurally and displayed for 10 s. Participants were asked to indicate whether or not the adjective had been replayed during the offline period. Discrimination task data are available in the supplementary materials.

Equipment

Experimental Tasks and TMR

All experimental tasks and TMR algorithms were implemented on a PC with MATLAB 2015a and Psychtoolbox version 3.0.13. In the wake group, adjective cues were presented via speakers connected to the task PC. In the nap group, cues were presented via a speaker mounted ~ 1.5 m above the bed, which was connected to an amplifier in a separate control room.

Polysomnography

An Embla N7000 PSG system with RemLogic version 3.4 software was used to monitor sleep. After the scalp was cleaned with NuPrep exfoliating agent (Weave and Company), gold plated electrodes were attached using EC2 electrode cream (Grass Technologies). EEG scalp electrodes were attached according to the international 10-20 system at 8 locations: frontal (F3, F4), central (C3, C4), parietal (P3, P4) and occipital (O1, O2), and each was referenced to an electrode on the contralateral mastoid (A1 or A2). Left and right electrooculography electrodes were attached, as were electromyography electrodes at the mentalis and submentalis bilaterally, and a ground electrode was attached to the forehead. Each electrode had a connection impedance of < 5 k Ω and all signals were digitally sampled at 200 Hz. Sleep scoring was carried out in accordance with the criteria of the American Academy of Sleep Medicine⁶.

Data Analysis

Behaviour

Category recall was our primary measure of memory accuracy. We calculated for each participant: 1) the proportion of categories recalled at T1 that were subsequently recalled at T2, and 2) the proportion of categories recalled at T2 that were subsequently recalled at T3 (i.e. following a night of sleep). To avoid any ambiguity related to category memory, we excluded from our analyses any item that was incorrectly classified during the object/scene categorisation task. Across all participants, we excluded 162 items out of a possible 4600 (3.52%). Category recall scores were subjected to a 2 (TMR: Cued/Not-Cued) X 2 (Group: Nap/Wake) mixed analysis of variance (ANOVA).

EEG

EEG data were analysed with MATLAB (MathWorks), using the FieldTrip⁷ and CircStat⁸ toolboxes. The continuous sleep data were segmented into epochs from -1 s to 3 s around cue onset and subjected to a two-step artifact rejection procedure. In the first step, artifacts were automatically detected and removed based on the median plus/minus 3.5 inter-quartile ranges of both signal amplitude and gradients (the difference between two adjacent samples) of all epochs. In the second step, the remaining epochs were manually screened via FieldTrip's visual summary functions and epochs containing amplitude, variance or kurtosis outliers were additionally removed. For TMR-locked analysis of event-related potentials (ERPs), data were high-pass filtered at 0.5 Hz and baseline-corrected with respect to the -200 ms to 0 ms window before cue onset. For time-frequency representations (TFRs), data were convolved with a 5-cycles hanning taper and spectral power was obtained from 4-30 Hz in 0.5 Hz frequency steps and 5 ms time steps. For analyses,

participant-specific TFRs were converted into percent power change relative to a -300 ms to -100 ms pre-cue window.

For representational similarity analysis (RSA) of within- vs. between-category processing, a sliding window of 200 ms (in steps of 10 ms) was used to obtain, for each trial, a series of 8-channel-by-41-timepoints (200 Hz/5 ms sampling rate) EEG feature vectors. Using these feature vectors, Spearman correlations were then used to quantify, for each time point, the representational similarity across all pairwise combinations of trials. Next, we collapsed similarity values to capture (i) the average similarity of an object-related cue with all other object-related cues (i.e., within-object similarity), (ii) the average similarity of a scene-related cue with all other scene-related cues (i.e., within-scene similarity) and (iii) the average similarity of all object-related cues with all scene-related cues (i.e., between-object-scene similarity). Category-specific information processing was quantified as the difference between within-category similarity ((i) and (ii) combined) and between-category similarity (iii). Finally, the measure “1 - between-category similarity” was used to index the distinctiveness of category-specific information processing per participant (the greater the between-category similarity, the smaller the discriminability of object-vs. scene-related information).

All ERP, TFR and RSA analyses were corrected for multiple comparisons using FieldTrip’s cluster-based permutation method (1000 randomisations), including channel x time (ERP), channel x time x frequency (TFR) and time (RSA) as cluster-defining features.

Event Detection

Sleep data were partitioned according to the time (minutes) spent in each stage of sleep (N1, N2, N3 and REM sleep). Data scored as N2 or N3 were extracted from all EEG channels for spindle and slow oscillation analysis. For spindles, data were first bandpass filtered from 10-13 Hz (slow spindles) or 13-16 Hz (fast spindles) using a 4th order two-pass Butterworth filter. Next, we took the envelope of the resulting signal and determined an amplitude threshold as mean + 1.25 SD. A spindle was then defined as an event that surpassed that threshold for a minimum of 0.5 s and a maximum of 3 s. For SO detection, data were filtered from 0.8-2 Hz using a 4th order two-pass Butterworth filter. Next, zeros crossings were detected in the resulting signal and events with two successive positive-to-negative crossings spanning 0.8-2 s were taken forward to the next step. Here, the resulting candidate events’ trough and trough-to-peak amplitude were calculated and events surpassing mean + 1.25 SD of both these metrics were considered SOs. In both event detection procedures, automatically detected artifact samples (see above) were padded for ± 1 s and those samples were excluded prior to event detection.

For determining the preferred phase of SO-spindle modulation, we first identified spindles whose maximum occurred from 1-2.5 sec after onset of old memory cues (encompassing the interval in which we observed the spindle increase, Figure 3B). We then extracted a ± 1.5 s raw data segment around the spindle maximum (accommodating the maximum spindle duration of 3 s) and created one signal by filtering the data between 0.5 and 2 Hz and another signal by filtering the data between 13 and 16 Hz. For the lower frequency signal, instantaneous phase was extracted via the Hilbert transform. For the higher frequency signal, phase of the power envelope was extracted, again using the Hilbert transform. For each sample (601 samples, i.e. 3 s at 200 Hz sampling rate), the circular distance between the two phase time series was calculated and the mean resulting angle determined. For instance, if the spindle amplitude were to systematically peak at the SO down state (trough), the mean angle of the phase differences would be 180° . Conversely, if the spindle amplitude was to – as hypothesised – systematically peak at the SO up state (peak), the mean angle of the phase differences would be 0° (see also^{9,10}).

References

1. Staresina, B.P., Duncan, K.D. & Davachi, L. Perirhinal and Parahippocampal Cortices Differentially Contribute to Later Recollection of Object- and Scene-Related Event Details. *The Journal of Neuroscience* 31, 8739-8747 (2011).
2. Konkle, T., Brady, T.F., Alvarez, G.A. & Oliva, A. Scene Memory Is More Detailed Than You Think. *Psychological Science* 21, 1551-1556 (2010).
3. Staresina, B.P. & Davachi, L. Mind the Gap: Binding Experiences across Space and Time in the Human Hippocampus. *Neuron* 63, 267-276 (2009).
4. Rudoy, J.D., Voss, J.L., Westerberg, C.E. & Paller, K.A. Strengthening Individual Memories by Reactivating Them During Sleep. *Science* 326, 1079-1079 (2009).
5. Schreiner, T. & Rasch, B. Boosting Vocabulary Learning by Verbal Cueing During Sleep. *Cerebral Cortex* 25, 4169-4179 (2015).
6. Iber, C., Ancoli-Israel, S., Chesson, A. & Quan, S.F. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification*. (Westchester (IL): American Academy of Sleep Medicine, 2007).
7. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience* 2011, 156869 (2011).
8. Berens, P. CircStat: A MATLAB Toolbox for Circular Statistics. 2009 31, 21 (2009).

9. Cohen, M.X. Assessing transient cross-frequency coupling in EEG data. *Journal of Neuroscience Methods* 168, 494-499 (2008).
10. Staresina, B.P., et al. Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nat Neurosci* 18, 1679-1686 (2015).

Supplementary Materials

Adjective Recognition

In each test, participants made recognition judgements (old or new) for adjectives before recalling the associated image category. The sensitivity index (d') [Normalized (hits/(hits + misses)) – Normalized (false alarms/(false alarms + correct rejections))] was calculated for adjective recognition memory. We adopted a log-linear approach to safeguard this analysis against errors arising from 0 and 1 values: for each participant, 0.5 was added to the total hits and total false alarms, and 1 was added to the total signal (old) trials and total noise (new) trials¹. Adjective recognition scores (d') were applied to a 3 (Test: T1/T2/T3) X 2 (Group: Nap/Wake) mixed ANOVA. Recognition performance initially declined between T1 and T2, but then stabilised between T2 and T3 (Test main effect [Huynh-Feldt corrected]: $F(1,78,88)=9.00$, $P=.001$). However, these changes in performance were not modulated by the type of offline activity (Group*Session interaction [Huynh-Feldt corrected]: $F(1,78,88)=2.36$, $P=.11$). There was also no main effect of Group ($F(1,44)=0.53$, $P=.47$), indicating that overall adjective recognition performance was equivalent in the nap and wake groups. Collectively, these findings imply that the memory effects of sleep and TMR observed in this study were specific to category recall. Adjective recognition data is available in Table S1.

Category Recall

Cue Frequency

In the nap group, category recall performance was superior for cued relative to non-cued items at T3. However, this behavioural benefit of TMR indexed as [cued category recall – non-cued category recall], was not predicted by the frequency of individual cue presentations in sleep (across all participants and cues, mean \pm SD = 5.96 ± 5.85 , $r=.19$, $P=.34$).

Objects vs. Scenes

We repeated the analyses reported in the main text to examine objects and scenes separately. A 2 (Type: Object/Scene) X 2 (Group: Nap/Wake) mixed ANOVA showed that object and scene recall did not differ at T1 (Type main effect: $F(1,44)=0.23$, $P=.64$), for either the sleep group or the wake group (Type*Group interaction: $F(1,44)=1.54$, $P=.22$). Again, there was no main effect of Group ($F(1,44)=0.59$, $P=.45$).

The factor 'Type' was added to our analysis of category retention at T3. A main effect of Type ($F(1,44)=6.88$, $P=.01$) indicated that objects were generally better recalled than scenes. However, there was no interaction between Type and any other factor(s) (all $P>.05$), suggesting that the

memory effects of sleep and TMR observed in this study did not vary according to category membership.

There was also a main effect of Type when this factor was included in our analyses of category retention at T2 ($F(1,44)=4.11$, $P=.05$). Again, there was no interaction between Type and any other factor(s) (all $P>.05$). Category recall data separated for objects and scenes is available in Table S2.

		<i>T1</i>				
		<i>Hit</i>	<i>Miss</i>	<i>CR</i>	<i>FA</i>	<i>d'</i>
<i>Nap</i>		84.04 (± 2.10)	15.96 (± 2.10)	96.37 (± 0.69)	3.63 (± 0.69)	2.89 (± 0.10)
<i>Wake</i>		89.74 ± (1.71)	10.26 ± (1.71)	96.42 (± 0.77)	3.58 (± 0.77)	3.15 (± 0.14)
		<i>T2</i>				
		<i>Hit</i>	<i>Miss</i>	<i>CR</i>	<i>FA</i>	<i>d'</i>
<i>Nap</i>		77.70 (± 2.54)	22.30 (± 2.54)	96.67 (± 0.72)	3.33 (± 0.72)	2.67 (± 0.11)
<i>Wake</i>		79.32 (± 3.27)	20.68 (± 3.27)	94.95 (± 1.98)	5.05 (± 1.98)	2.78 (± 0.16)
		<i>T3</i>				
		<i>Hit</i>	<i>Miss</i>	<i>CR</i>	<i>FA</i>	<i>d'</i>
<i>Nap</i>		79.74 (± 2.86)	20.26 (± 2.86)	96.52 (± 0.87)	3.48 (± 0.87)	2.83 (± 0.10)
<i>Wake</i>		80.11 (± 3.38)	19.89 (± 3.38)	94.95 (± 2.46)	5.05 (± 2.46)	2.78 (± 0.11)

Table S1: Adjective recognition performance (%) at each test phase. The data refer to hits, misses, correct rejections (CR) and false alarms (FA). The sensitivity index (d') was calculated for each test. Data are shown as means ± SEM.

Discrimination Task

At the end of the experiment, participants were re-presented with all of the adjectives from the encoding phase and, for each, asked to indicate whether or not it was replayed in the offline period. The discrimination task analysis was restricted to items that were correctly recalled at T1 (and thus assigned to the cued and non-cued conditions). Cued stimuli that were and were not correctly identified as such were marked as hits and misses, respectively. Non-cued stimuli that were and were not correctly identified as such were marked as correct rejections or false alarms, respectively. A discrimination index was then calculated for each participant by subtracting the proportion of non-

cued trails marked as false alarms from the proportion of cued trials marked as hits. The discrimination index was not significantly different from zero in either the nap group ($t(26)=0.07$, $P=.95$) or the wake group ($t(18)=1.67$, $P=.11$), implying that any memory effects of TMR were not due to conscious processing of the cues. It should be noted that while some of the wake group participants claimed to have heard the cues in the offline period, none of the sleep group participants professed any knowledge of adjective replay.

Objects	<i>T1</i>	<i>T2*</i>		<i>T3**</i>	
		<i>Not Cued</i>	<i>Cued</i>	<i>Not Cued</i>	<i>Cued</i>
<i>Nap</i>	48.96 (± 2.92)	80.86 (± 2.36)	80.20 (± 3.16)	90.09 (± 2.03)	96.14 (± 1.76)
<i>Wake</i>	47.37 (± 3.51)	67.47 (± 4.51)	68.01 (± 4.32)	84.54 (± 4.02)	88.72 (± 3.55)
Scenes	<i>T1</i>	<i>T2*</i>		<i>T3**</i>	
		<i>Not Cued</i>	<i>Cued</i>	<i>Not Cued</i>	<i>Cued</i>
<i>Nap</i>	51.33 (± 3.18)	78.14 (± 2.70)	80.78 (± 2.69)	87.77 (± 1.85)	92.27 (± 1.50)
<i>Wake</i>	46.32 (± 2.77)	61.40 (± 4.57)	61.61 (± 4.53)	79.14 (± 4.90)	76.14 (± 5.62)

Table S2: Category recall performance (%) at each test phase for objects and scenes. Data are shown as means ± SEM. *Refers to the proportion of T1-recalled categories that were also recalled at T2. **Refers to the proportion of T2-recalled categories that were also recalled at T3.

Supplementary Figures

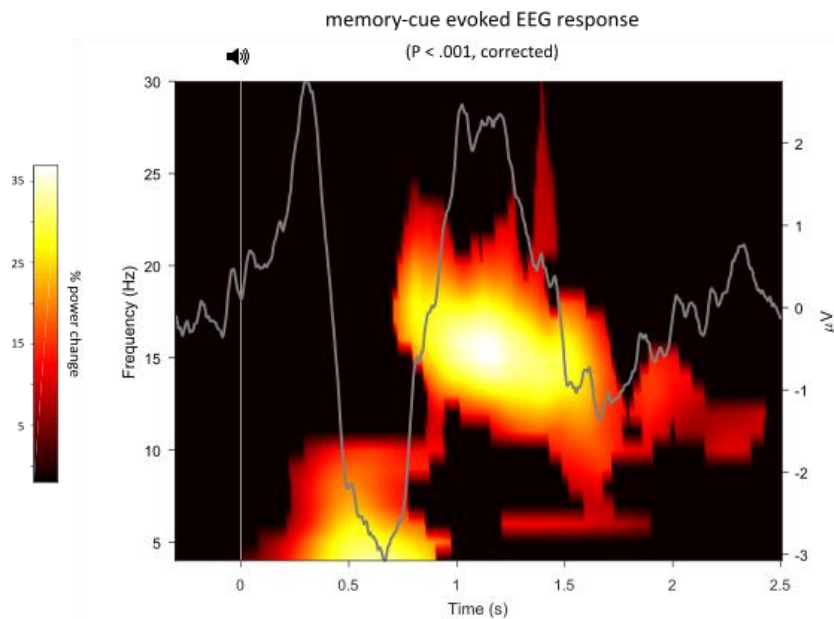


Figure S1: Significant increases in cue-evoked oscillatory power relative to a pre-cue baseline (-.3 to -.1 s). Results are thresholded at $P < .001$ (both for the initial, cluster-defining t test and for the final permutation test). Grey line reflects the corresponding ERP for old cues (cf. Figure 2B).

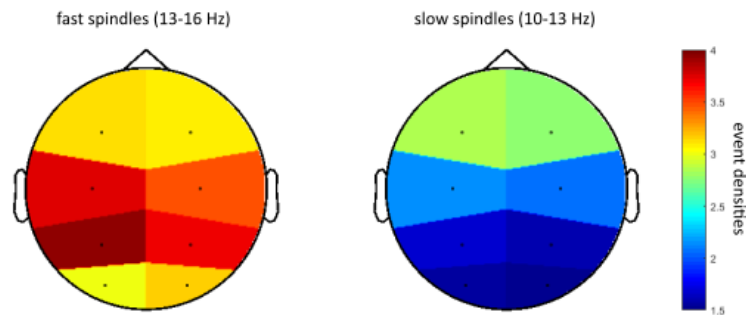


Figure S2: Event densities of fast (left; 13-16 Hz) and slow (right; 10-13 Hz) spindles detected algorithmically during NREM sleep (stages N2 and N3). Note the relative prevalence of fast spindles over centroparietal sites and that of slow spindles over frontocentral sites.

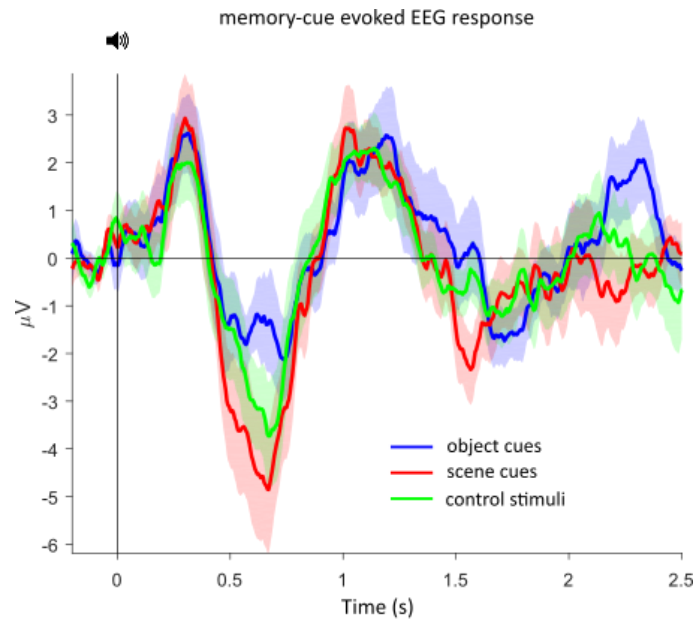


Figure S3: Event related potentials (ERPs) in response to adjective stimuli. Traces show mean \pm SEM across participants, collapsed across all electrodes. A repeated-measures ANOVA revealed no significant condition effects comparing object cues, scene cues and control adjectives after controlling for multiple comparisons across time and electrodes.

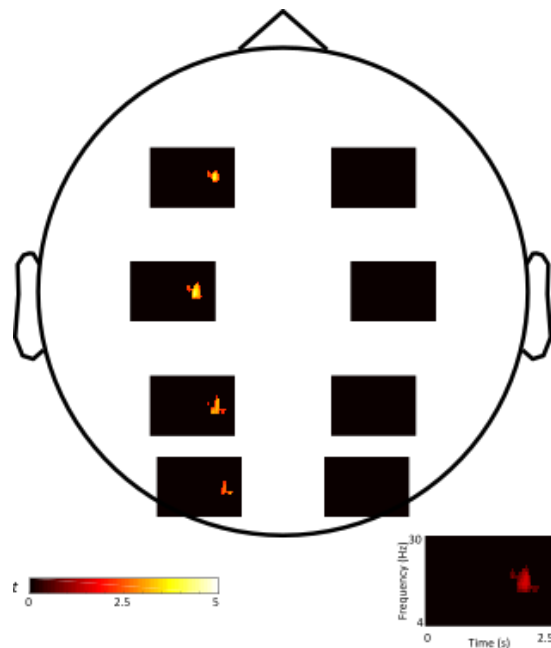


Figure S4: Topographical representation of significant clusters resulting from the contrast old TMR cues vs. novel control adjectives. Note the specificity of the effect in left hemisphere sites, with the maxima at C3 and P3. Right bottom: Effect integrated across all significant sites (cf. Figure 3B).

References

1. Hautus, M.J. Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instruments & Computers* 27, 46-51 (1995).