nature human behaviour

Supplementary information

https://doi.org/10.1038/s41562-025-02117-5

Sleep selectively and durably enhances memory for the sequence of real-world experiences

In the format provided by the authors and unedited

Table of Contents

Supplementary Methods

Participants

Study 1

Fifty-seven healthy adults (14 males, 43 females; 18-35 years old, M_{Age} = 22.74 years ± SD_{Age} = 3.59 years) were recruited from the Baycrest participant registry and the community. All participants were naïve to the second floor of Baycrest, where the tour took place. Four participants were excluded: one for failing to complete one of the remote memory tests within the required time interval, one for technical issues in the headphones used for the tour audio guide, and two for test performance below 2.5 standard deviations from the sample mean on any of the four memory tests (averaged across featural and sequence items), which was suggestive of inattentive responding. After exclusions, data from 53 younger adults were included in analyses (M_{Age} = 22.60 years ± SD_{Age} = 3.23; $M_{Education}$ = 15.40 ± $SD_{Education}$ = 1.67). Participants in all studies were compensated for their time according to institutional guidelines.

Study 2a

Ninety healthy adults (34 males, 56 females; 18 – 45 years old) were randomly assigned to either a Sleep group (n = 44, M_{Age} ± SD_{Age} = 24.09 ± 5.21 years; $M_{\text{Education}}$ ± $SD_{\text{Education}}$ = 15.66 ± 2.82 years) or a Wake group (n = 46, M_{Age} ± SD_{Age} = 24.57 ± 6.20 years; $M_{\text{Education}}$ ± $SD_{\text{Education}}$ = 15.43 ± 2.52 years). Participants underwent the same audio-guided walking tour as in Study 1 (a complete list of all audio tracks can be found on our OSF page: https://osf.io/bxm5w/). Five participants in the Wake group were excluded: three for failing to complete the second follow-up memory test before they went to sleep, one for later disclosing that they had been exposed to the second floor of Baycrest prior to the experiment, and one for taking a nap during the retention interval. One participant in the Sleep group who slept for only two hours was excluded. One participant in the Sleep group withdrew from the study. Four participants (three from the T1-T2 analysis, one from the T1-T4 analysis) were excluded for test performance below 2.5 standard deviations from the sample mean on any of the four memory tests (averaged across featural and sequence items). Finally, two participants failed to complete testing at T4 (these participants were retained for analyses involving T1-T2). As a result, 80 participants (n = 39 Wake, n = 41 Sleep) were included in the total sample, but only 77 (n = 38 Wake, n = 39 Sleep) were included in the analysis exploring memory change across T1-T4.

Participants in the Sleep and Wake groups were matched in terms of chronotype and self-reported sleep habits. Chronotype was assessed with the *Morningness-Eveningness Questionnaire Self-Assessment* (1), a 19-item survey that measures individual differences in human circadian rhythms. Participants' ratings on this instrument did not differ across the Wake and Sleep groups (M's = 45.66 and 44.97, SD's = 6.89 and 8.17 for the Wake and Sleep groups, respectively, $t(75) = 0.40$, p =

.693). There were no significant group differences in self-reported sleep habits such as sleep quality, sleep disturbances, and use of sleep medication as rated on the *Pittsburgh Sleep Quality Index (*PSQI)(2)*,* a nine-item instrument that assesses average sleep quality, patterns, and efficiency over the past month (M's = 9.95 and 9.28, SDs = 6.89 and 4.91 for the Wake and Sleep groups, respectively, $t(75) = 0.49$, $p = 0.627$). A global score of five or more indicates poor sleep quality. The higher the score, the worse the self-reported sleep quality (3).

Study 2a participants who completed the T5 assessment were representative of the full Study 2a sample

The ancillary T5 assessment to assess remote memory for the Baycrest tour was completed by 49 of the original 77 Study 2a participants (26 Sleep and 23 Wake). We compared those who completed this assessment to the original Study 2a sample on demographics and memory performance. Those who completed the T5 assessment did not statistically differ in age ($t(65) = 0.41$, $p = .685$) or years of education ($t(61) = 0.30$, $p = .768$) from the 28 Study 2a participants who did not participate in this long-term follow-up test. Considering memory performance, we compared participants who completed T5 to those who did not (i.e. Attrition Group) in an ANOVA that included effects of Group (sleep vs. wake), Time (T1-T4), and Memory Type (featural vs. sequence). There was no significant effect of Attrition Group ($\chi^2(1) = 0.03$, $p = .862$), nor were there any significant interactions between Attrition Group and the other factors ($\chi^2(1)$'s > 0.72, p's > .093). Finally, the critical Time (T1-T4) X Retrieval Type (featural vs. sequence) interaction observed in Studies 1 and 2a held for the Sleep group who completed the T5 test, even though both sequence and featural memory improved from T1 to T2 (sequence: $z = 3.03$, $p = .002$, $d = 0.986$; featural: $z = 2.08$, $p = .038$, $d = 0.545$). In sum, participants who completed the T5 test were representative of the full sample in terms of demographics, memory performance, and the Sleep group's relative preservation of sequence memory over time.

Study 2b

Eighteen additional participants were recruited to increase the sample size of the sleep group for testing individual differences, prior to halting of the study due to the onset of the COVID-19 pandemic. Although these participants underwent the same procedures as the Study 2a Sleep group, they were not included in behavioral analyses of T1-T4 as they were not randomly assigned to the sleep vs. wake groups. Of the 62 participants who completed PSG, 13 were excluded due to study withdrawal (1), technical issues of sweat sway affecting PSG recording (4), sleep efficiency scores below 60% (6), suspected sleep apnea based on the technologist's report (1), and overall memory performance more than 2.5 standard deviations below the group average at baseline (1). These exclusion criteria were determined prior to data analysis. The final sample of 49 participants included 29 females and 20 males (M_{Age} + SD_{Age} = 24.29 ± 5.03 years); $M_{Education}$ + $SD_{Education}$ = 15.47 ± 2.94 years).

Polysomnography (PSG) Methods

The PSG montage included electrooculography (EOG), EEG, electromyography (EMG), and electrocardiographic (ECG) activity. Electrode placements followed the international 10-20 system and were referenced to the contralateral mastoid (A1 and A2). Vertical and horizontal eye movements were recorded via EOG electrodes that were placed on the left outer canthus (LOC) and right outer canthus (ROC). EEG scalp electrodes were placed at six different locations: frontal (F3, F4), central (C3, C4), and occipital (O1, O2) since these areas are ideal for measuring cortical neural activity such as spindles and slow waves (4). Two EMG electrodes were placed on the chin and the anterior tibial muscles to assist with AASM scoring. This allowed us to extract macrostructure measures of interest including sleep duration, efficiency, latency, and wake time after sleep onset. Data from each EEG channel was collected with a 256-Hz sampling rate, high- and low-pass filtered at 0.3 and 35 Hz, respectively, with a 60-Hz notch filter before being exported from the Compumedics system into European Data Format (EDF) files.

Spindle detection

Spindle detection was carried out using a previously published method that was validated against both multiple expert scorers and crowd-sourced from a large sample of non-experts (5). Prior to automatic detection, movement and arousal artifact appearing in N2 and N3 periods were manually rejected by three expert raters from all participants. The spindle data were then extracted from movement artifact-free, N2 and N3 sleep epochs using the EEGlab plugin (6) "*detect_spindles2.2*" written for MATLAB R2019a (The MathWorks Inc., Natick, MA, United States). This detection method (5) uses a complex demodulation transformation of the EEG signal with a bandwidth of 5 Hz centred about a carrier frequency of 13.5 Hz (i.e., 11–16 Hz). This allowed us to extract the number, duration (s), amplitude (μV), and frequency (Hz) of spindles at each central derivation (C3 and C4). This spindle detection method was performed at C3 and C4 channels for slow (11–13.5 Hz), fast (13.5–16 Hz), and total bandwidth spindles (11–16 Hz). Total spindle counts (averaged across both central channels) were used for our main analyses of spindles. Due to an equipment issue, data for two participants was referenced to the same electrode (e.g., C3-A2 and C4-A2). In this event, we only selected data from the single contralaterally referenced electrode (e.g., C3-A2).

Half-wave detection

The period amplitude analysis (PAA) method is adapted from well-established published methods (7–9). The detection method requires narrow band-pass filtering to detect waves between 0.5 - 4 Hz, thus, derivations from F3 and F4 were band-pass filtered 0.46 Hz (64th-order Chebyshev type II high-pass filter, -80 dB stopband attenuation) and 4.30 Hz (32nd-order Chebyshev type II low-pass filter, -80 dB stopband attenuation) to achieve minimal attenuation in our 0.5 - 4 Hz band of interest and good attenuation at neighbouring frequencies. The filters were applied in the forward and reverse directions to achieve zero-phase distortion resulting in a doubling of the filter

order. Half-waves were determined as negative or positive deflections between two consecutive zero crossings in the band-pass filtered signal. In line with scoring rules for slow waves, the peak-to-peak amplitude threshold of 75 μV (10) was applied to each half-wave and its neighbouring half-wave (i.e., half-wave pairs with < 75 μV peak-to-peak amplitude difference were excluded from further analysis). The duration of each half-wave that met the amplitude threshold was determined by the length of time (in seconds) between zero crossings. Here, we report the number of half-waves averaged over both channels as our proxy measure of SWA. Data from the two participants with referencing issues was treated in the same manner as detailed above.

Supplemental Statistical Analyses

Using the *performance* package in R, we confirmed that all of the effects in the linear mixed effect models were below a variance inflation factor of 5, indicating no issues with multicollinearity (11) (see tables below). For *a priori t*-tests, we used Welch's *t*-tests and reported Cohen's *d* effect sizes. Steiger's Z tests were also conducted to test statistical differences between correlations. Note that although we conducted our statistical tests on model fits, we visualized descriptive statistics (i.e., "raw" means from individual participant data) for ease of interpretability.

Supplementary Results

Time elapsed between encoding and follow-up tests

The follow-up tests (T2-T5) conducted outside of the laboratory were completed online, unsupervised. We confirmed that participants were compliant with the instructions for completing the follow-up tests within the requested time interval. In Study 2a, there were no group (Wake vs. Sleep) differences in the time elapsed between tour encoding and the 12-hour (T1-T2: $t(75) = 1.39$, $p = .170$, $d = 0.316$), 1-week (T1-T3: $t(75) = 1.03$, $p = .307$, $d = 0.234$), 1-month ($t(75) = -0.28$, $p = .782$, $d = -0.063$), or 1-year ($t(49) = 1.04$, $p = .302$, $d = 0.294$) delays.

Table S1. Mean time (SD) elapsed between encoding and follow-up tests for participants in Studies 1 ($N = 53$) and 2a (Sleep $N = 39$; Wake $N = 38$).

* Subset of Study 2a participants who performed T5 (Sleep $N = 26$, Wake $N = 23$).

Study	Delay between encoding and T2 (hours)	Delay between encoding and T3 (days)	Delay between encoding and T4 (days)	Delay between encoding and T5 (days)*
Study 1	24.52 (3.64)	6.98 (.24)	28.31 (1.72)	--
Study 2a Sleep	10.97 (0.26)	7.05 (0.17)	28.02 (0.10)	366.64 (7.95)
Study 2a Wake	11.24 (1.20)	7.61 (3.40)	28.01 (0.18)	368.89 (7.32)

Differential effects of sleep on sequence and featural memory not accounted for by psychometric characteristics

Our four *a priori* test forms were constructed to be equivalent in terms of item characteristics and difficulty across forms and administration sessions. Moreover, there was no differential effect of item difficulty, validity (true/false status) or content repetition on sequence versus featural memory.

A 2 (Retrieval Type: detail vs. sequence) × 2 (item validity: true vs. false) ANOVA revealed neither effects of Retrieval Type ($F(1, 269) = 2.67$, $p = .103$) nor item validity ($F(1, 269) = .38$, $p = .540$) on overall item accuracy across participants (see Figure S1).

Fig. S1. **Average accuracy of responses to test items.** Dots depict mean accuracy for 135 Featural and 138 Sequence items, according to whether the item statement was true or false. Dots represent individual test items, averaged across participants and timepoints (dots reflecting items with overlapping mean values are not fully separated). Shaded half-violin plots show the distribution of item-level accuracy across Retrieval Type and item validity. Boxplots display the medians, lower, and upper quartiles.

We included item validity in our main model of sequence vs. featural memory change. Consistent with the above item-level analysis, at the participant level, item validity did not predict response accuracy ($\chi^2(1) = .289$, $p = .591$) and did not interact with Time ($\chi^2(1) = .463$, $p = .496$) nor Retrieval Type ($\chi^2(1) = .017$, $p = .896$), nor was there a three-way interaction ($\chi^2(1) = .739$, $p = .390$). This means that the decline in detail memory over time, and the difference between featural and sequence memory change, were not driven differently by misses (responding 'false' to a true item) and false alarms (responding 'true' to a false item).

We considered that the contrasting forgetting curves for featural and sequence memory could be partly due to differential effects of repeated testing, rather than by time per se. Sequence statements, though trial-unique, comprise pairwise recombinations of a common set of tour events (e.g., "You encountered X before Y" and "You encountered Z before X"). Conversely, feature statements each referred to a specific item or tour feature, though for each target item in the tour we probed multiple features (e.g., its colour, size, etc.). If sequence memory benefited uniquely across the testing procedure

via latent reactivation of intervening items (12), then sequence memory performance would increase across trials *within* a test, either in absolute terms or relative to featural memory performance. We did not find evidence for differential featural and sequence memory performance across trials within T1: there was neither a main effect of trial number (β = 0.002, z = -.84, p = .40) on memory accuracy across successive T1 trials, nor an interaction between trial number and retrieval type (β = 0.01, z = 1.12, p = .234). These results suggest that trial-to-trial effects within T1 cannot account for our observed effects of change in sequence versus featural memory from T1-T2. We cannot rule out an effect of repeated testing on memory performance across T2-T4, which would require separate study assessing delayed recall in the absence of repeated testing.

Sequence lag effects

As expected, sequence memory accuracy increased as the inter-item lag between items in question increased (main effects of lag in Study 1 and Study 2a: $\chi^2(2) = 8.802$, p = .012, and $\chi^2(2) = 8.59$, p = .014) (see Figure S2). In Study 1, there was no significant effect of Time ($\chi^2(1) = .466$, p = .495), but in Study 2a there was ($\chi^2(1) = 12.59$, p = < .001); compared to T1, sequence memory across all lag levels was preserved after 12 hours (z = 1.56, p = 0.238) and 1 week (z = 0.71, p = 0.476), and only showed signs of decline after one month (z = 4.41, p < 0.001).

We considered the possibility that shorter lag pairs – requiring finer-grained spatiotemporal context retrieval – would decline more over time than coarser-grained far pairs, which could indicate a transformation from more specific to more gist-like memory for sequence information. We did not observe evidence for this effect: the Lag × Time interaction effect was neither significant in Study 1 ($\chi^2(2) = 1.74$, p = 0.418), nor in Study 2 ($\chi^2(2) = 1.79$, p = 0.408), nor did lag interact with Group ($\chi^2(2) = 2.55$, p = 0.279) or Group × Time in Study 2a ($\chi^2(2) = 2.08$, p = 0.353). These findings suggest that although finer-grained spatiotemporal context representations are more difficult to retrieve overall, they are not forgotten significantly faster than coarser-grained representations, nor are they differentially affected by sleep per se.

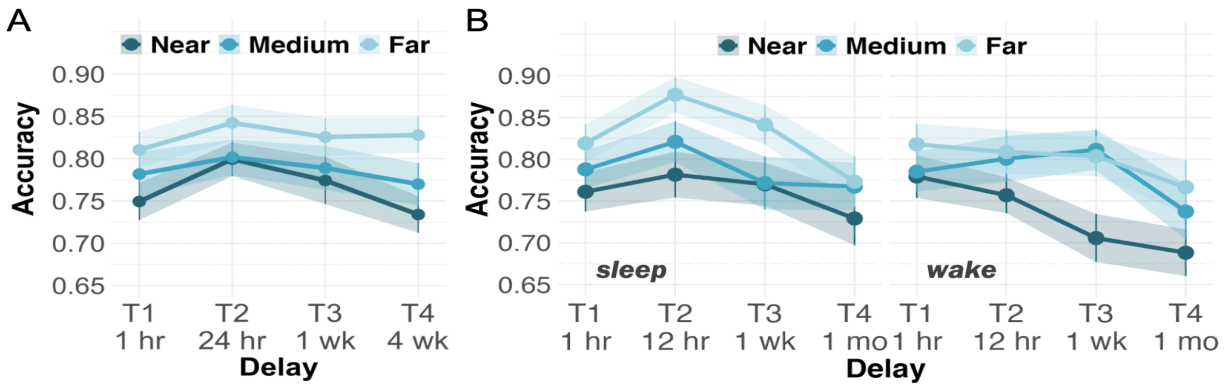

Fig. S2. Effects of inter-item lag on sequence memory accuracy in Study 1 (panel A; *N* = 53), and Study 2a (panel B; *N* = 39, *N* = 38 for Sleep and Wake groups, respectively). Dots depict means, vertical lines and shaded regions depict between-subjects standard errors.

Study 2a: Individual differences in overnight memory change

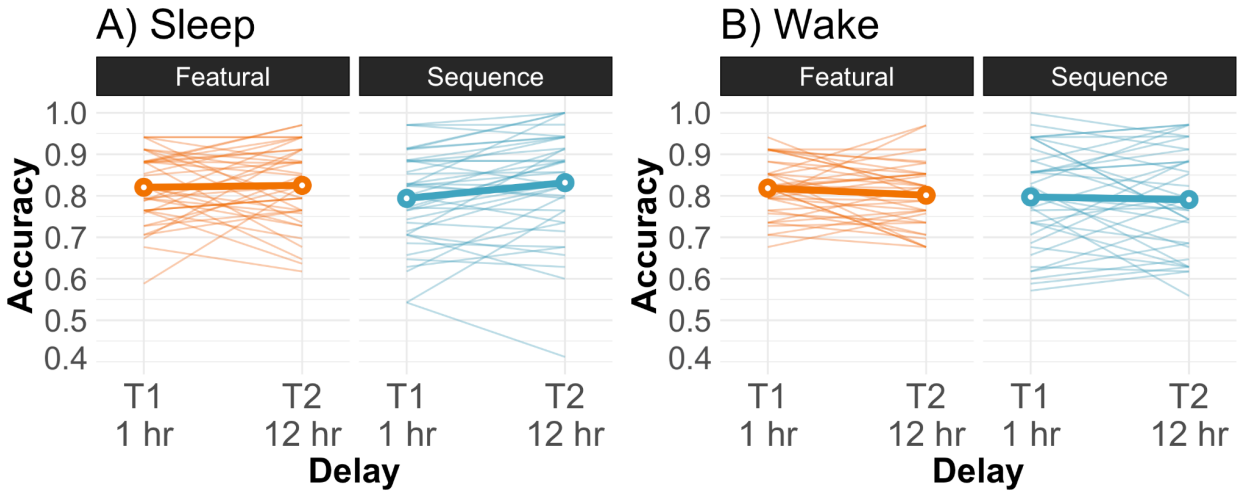

Fig. S3. Individual differences in overnight featural (orange) and sequence (blue) memory either after a night of sleep (panel A; *N = 39*) or an equivalent period of wakefulness (panel B; *N = 38*) for participants in Study 2a. Thick circles with white fill and thick solid lines depict group averages; each thin line depicts an individual participant.

Study 2b Results

Table S2. Descriptive statistics of polysomnography and memory performance (T2-T1) data from the larger sleep sample (*N* = 49). Standard deviations are included in parentheses.

	Mean (SD)	**Minimum**	**Maximum**
Total sleep time (min)	420.14 (35.08)	312.5	469
Sleep efficiency (%)	87.64 (7.09)	68.9	96.8
N1 duration (min)	41.14 (19.57)	15	103
N2 duration (min)	224.76 (38.10)	147.5	302
SWS duration (min)	87.40 (29.96)	24	166.5
REM duration (min)	65.22 (20.18)	30	118.5
Non-REM duration (min)	353.30 (35.31)	259.5	396.5
Average spindle count (raw value)	1505 (259)	1024	1985
Average half-wave count (raw value)	10,635 (3627)	2322	18,579
Overnight feature change (T2-T1)	-0.002 (0.102)	-0.191	0.242
Overnight sequence change (T2-T1)	0.026 (0.073)	-0.137	0.225

Table S3. Effects from linear mixed effects model assessing relationship of sleep macrostructure variables to memory performance on T1 and T2 (df = 1,131). The dependent variable was each participant's average memory score at each timepoint split by featural vs. sequence (where memory change is assessed via interactions with Time); therefore, only effects involving the Time factor (i.e., T1, T2; immediate vs 12-hr testing) are listed. The interaction between Time and Total Sleep Time (TST) was included as a covariate of non-interest. Asterisks indicate statistically significant effects. VIF = Variance Inflation Factor.

Effect	***F* value**	***p* value**	η^2	**VIF**
Time x Retrieval Type	1.85	.176	0.01	2.00
Time x N1 duration	1.09	.298	< 0.01	3.77
Time x N2 duration	1.94	.167	0.01	8.67
Time x SWS duration	4.51	.036 *	0.03	7.25
Time x REM duration	3.84	.052	0.03	2.95
Time x TST	0.82	.368	< 0.01	7.69
Time x Retrieval Type x N1 duration	0.09	.766	< 0.01	3.14
Time x Retrieval Type x N2 duration	0.21	.643	< 0.01	2.66
Time x Retrieval Type x SWS duration	0.01	.934	< 0.01	3.34
Time x Retrieval Type x REM duration	0.07	.787	< 0.01	2.17

Table S4. Effects of linear mixed effects model assessing the relationships of sleep spindles and half-waves to memory performance (in separate models; df = 1,141). Only effects involving the Time factor (i.e., T1, T2; immediate vs 12-hr testing) are listed. Asterisks indicate statistically significant effects. VIF = Variance Inflation Factor.

Effect	*F* value	*p* value	η^2	VIF
Spindles				
Time x Retrieval Type	1.92	0.168	0.01	2
Time x Average Spindle Count	8.66	.004 *	0.06	1.18
Time x Retrieval Type x Average Spindle Count	0.06	0.812	< 0.01	2
Half-waves				
Time x Retrieval Type	1.91	0.169	0.01	2
Time x Average Half-wave Count	7.43	.007 *	0.05	1.18
Time x Retrieval Type x Average Half-wave Count	0.06	0.806	< 0.01	2

Table S5. Effects of linear mixed effects model assessing the relationships of coupled vs. uncoupled spindles to memory performance (df = 1,138). Only effects involving the Time factor (i.e., T1, T2; immediate vs 12-hr testing) are listed. Asterisks indicate statistically significant effects. VIF = Variance Inflation Factor.

Effect	*F* value	*p* value	η^2	VIF
Time x Retrieval Type	1.90	0.170	0.01	2.00
Time x Average Coupled Spindles	9.32	0.003 *	0.06	1.21
Time x Average Uncoupled Spindles	0.07	0.792	< 0.01	1.21
Time x Retrieval Type x Average Coupled Spindles	0.01	0.909	< 0.01	2.02
Time x Retrieval Type x Average Uncoupled Spindles	0.22	0.640	< 0.01	2.02

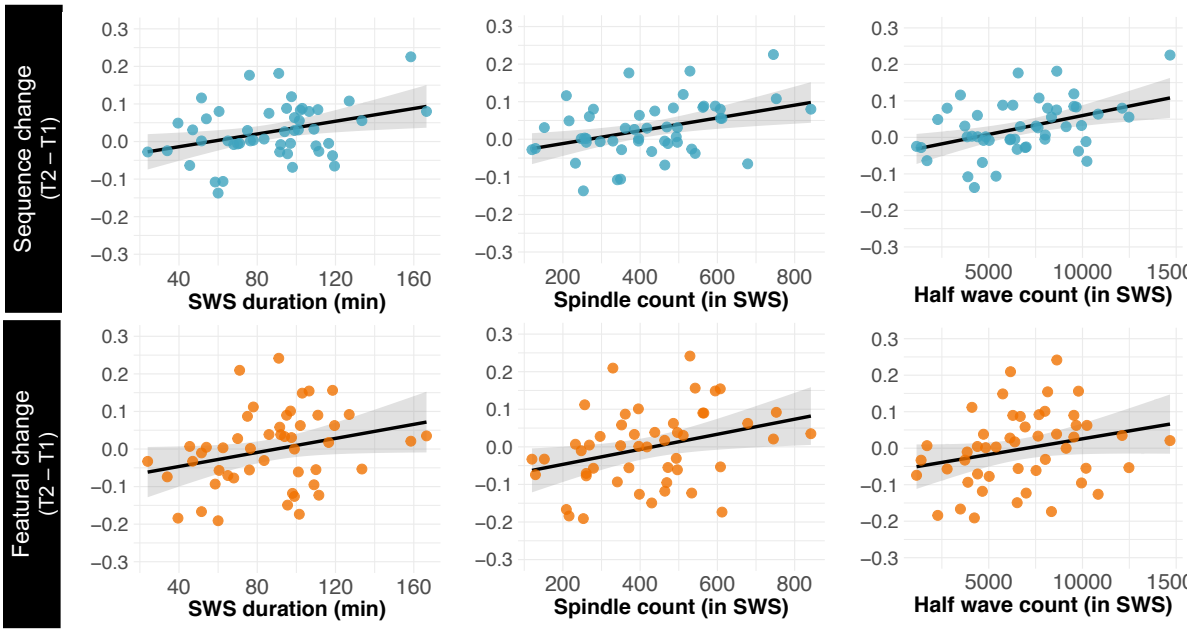

Fig. S4. Pearson correlations following up on significant effects from models testing the relationship of (T1 to T2) change in memory for sequence (top panel) and features (bottom panel) to macrostructural and microstructural sleep measures (N = 49; see Tables S3-S5). The line reflects the best fit with 95% confidence intervals displayed. Sequence memory was positively correlated with SWS (r(47) = 0.35, p = .015, 95% CI [0.073 – 0.573]; uncorrected, two-sided significance test), spindle count (r(47) = 0.38, p = .006, 95% CI [0.116 – 0.600]), and half-wave count (r(47) = 0.42, p = .002, 95% CI [0.162 – 0.630]). Featural memory was positively related to SWS (r(47) = 0.28, p = .055, 95% CI [-0.006 – 0.517]), spindle count (r(47) = 0.33, p = .022, 95% CI [0.049 – 0.556]) and half-wave count (r(47) = 0.25, p = .077, 95% CI [-0.028 – 0.500]). Overall memory (not pictured) was positively correlated with SWS duration (r(47) = 0.38, p = .007, 95% CI [0.112 – 0.599]), spindle count (r(47) = 0.44, p = .002, 95% CI [0.179 – 0.640]), and half-wave count (r(47) = 0.41, p = .004, 95% CI [0.142 – 0.618]).

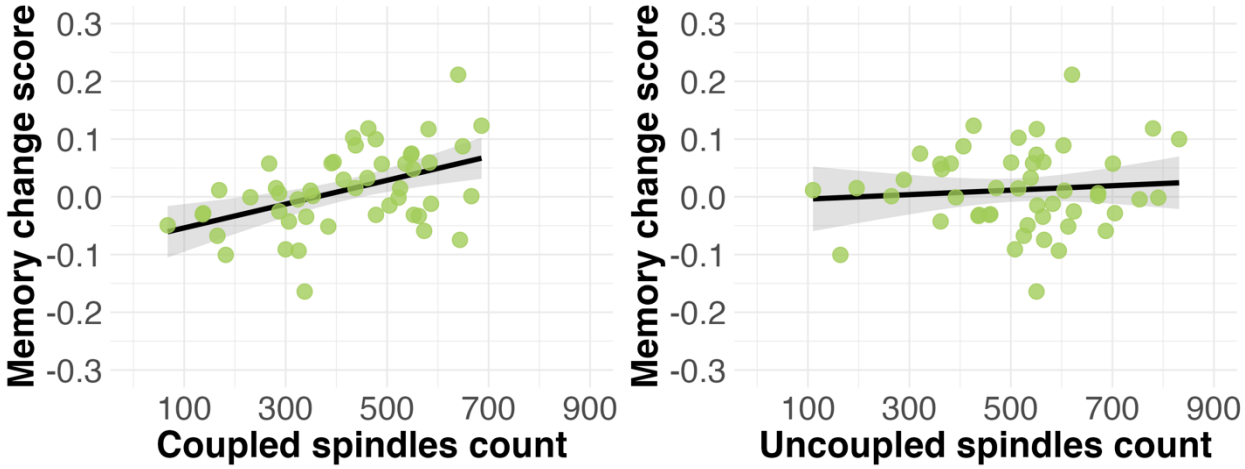

Fig. S5. Relationship between coupled vs. uncoupled spindles and overnight memory change. Pearson correlation between coupled vs. uncoupled spindles in SWS and overnight (T1 to T2) overall memory change (N = 49). The line reflects the best fit with 95% confidence intervals displayed. The correlations for coupled spindles (r(47) = 0.46, p < .001, 95% CI [0.209 – 0.658]; uncorrected two-sided significance test) and uncoupled spindles (r(47) = 0.09, p = .545, 95% CI [0.199 – 0.360]) are statistically different from each other, Steiger's z = 2.11, p = .038.

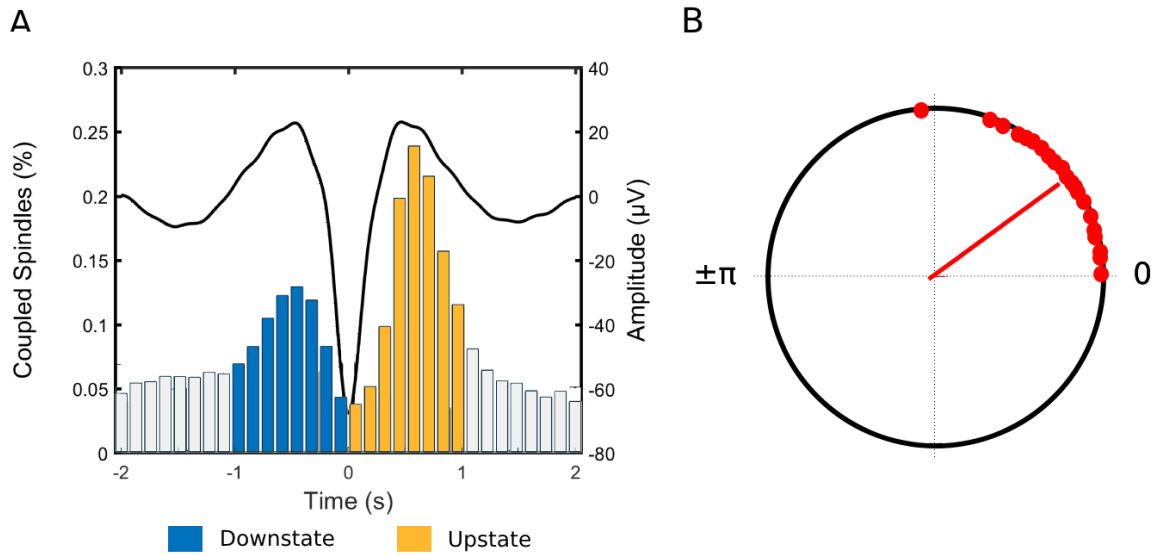

Fig. S6. Temporal and phase distribution of spindle-slow wave coupling. *A.* Coupled slow wave-spindle histogram for all participants. Each bar represents the number of coupled spindles detected in an interval of 125 milliseconds divided by the total number of spindles. Average slow wave oscillation for all participants is superimposed in black. ***B.*** Circular plot of preferred phase for each individual (slow wave phase at spindle amplitude maximum). Red dots denote an individual preferred phase (0° slow wave down-to-up state, ± 180° slow wave up-to-down state). All

participants showed a non-uniform distribution (Rayleigh test $p < 0.001$ for all 49 participants). The direction of the line indicates the preferred direction of the grand average. All individuals exhibit spindles adjacent to, or immediately preceding, the positive slow wave peak at 0^o (V-test; $p < 0.001$ for all 49 participants).

Spindle-SO coupling results using a refined method targeting SO peak (upstate) vs. trough (downstate)

To augment our spindle-SO coupling results reported in the main manuscript, we developed a refined method to detect spindles that are specifically coupled to either the upstate or downstate of the slow wave, using the same electrode pairings (F3-C3 and F4-C4) as in the main analysis. Instead of relying on a fixed time window of ±2 seconds around the slow wave's negative peak, we created individualized time windows for each positive half-wave (upstate) and negative half-wave (downstate)—based on their specific durations. This approach ensures that spindles occurring within these windows are tightly coupled to the corresponding part of the slow wave. As with the fixed window length approach, spindle peak latency served as the reference point. In rare cases where the spindle peak latency occurred during the transition from the down to upstate, the event was considered coupled to both segments. On average, 12 spindle events (SD = 7) were classified as coupled to both parts of the slow wave. This method offers an advantage over the fixed ±2 seconds time window approach by either reclassifying events located at the tails of the slow wave ±2 seconds time window (indicated by the grey bars in Figure S6A) into more appropriate half-wave segments or removing them altogether, thus enhancing detection accuracy.

As reported in the main manuscript, coupled, but not uncoupled spindles were related to overnight change in overall memory performance. Considering the SO positive half-waves, spindles coupled to slow waves' upstate—but not uncoupled spindles—were associated with overall memory enhancement (interactions with Time: $F(1, 138) = 8.14$, $p = .005$, partial eta squared = 0.06; $F(1, 138) = 0.49$, $p = .484$, eta squared ~ 0.00, for coupled and uncoupled spindles respectively. Similarly, spindles coupled to slow waves' downstate—but not uncoupled spindles—were associated with overall memory enhancement (interactions with Time: $F(1, 138) = 4.07$, $p = .046$, partial eta squared = 0.03; $F(1, 138) = 0.41$, $p = .521$, eta squared ~ 0.00, for coupled and uncoupled spindles respectively). When spindles coupled to the upstate and downstate were considered in relation to overnight change in memory performance in a single model, only the effect for coupled spindles in the upstate was significant ($F(1, 138) = 5.04$, $p = .026$, partial eta squared = 0.04, vs $F(1, 138) = 1.04$, $p = .311$, partial eta squared ~ 0.00 for coupled spindles in the downstate), suggesting coupled spindles in the upstate uniquely relate to overnight change in memory performance.

Exploratory analysis of slow versus fast spindles

Exploratory analyses of slow (11–13.5 Hz) versus fast (13.5–16 Hz) spindle counts in relation to memory performance, in separate mixed effects models, indicated that slow spindles were associated with significant overnight improvements in overall memory

performance in a similar manner to total spindle counts (interaction with Time: $F(1, 141) = 10.26$, $p = .002$, partial eta squared = 0.07; no significant three-way interaction with Time and Retrieval Type: $F(1, 141) = 0.11$, $p = .738$, partial eta squared < 0.01). Fast spindles, on the other hand, were not associated with overnight memory change (no significant interaction with Time ($F(1, 141) = 0.07$, $p = .787$, partial eta squared < 0.01) nor Time and Retrieval Type ($F(1, 141) = 0.66$, $p = .416$, partial eta squared < 0.01)). Considering bivariate correlations, overnight memory change was positively correlated with the average number of slow spindles in SWS ($r(47) = 0.47$, $p < .001$, 95% CI [0.223 – 0.667]) but not fast spindles ($r(47) = -0.04$, $p = .777$, 95% CI [-0.319 – 0.242]). These correlations are significantly different from each other: Steiger's $z = 2.38$, $p = .021$).

References

1. J. A. Horne, O. Ostberg, A Self Assessment Questionnaire to Determine Morningness Eveningness in Human Circadian Rhythms. *International journal of chronobiology* **4**, 97–110 (1976).

2. D. J. Buysse, C. F. Reynolds, T. H. Monk, S. R. Berman, D. J. Kupfer, The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research* **28**, 193–213 (1989).

3. C. A. Smyth, Evaluating sleep quality in older adults: the Pittsburgh Sleep Quality Index can be used to detect sleep disturbances or deficits. *Am J Nurs* **108**, 42–50; quiz 50–51 (2008).

4. R. K. Malhotra, A. Y. Avidan, "Sleep Stages and Scoring Technique" in *Atlas of Sleep Medicine*, (Elsevier, 2014), pp. 77–99.

5. L. B. Ray, *et al.*, Expert and crowd-sourced validation of an individualized sleep spindle detection method employing complex demodulation and individualized normalization. *Front Hum Neurosci* **9**, 507 (2015).

6. A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* **134**, 9–21 (2004).

7. A. Bersagliere, P. Achermann, Slow oscillations in human non-rapid eye movement sleep electroencephalogram: effects of increased sleep pressure. *J Sleep Res* **19**, 228–237 (2010).

8. I. Feinberg, *et al.*, Period and amplitude analysis of 0.5-3 c/sec activity in NREM sleep of young adults. *Electroencephalogr Clin Neurophysiol* **44**, 202–213 (1978).

9. B. A. Geering, P. Achermann, F. Eggimann, A. A. Borbély, Period-amplitude analysis and power spectral analysis: a comparison based on all-night sleep EEG recordings. *J Sleep Res* **2**, 121–129 (1993).

10. A. Rechtschaffen, A. Kales, *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects* (Government Printing Office, 1968).

11. J. Rasbash, W. J. Steele, H. Browne, A. Goldstein, "A User's Guide to MLwiN: Version 2.26. Centre for Multilevel Modelling" (University of Bristol).

12. S. DuBrow, L. Davachi, The influence of context boundaries on memory for the sequential order of events. *J Exp Psychol Gen* **142**, 1277–1286 (2013).