Supplementary Materials for

Item-specific memory reactivation supports overnight memory consolidation

Jing Liu, Tao Xia, Danni Chen, Ziqing Yao, Minrui Zhu,

James W. Antony, Tatia M.C. Lee*, Xiaoqing Hu*

*Corresponding author. Email: tmclee@hku.hk; xiaoqinghu@hku.hk;

This PDF file includes:

Fig. S1 to S9

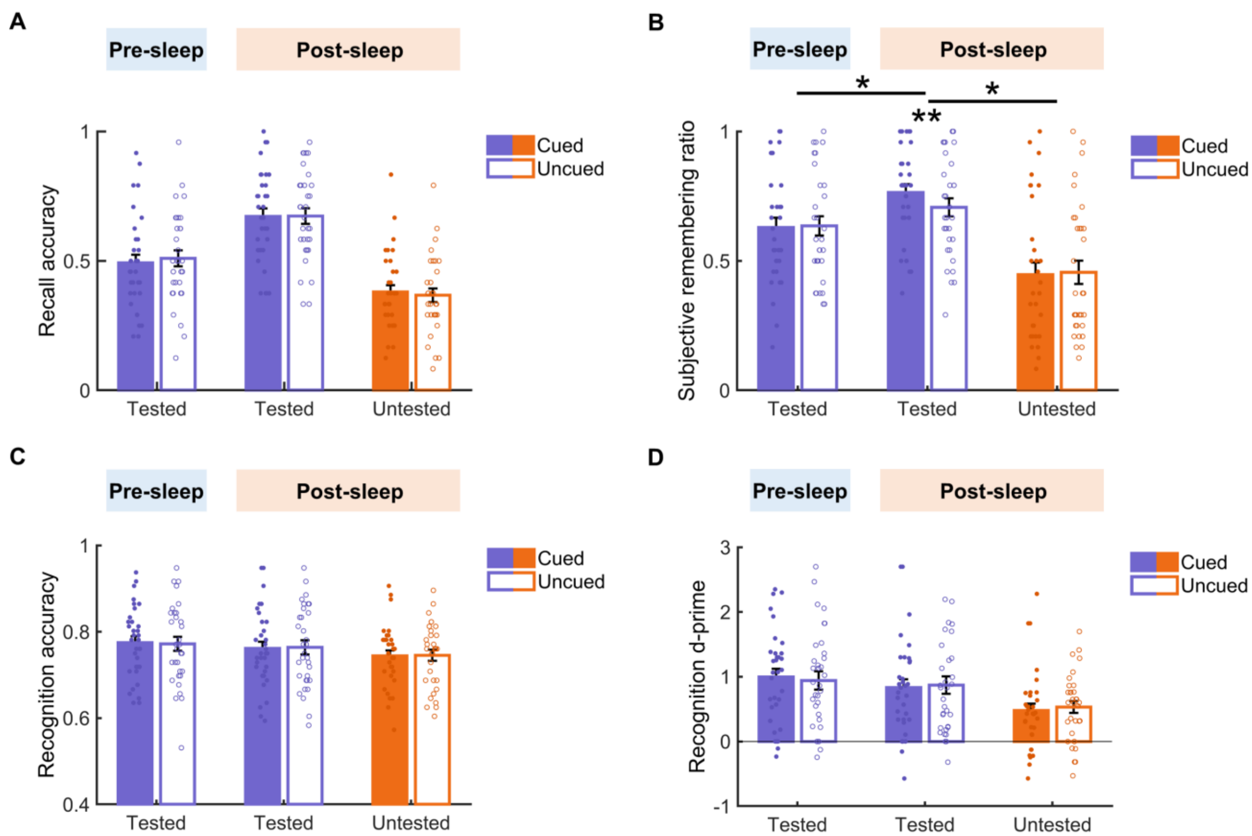

Fig S1. Behavioral results. (A) Category-report accuracy during pre- and post-sleep cued-recall tasks. In the cued recall task, auditory cue words were presented, and participants were asked to report both the category of the associated pictures and their subjective remembering of the associated pictures. Regardless of TMR cueing, category-report accuracy for tested items was greater in the post-sleep test than that in the pre-sleep test ($t(32) = 9.60$, $p < 0.001$), and memory accuracy for the tested items was greater than the untested items ($t(32) = 14.50$, $p < 0.001$) in the post-sleep test. Further examining how pre-sleep testing impact the TMR benefits by performing a 2 (cued vs. uncued) by 2 (tested vs. untested) repeated measures ANOVA on the post-sleep category-report accuracy, we found neither a significant interaction effect ($F(1,32) = 0.23$, $p = 0.637$) nor a significant main effect of TMR benefits (cued vs. uncued, $F(1,32) = 0.16$, $p = 0.690$). In addition, there was no significant TMR benefits (cued vs. uncued) by time (pre-sleep vs. post-sleep) interaction effect for tested items ($F(1,32) = 0.74$, $p = 0.397$). **(B)** Subjective remembering during pre- and post-sleep cued-recall tasks. Subjective remembering was measured by the ratio of trials that participants reported "remember" during the cued-recall test regardless of the following category report accuracy. The results found that subjective remembering for the post-sleep tested items was greater than both the pre-sleep tested items and post-sleep untested items (ps < 0.001) regardless of TMR cueing. Further analysis revealed a significant TMR benefits by pre-sleep testing (tested vs. untested) interaction effect on the post-sleep test and a significant TMR benefits (cued vs. uncued) by time (pre- vs. post-sleep) interaction effect for tested items (ps < 0.019). Simple-effects analysis revealed that both interaction effects were driven by greater subjective remembering for cued than uncued items among the tested items during the post-sleep test ($t(32) = 3.60$, $p = 0.001$). **(C)** Recognition accuracy. Both the pre-sleep tested items and post-sleep tested items showed greater accuracy than the post-sleep untested items (ps < 0.032) regardless of TMR cueing, while no significant difference between pre-sleep tested items and post-sleep tested items ($t(32) = 1.76$, $p = 0.089$). Further analyses revealed neither a significant TMR by pre-sleep testing interaction effect on the post-sleep test nor a significant TMR by time interaction effect for tested items (ps > 0.610). **(D)** Recognition d-prime. Both the pre-sleep tested items and post-sleep tested items showed greater d-prime in the recognition test than the untested items (ps < 0.001) when regardless of TMR cueing, while no significant difference between pre-sleep tested items and post-sleep tested items ($t(32) = 1.58$, $p = 0.123$). Further analyses revealed neither a significant TMR by pre-sleep testing interaction effect on the post-sleep test nor a significant TMR by time interaction effect for tested items (ps > 0.475).

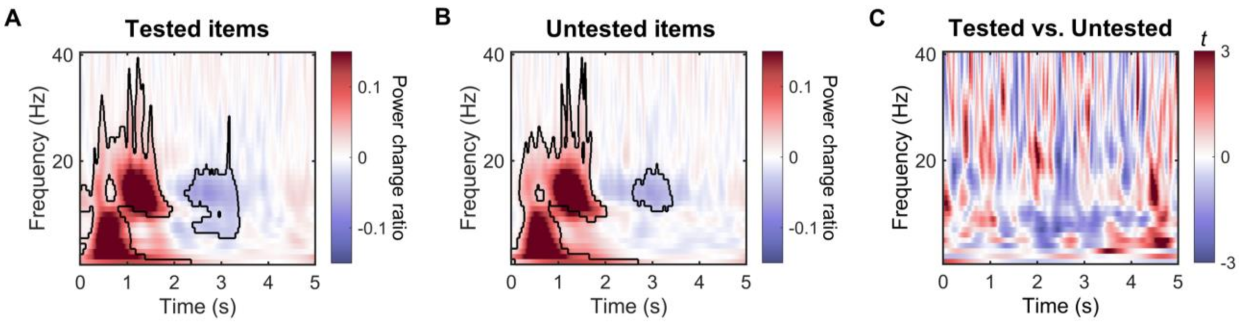

Fig S2. TMR cue-elicited EEG activities for the tested and untested items. (A) For tested items, TMR cues elicited increased EEG activities from 0 to 2360 ms post cue onset in a broad range of frequencies (i.e., 1-40 Hz) ($p_{corr} < 0.001$), followed by a decreased sigma band activity (2230-3390 ms) ($p_{corr} = 0.006$) as compared with the pre-cue baseline (-1 s to -0.5 s relative to cue onset). **(B)** For untested items, TMR cues elicited increased EEG activities from 0 to 2700 ms post cue onset in a broad range of frequencies (i.e.,1-40 Hz) ($p_{corr} < 0.001$), followed by a decreased sigma band activity (2590-3430 ms) ($p_{corr} = 0.017$). **(C)** No significant difference was found between the TMR cue-elicited power for tested and untested items ($ps_{corr} > 0.335$).

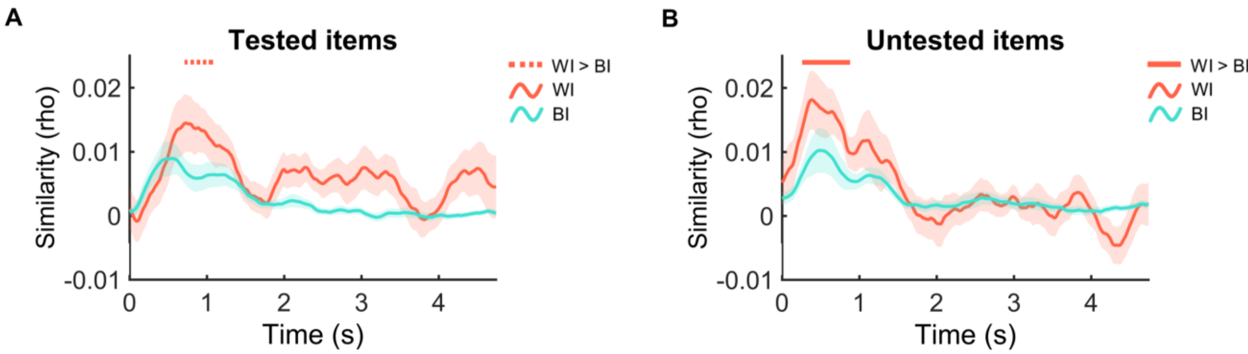

Fig S3. Item-specific representations for tested and untested items during TMR regardless of subsequent memory performance. Item-specific representations were examined by contrasting the within-item (WI) similarity versus between-item (BI) similarity. **(A)** For tested items, there was a marginally significant cluster showing greater WI similarity than BI similarity (720-1090 ms, $p_{corr} = 0.057$). **(B)** For untested items, there was a significant cluster showing greater WI similarity than BI similarity (270-890 ms, $p_{corr} = 0.009$). In addition, there was no significant interaction effect between item-specificity (WI vs. BI) and pre-sleep testing (tested vs. untested) ($p_{corr} > 0.097$). The marginally significant cluster is indicated by the dashed horizontal lines on the top, while the solid horizontal line on the top indicates the significant cluster.

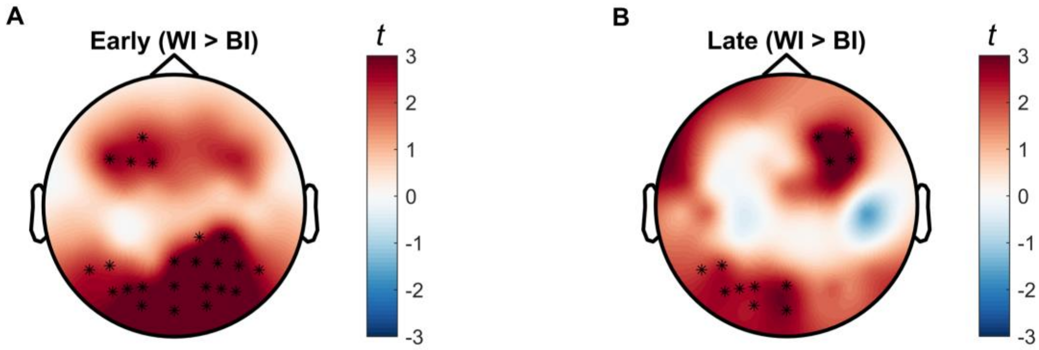

Fig S4. Topographies of item-specific representations for subsequently remembered items in the time windows of early (A) and late (B) clusters. Referring to Fig. 3C in the main text, item-specific representations were found for subsequently remembered items in an early (620-1350 ms) and a late cluster (2270-3190 ms). Contributing channels for the item-specific representations in the early and late clusters were examined by contrasting the WI similarity versus BI similarity at the single-channel level. The channel clusters were obtained via the non-parametric permutation test. **(A)** Two significant spatial clusters that showed the EEG channels on the left frontal and posterior part of the scalp contributed to the item-specific representations in the early cluster ($p_{corr} < 0.049$). **(B)** Another two significant spatial clusters that showed the EEG channels on the right frontal and left posterior part of the scalp contributed to the item-specific representations in the late cluster ($p_{corr} < 0.029$). Channels in the clusters that showed significantly greater WI similarity than BI similarity were highlighted by stars (*).

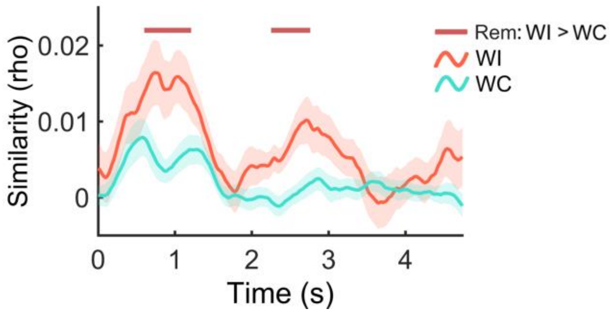

Fig S5. Item-specific representations for subsequently remembered items after constraining the between-item similarity to the similarity of trial pairs with pictures from the same category (i.e., within-category similarity). Item-specific representations were then examined by contrasting the within-item (WI) similarity versus within-category (WC) similarity. The results revealed two significant clusters that showed greater WI similarity than WC similarity (i.e., 610-1220 ms, $p_{corr} = 0.011$; 2260-2770 ms, $p_{corr} = 0.024$) for subsequently remembered items, a pattern highly consistent with Fig. 3C in the main text.

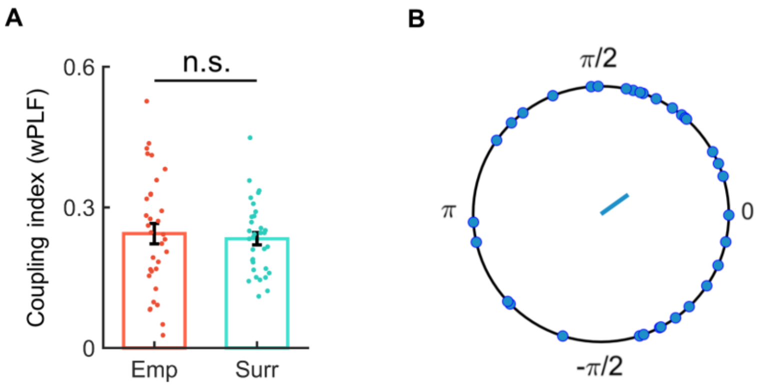

Fig S6. Representational strength was not coupled to the phase of continuous slow oscillations for subsequently forgotten items. The coupling effect was examined by computing the weighted phase locking factor (wPLF) between the phase of slow oscillations and the concurrent representational strength (i.e., WI similarity) for the empirical data. We then contrasted the empirical wPLF with the surrogate wPLF that was generated by shuffling the corresponding relationship between the empirical phase data and representational strength. **(A)** No significant difference was found between the empirical and surrogate coupling index for forgotten items ($t(32) = 0.54$, $p = 0.592$). **(B)** Directly examining the preferred phase when representational (re)activations occurred by extracting the phase from the wPLF for each participant and employing the Rayleigh Z test to examine the distribution of phases against the uniform distribution across participants, we found no significant phase clustering ($z = 2.25$, $p = 0.105$; mean vector length (*mlv*) = 0.26, preferred phase: 35.55°). n.s.: not significant.

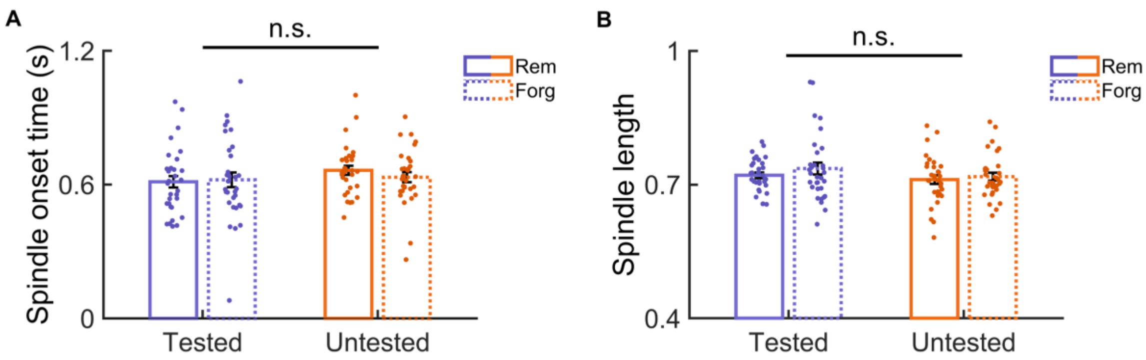

Fig S7. The onset time and duration of the spindles that occurred within the first 2s following TMR cues were not different either between tested and untested items or between subsequently remembered and forgotten items. We extracted the spindle onset time and duration for the spindles that occurred within the first 2s and then performed testing (tested vs. untested) by subsequent memory (remember vs. forget) repeated measures ANOVA respectively. The results found that neither a significant interaction for the spindle onset time ($F(1, 31) = 0.73$, $p = 0.400$, **A**) nor a significant interaction for the spindle length ($F(1, 31) = 0.37$, $p = 0.546$, **B**). n.s.: not significant.

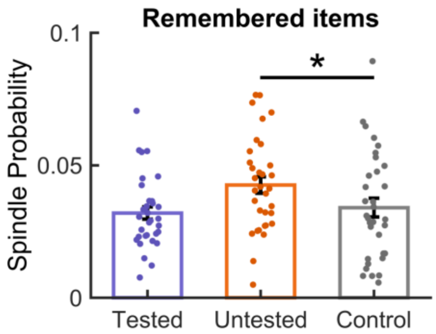

Fig S8. Greater spindle activities for subsequently remembered untested items than control cues in the time windows (2352 to 3020 ms post cue). Referring to the temporal cluster (i.e., 2352 to 3020 ms post cue) that showed a significant interaction effect between testing (tested vs. untested) and subsequent memory (remember vs. forget) in Fig. 5C of the main text, we further directly contrasted the spindle activity in this cluster between subsequently remember tested (or untested) items versus the control cues. The results revealed greater spindle probability for subsequently remembered untested items than the control cues ($t(32) = 2.27$, $p = 0.030$), while no significant difference between subsequently remembered tested items and control cues ($t(32) = -0.59$, $p = 0.561$). *: $p < .05$.

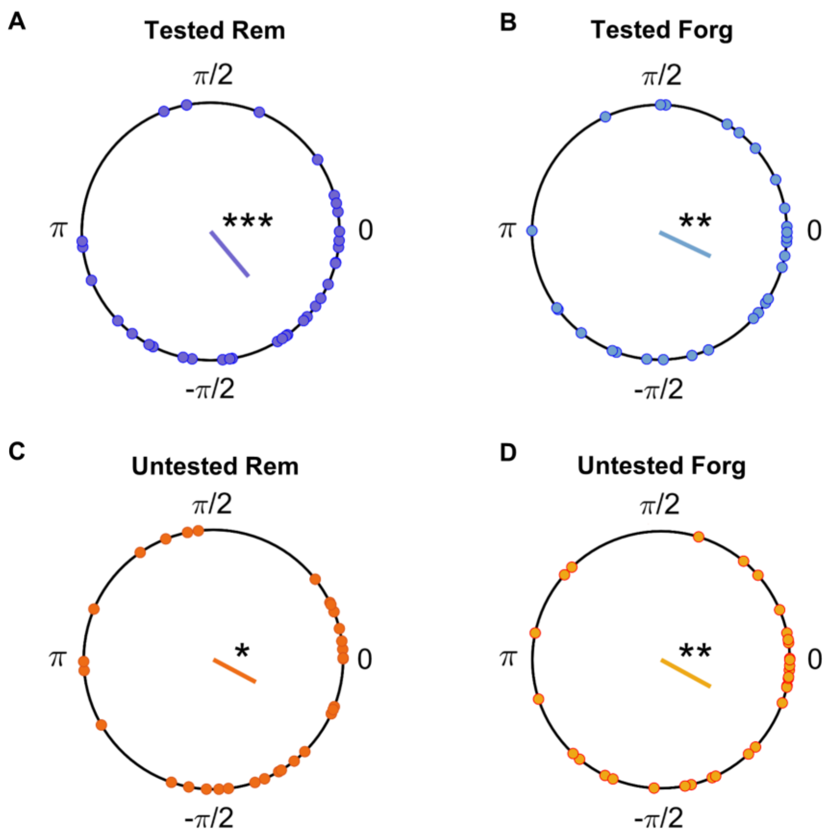

Fig S9. Spindles in overlapped time windows (2440-2960 ms) that showed item-specific memory reactivation and increased spindle activities were coupled to the up-state of slow oscillations. For trials that showed both the spindles and slow oscillations in the time windows (i.e., 2440-2960 ms), we first extracted the preferred phase of the SO for each spindle in an extended time window (i.e., 2-4 s) and then the preferred phase for each participant was obtained by computing the mean phase across trials (CircStat toolbox). We then employed the Rayleigh Z test to examine whether the distribution of phases would be against the uniform distribution across participants. The results showed that spindle activities for subsequently remembered and forgotten tested items **(A-B)** as well as for subsequently remembered and forgotten untested items **(C-D)** were all significantly coupled to the up-state of SOs (*ps* < 0.015). *: $p < .05$; **: $p < .01$; ***: $p < .001$.