The Role of Rapid Eye Movement Sleep in Neural Differentiation of Memories in the Hippocampus

Elizabeth A. McDevitt[1], Ghootae Kim[2], Nicholas B. Turk-Browne[3], and Kenneth A. Norman[1]

Abstract

■ When faced with a familiar situation, we can use memory to make predictions about what will happen next. If such predictions turn out to be erroneous, the brain can adapt by differentiating the representations of the cue from the mispredicted item itself, reducing the likelihood of future prediction errors. Prior work by Kim, G., Norman, K. A., and Turk-Browne, N. B. Neural differentiation of incorrectly predicted memories. *Journal of Neuroscience*, *37*, 2022–2031 [2017] found that violating a sequential association in a statistical learning paradigm triggered differentiation of the neural representations of the associated items in the hippocampus. Here, we used fMRI to test the preregistered hypothesis that this hippocampal differentiation occurs only when violations are followed by rapid eye movement (REM) sleep. Participants first learned that some items predict others (e.g., A predicts B) and then encountered a violation in which a predicted item (B) failed to appear when expected after its associated item (A); the predicted item later appeared on its own after an unrelated item. Participants were then randomly assigned to one of three conditions: remain awake, take a nap containing non-REM sleep only, or take a nap with both non-REM and REM sleep. While the predicted results were not observed in the preregistered left CA2/3/dentate gyrus (DG) ROI, we did observe evidence for our hypothesis in closely related hippocampal ROIs, uncorrected for multiple comparisons: In right CA2/3/DG, differentiation in the group with REM sleep was greater than in the groups without REM sleep (wake and non-REM nap); this differentiation was item-specific and concentrated in right DG. REM-related differentiation effects were also greater in bilateral DG when the predicted item was more strongly reactivated during the violation. Overall, these results provide initial evidence linking REM sleep to changes in the hippocampal representations of memories in humans. ■

INTRODUCTION

When we retrieve a memory, related memories often come to mind. In some situations, this may be helpful: For example, you enter a familiar environment and can predict who or what you will encounter. But what if your prediction is wrong and instead becomes a source of interference for memory retrieval? One way the brain might mitigate prediction errors is by adaptively disconnecting the mispredicted item from its old context and binding it to a new, updated context, effectively pushing a mispredicted memory away from its old cue in representational space (i.e., neural differentiation). Here, we investigate whether this process of prediction-based neural differentiation is supported by a period of optimized memory consolidation that includes sleep.

Our work builds on a prior fMRI study by Kim, Norman, and Turk-Browne (2017) that found that prediction errors lead to neural differentiation in the hippocampus. Specifically, they found that when an item predicted in a particular context (e.g., A predicts B) failed to appear and was later restudied in a different context, the neural representations of A and B became less similar in the left CA2/3/dentate gyrus (DG) subregion of the hippocampus. Kim et al. (2017) explained this result in terms of an unsupervised learning principle called the "nonmonotonic plasticity hypothesis" (NMPH; Ritvo, Turk-Browne, & Norman, 2019), which posits a U-shaped relationship between the coactivation of two memories (A and B) and learning; according to the NMPH, strong coactivation of the B memory while retrieving the A memory will lead to strengthening of the connections between A and B, moderate coactivation of the B memory will lead to synaptic weakening, and little or no activation of B will lead to no change in the synaptic connections between A and B (Detre, Natarajan, Gershman, & Norman, 2013; Newman & Norman, 2010). Kim et al. (2017) argued that when A predicts B but B does not appear, this unconfirmed prediction leads to moderate activation of B, which—according to the NMPH—weakens the connections between the unique features of B and the features it formerly shared with A (for further evidence that B is weakened, see Kim, Lewis-Peacock, Norman, & Turk-Browne, 2014); when item B is presented on a subsequent trial, it then activates a different set of features (not shared with A) and incorporates these new features into its neural representation (Ritvo, Nguyen,

[1]Princeton University, Princeton, NJ, [2]Korea Brain Research Institute, Daegu, Republic of Korea, [3]Yale University, New Haven, CT

Turk-Browne, & Norman, 2024; Ritvo et al., 2019; Hulbert & Norman, 2015). The overall effect of this process is neural differentiation—decreased overlap in the populations of neurons that encode A and B.

An important detail of the Kim et al. (2017) study (and some other fMRI studies that have found differentiation, e.g., Favila, Chanales, & Kuhl, 2016) is that the interval between learning and the final measurement of neural representations contained a night of sleep, raising the question of how offline consolidation processes might contribute to the observed representational changes. Numerous studies have found that neural activity is reactivated (i.e., replayed) during sleep (Ji & Wilson, 2007; Louie & Wilson, 2001; Wilson & McNaughton, 1994), and this is thought to be a critical mechanism underlying sleep-dependent memory consolidation (Klinzing, Niethard, & Born, 2019; Diekelmann & Born, 2010). Most work has focused on understanding reactivation during non-rapid eye movement (NREM) sleep and how it supports systems-level consolidation via hippocampo-cortical interactions. There is evidence that reactivation also happens during rapid eye movement (REM) sleep (Abdellahi, Koopman, Treder, & Lewis, 2023; Schönauer et al., 2017; Louie & Wilson, 2001), but results linking REM reactivation and memory consolidation are mixed. Neural network modeling suggests that REM sleep, in particular, serves as a focused period of interleaved replay of related memories (Guerreiro & Clopath, 2024; Singh, Norman, & Schapiro, 2022; Norman, Newman, & Perotte, 2005); accordingly, looking at how representations change locally, relative to one another, might be a better target than looking at systems-level changes when trying to understand the role of REM sleep. During REM sleep, brain activity is not guided by environmental stimuli, and the hippocampus and cortex are relatively uncoupled (Diekelmann & Born, 2010; Cantero et al., 2003). This allows the hippocampus (and cortex) to autonomously rehearse stored memories in an unsupervised, interleaved manner, meaning that spreading activation within each network selects the "targets" for learning, while coactivating related memories in the process. This is in contrast to NREM sleep, when neural activity between the hippocampus and cortex is tightly coupled and the hippocampus is replaying information to cortex (Diekelmann & Born, 2010). One way to think about this is that, during NREM sleep, the hippocampus is busy training cortex; during REM sleep, the hippocampus is freed of that obligation and can focus on fine-tuning its own network of information. This led us to hypothesize that the brain identifies memories vulnerable to prediction errors during wake and then implements the restructuring needed to address those prediction errors during sleep. Specifically, we hypothesized that REM sleep should be the critical sleep stage driving hippocampal neural differentiation.

In a preregistered study, we tested this hypothesis in a day-long experiment using fMRI to measure how neural representations differentiate across periods of wake and sleep. We used the same task and the same general pre/post design as Kim et al. (2017). In the morning, we first obtained prelearning fMRI "snapshots" of each item's initial neural representation by showing participants all relevant items and extracting the spatial pattern of BOLD activity corresponding to each item. Participants then completed the learning task that was previously shown to induce neural differentiation (Kim et al., 2017). Next, participants were randomly assigned to one of three EEG-recorded offline conditions: a nap composed of NREM sleep only (NREM group), a nap with both NREM and REM sleep (REM group), or a period of quiet wakefulness (Wake group). Later the same day, we obtained postlearning (and postsleep, in the case of the nap groups) fMRI snapshots of each item's neural representation. Comparing pre- and postlearning snapshots allows us to assess how the intervening task and subsequent sleep or wake conditions altered the neural representations.

In our preregistration, the overarching prediction was that violations during learning (i.e., instances when A predicted B but B did not appear), followed by REM sleep—versus not having REM sleep (i.e., Wake or NREM only)—would lead to neural differentiation of A and B. Kim et al. (2017) also found that when A predicted B but B did not appear, stronger activation of memory B predicted higher levels of (subsequent) neural differentiation in the hippocampus; they explained this in terms of low activation being associated with no synaptic change and higher (moderate) levels of activation being associated with synaptic weakening and, through this, differentiation. We also tested for this relationship in our study, predicting that it would be more evident in the REM group than in the other two groups. Because the Kim et al. (2017) study found these differentiation effects in the left CA2/3/DG subfield of the hippocampus, we focused initially on this specific subregion.

As described below, we did not find evidence for the predicted effects in the left CA2/3/DG, but we did find the predicted pattern of results (uncorrected for multiple comparisons) in the same hippocampal subfield, lateralized to the right hemisphere. Compared with the Wake and NREM groups, the REM group showed more neural differentiation in the right CA2/3/DG (and right DG alone). Additionally, there was a stronger relationship between B activation and differentiation in bilateral DG in the REM group. Together, these findings provide support for our hypothesis that REM sleep facilitates learning-dependent representational change in the hippocampus, although more work is needed to firmly establish this point and to understand whether hemispheric differences are real or reflect noisy fMRI measurements from small and highly precise anatomical regions.

METHODS

Preregistration

Study procedures, planned analyses, and predictions were preregistered (https://osf.io/p953t; Nemeth et al., 2024).

We originally planned to test 34 participants in each of three experimental groups (total $n = 102$) as stated in the preregistration; this was based on an a priori power analysis. However, data collection was interrupted by the COVID-19 pandemic, and we made the decision to permanently stop data collection so that we could begin data analysis while in-person human participant research was suspended.

Participants

We collected complete data sets from 74 healthy, nonsmoking adults (42 female, eight left-handed, mean age = 20.4 years, range = 18–35 years) from the Princeton University community to participate in exchange for monetary compensation ($20/hr). All procedures were approved by the Princeton University Institutional Review Board for Human Subjects. All participants provided informed consent.

Following our preregistered criteria, two participants were excluded due to poor performance (2.5 *SD* below the mean) on the Session 1 subcategory judgment task, and one participant was excluded due to poor performance (2.5 *SD* below the mean) on the Session 2 reward learning task. Two participants did not agree to share their data publicly and were not included in our analysis. Our final sample included 69 participants ($n = 23$ per experimental group).

Participants reported normal or corrected-to-normal vision, with no history of neurological disorders, psychiatric disorders, major medical issues, or use of medication known to interfere with sleep. Participants also reported normally obtaining 6–9 hr of sleep per night (with weekday bedtime no later than 2 a.m. and wake time no later than 10 a.m.). The Epworth Sleepiness Scale (ESS; Johns, 1992) and the reduced Morningness–Eveningness Questionnare (rMEQ; Adan & Almirall, 1991) were used to screen for excessive daytime sleepiness (ESS score >10 excluded) and extreme chronotypes (rMEQ <8 or >21 excluded). Heavy caffeine users (>3 servings per day) were not enrolled in the study.

Participants were instructed to follow their regular sleep/wake schedule for 1 week before their study and to spend at least 8 hr in bed the night before the study. To confirm adequate sleep was obtained the night before the study, participants completed an online, time-stamped sleep diary; if less than 6.5 hr of sleep was reported, participants were not tested and rescheduled for another day. Participants were asked to abstain from caffeine and alcohol starting at noon the day before the study.

Stimuli

The stimulus materials were the same as in Kim et al. (2017) and consisted of color photographs of indoor and outdoor scenes, male and female faces, and natural and manmade objects presented on a gray background. Stimuli were projected on a screen behind the scanner and viewed with a mirror on the headcoil. Stimuli were presented using the Psychophysics Toolbox for MATLAB (https://psychtoolbox.org).

Experimental Procedures

Prestudy Orientation and fMRI Scan

Approximately 1–14 days before the scheduled study day, participants came to the lab for a study orientation appointment and prestudy fMRI scan. During this appointment, they were informed of all study procedures, provided informed consent, and completed study paperwork.

Participants also completed two runs of a functional localizer task in the scanner. Each run consisted of 15 blocks, with five blocks from each of three categories: faces, scenes, and objects. Participants categorized faces as male or female, scenes as indoor or outdoor, and objects as manmade or natural. Each stimulus was presented for 500 msec, followed by a blank interval of 1000 msec. Each block was 10 trials, and each 15-sec block was followed by 15 sec of fixation (i.e., rest). One run lasted approximately 7.8 min.

This scan served three main purposes: (1) to help participants become acclimated to the scanner environment before the full study day, (2) to help participants learn the button-press response mappings that were also used in the task on the full study day, and (3) to obtain data for training category-specific classifiers to potentially be used in our analyses (however, we do not report results using these data).

fMRI Session 1

An overview of the study is illustrated in Figure 1. Session 1 began at 10:00 a.m. Participants entered the MRI scanner and completed six runs of an incidental encoding task in which they viewed streams of scene and face images with embedded regularities while performing a cover task. Scenes and faces were presented one at a time, and participants performed a subcategory judgment task ("Is the scene indoor or outdoor?" "Is the face male or female?"). Participants fixated on a black central dot that changed to white when a response was recorded. Each trial began with a blink of the fixation dot to signal an upcoming stimulus, followed by the stimulus for 1000 msec and a blank interstimulus interval of 2000 msec. Each run consisted of 192 trials and lasted approximately 10 min.

The sequence of images followed an A–B pair structure: Each pair had scene A as the first item and a different scene B as the second item. These A–B pairs were inserted in the stream continuously among the other pairs, and participants were not made aware of this pair structure. Within each run, there were eight unique pairs for each of two task conditions (violation and nonviolation, see below), for a total of 16 pairs per run (16 pairs × 6 runs = 96 pairs

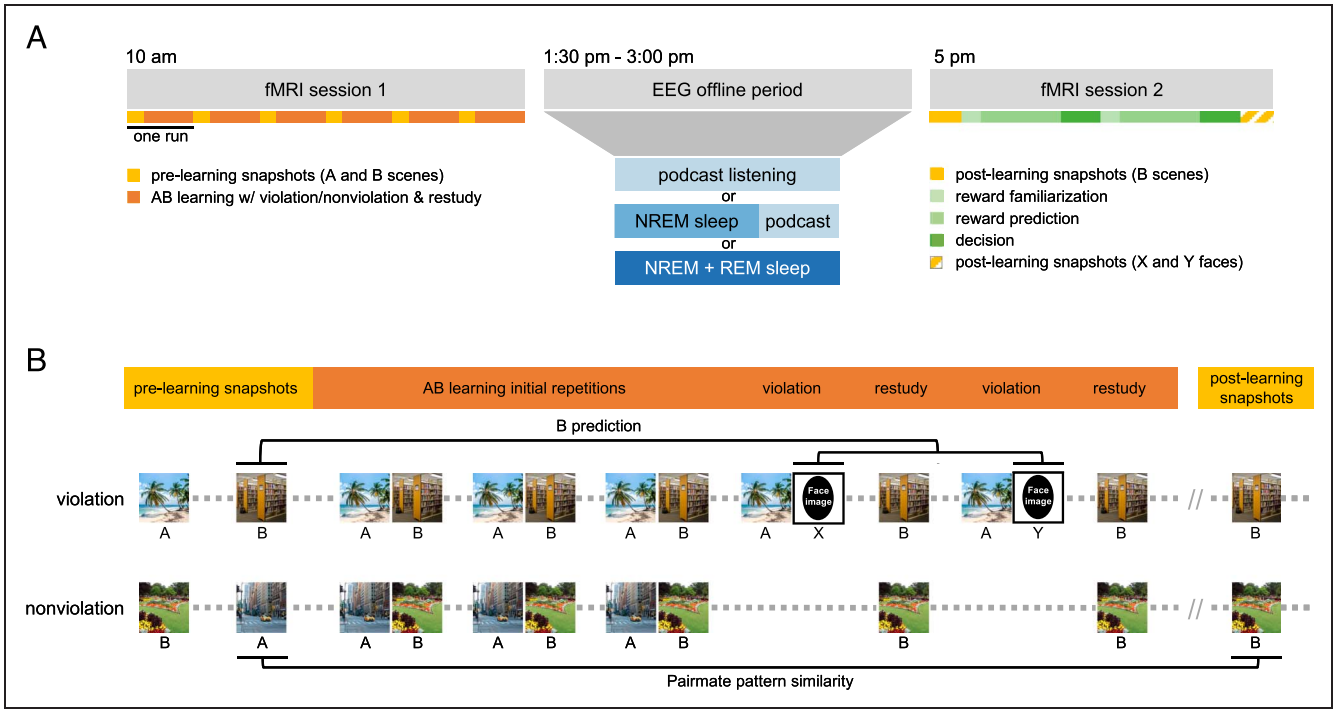

Figure 1. Experimental design and methods. (A) Study day timeline. Participants entered the MRI scanner at 10:00 a.m. and completed an incidental encoding task. Next, participants were randomly assigned to one of three EEG-recorded offline conditions: wake, a short 50-min nap followed by podcast listening, or a long 90-min nap. Participants reentered the MRI scanner at 5:00 p.m. and completed one run of postlearning "B" scene snapshots, the reward association task, and one run of postlearning "X and Y" face snapshots. (B) Task design and analysis schematic. Participants viewed streams of scene and faces images presented one at a time. At the beginning of each run, the A and B members of each pair were shown once separately (i.e., B did not follow A) to obtain prelearning snapshots of each item. During the learning phase, pairs in the violation condition followed a sequence of three initial learning repetitions followed by two cycles of violation and restudy trials (AB-AB-AB-AX-B-AY-B). During violation events, A was followed by a novel face (X or Y), violating the expectation that B should follow A. B was subsequently restudied in a novel context, not preceded by its pairmate A. The nonviolation control condition did not have any violation events but did include two B restudy trials (AB-AB-AB-B-B) to match the frequency and context of B item exposures in the two conditions. Each pair's trials were interleaved with repetitions of other pairs (represented here as gray dots). In Session 2, the B scenes were presented one more time in a random order to take postlearning snapshots. To measure neural differentiation, we correlated voxel patterns for the prelearning snapshot of A and postlearning snapshot of B ("pairmate pattern similarity") for all pairs. To measure the amount of "B prediction" on violation trials, we correlated the prelearning snapshot of B and the pattern of activity evoked by the X and Y violation events, then averaged these values for one B prediction score per pair in the violation condition. The scene images in this figure were sourced from the internet and not used as actual stimuli in our experiment; face stimuli have been replaced with black ovals.

total, 48 pairs per task condition). Scenes assigned to each pair and condition were randomized for each participant. Pair presentation order was also randomized for each participant with the constraint that the minimum and maximum distance between repetitions of the same pair was 2 and 20 pairs, respectively.

At the beginning of each run, the A and B members of each pair were shown once separately (i.e., B did not follow A) and randomly intermixed with scenes from other pairs. This first presentation of each scene is used to estimate each scene's neural representation ("prelearning snapshot"), uncontaminated by its pairmate.

Next, the scenes were shown together as a pair (A followed by B) three times, interleaved with repetitions of other pairs, creating the expectation that B would follow A. Each pair was assigned to one of two within-subject task conditions: violation and nonviolation. For pairs assigned to the violation condition, after the initial three repetitions of the pair, there were two "violation" events in which B failed to follow A. Instead, A was followed once by X and once by Y, where X and Y were novel faces. Following each violation event, the B item was subsequently "restudied" on its own, meaning it appeared in a novel context, not preceded by its pairmate A. Across the entire learning phase, violation pairs followed the sequence AB-AB-AB-AX-B-AY-B (intermixed with other pairs). Pairs in the nonviolation condition also had three initial A-followed-by-B repetitions, but no violation events. To match the frequency of B item exposures in the two conditions, as well as the context in which those B items were exposed, the nonviolation condition also included two B "restudy" events. Across the entire learning phase, nonviolation pairs followed the sequence AB-AB-AB-B-B (intermixed with other pairs). In summary, B was presented five times during the learning phase in both conditions, three times following A and two times not following A. As such, the only difference between the conditions was that the A-followed-by-B prediction was violated in one condition

(by extra presentations of A followed by X and Y, respectively) and not the other.

Offline Period with EEG

Following Session 1, all participants experienced between 90 and 120 min of controlled, offline time. Participants were randomly assigned to one of three offline conditions: wake, a 50-min nap, or a 90-min nap. Given that shorter naps tend to have less REM sleep than longer naps, these nap durations were chosen to increase the likelihood of having naps with and without REM sleep (Schapiro, McDevitt, et al., 2017; McDevitt, Duggan, & Mednick, 2015; McDevitt, Rowe, Brady, Duggan, & Mednick, 2014; Mednick, Nakayama, & Stickgold, 2003). For eventual data analysis, the naps were scored for sleep stages by an expert scorer according to standard criteria (Berry et al., 2012), and participants were regrouped based on the content of their nap: NREM only naps (referred to as the NREM group) or naps with both NREM and REM (referred to as the REM group).

All participants returned to the lab at 12:45 p.m. and had electrodes attached for EEG recording. Participants in the wake condition experienced a period of "quiet wakefulness"; they sat in a chair in the EEG recording room and listened to podcasts for 90 min from approximately 1:30 p.m. to 3:00 p.m. This condition was chosen to control for the reduced visual input and motor movement experienced during a nap while still engaging participants enough to avoid them easily falling asleep. During this time, an experimenter monitored participants via EEG and a camera to make sure they remained awake, and alerted participants at the first sign of Stage 1 sleep.

Participants in the two nap conditions were given a nap opportunity beginning at approximately 1:30 p.m. Sleep was monitored and quantified in real time by an experimenter. Participants in the 50-min nap condition were woken after 50 min of total sleep time (TST) was obtained or at the first sign of REM sleep, whichever occurred first. The remaining amount of controlled offline time (up to 90 min) was filled with podcast listening. Participants in the 90-min nap condition were woken after 90 min of TST was obtained, but no later than 120 min after the beginning of their nap opportunity. For both nap conditions, the nap opportunity was ended if participants spent more than 30 consecutive minutes awake, and the remaining amount of offline time (up to 90 min) was filled with podcast listening.

Since it was expected that not all participants would fall asleep easily or stay asleep for the entire duration of their nap, we preregistered a contingency plan for how to adjust on the fly under very specific circumstances. We employed this adjustment procedure in nine of the 69 participants included in our analyses. If a participant in the 50-min nap condition had REM-onset sleep (i.e., REM preceded SWS), we did not wake the participant at the first sign of REM sleep. Instead, we let this participant sleep for up to 90 min and analyzed their data as part of the REM group ($n = 3$). If a participant in the 90-min nap condition did not obtain REM sleep, this data set was analyzed as part of the NREM group ($n = 2$). If a participant in either nap condition did not fall asleep or only had short, fragmented bouts of Stages 1 and 2 within the first 30 min of the nap period, the nap opportunity was ended and the participant listened to a podcast for the remaining amount of time. This data set was analyzed as part of the Wake group ($n = 4$).

fMRI Session 2

Session 2 began at 5:00 p.m. Participants reentered the MRI scanner and completed three separate tasks in the following order: (1) one run of postlearning B scene snapshots, (2) the reward association task (see Supplemental Materials for task details), and (3) one run of postlearning X and Y face snapshots. The postlearning snapshots followed the same procedure as the Session 1 task (1000 msec stimulus duration, 2000 msec interstimulus interval). For scene snapshots, all B scenes were shown again, in a random order, and participants made indoor/outdoor judgments. For face snapshots, all X and Y faces (used during the Session 1 violation events) were shown in a random order, and participants made male/female judgments. Each snapshot run consisted of 96 trials and lasted approximately 5.3 min.

Throughout the study day, when participants were not being scanned or participating in the offline EEG session, participants were able to leave the lab and carry out their normal daily activities but were instructed not to nap, consume caffeine, or exercise during this time. These breaks occurred from approximately 11:45 a.m. to 12:45 p.m. and 3:30 p.m. to 4:45 p.m.

fMRI Data

fMRI Data Acquisition

MRI data were acquired on a 3 T Siemens Skyra scanner using a 64-channel head coil at the Princeton Neuroscience Institute's Scully Center for the Neuroscience of Mind and Behavior. Functional scans used a T2*-weighted multiband EPI sequence (repetition time [TR] = 1500 msec, echo time [TE] = 40 msec, voxel size = 1.5 mm isotropic, flip angle = 64°, multiband factor = 6, 72 slices manually aligned to the anterior commissure [AC] - posterior commissure [PC] line, i.e., top-of-AC, bottom-of-PC). These slices comprised a partial volume fully covering the occipital and temporal lobes. For fieldmap correction, two spin-echo field map volumes (TR = 10330 msec, TE = 68 msec) were acquired in opposite phase encoding directions. We collected the following anatomical scans: three whole-brain T1-weighted (T1w) MPRAGE images (one collected during each fMRI scan session; TR = 2300 msec, TE = 2.98 msec, voxel size = 1 mm isotropic, flip angle = 9°, 176 slices, Generalized Autocalibrating Partially Parallel

Acquisitions [GRAPPA] acceleration factor = 2), one T2-weighted turbo spin-echo (TSE) image (acquired at the end of fMRI Session 1; TR = 11390 msec, TE = 90 msec, voxel size = 0.44 × 0.44 × 1.5 mm, flip angle = 150°, 54 slices acquired perpendicular to the long axis of the hippocampus, distance factor = 20%), and three coplanar T1 FLASH images (one acquired during each fMRI scan session), but the FLASH images were ultimately not used in our preprocessing or analysis pipeline.

fMRI Data Preprocessing

Data were preprocessed using fMRIPrep 1.2.3 (Esteban et al., 2018, 2019), which is based on Nipype 1.1.6-dev (Gorgolewski et al., 2011, 2018). Many internal operations of fMRIPrep use Nilearn 0.4.2 (Abraham et al., 2014), mostly within the functional processing workflow.

Anatomical data preprocessing. A total of three T1w images were included within the input BIDS data set (from the prestudy, Session 1, and Session 2 scans). All of them were corrected for intensity nonuniformity (INU) using *N4BiasFieldCorrection* (ANTs 2.2.0; Tustison et al., 2010). A T1w reference map was computed after registration of three T1w images (after INU correction) using *mri_robust_template* (FreeSurfer 6.0.1; Reuter, Rosas, & Fischl, 2010). The T1w reference was then skull-stripped using *antsBrainExtraction.sh* (ANTs 2.2.0), using OASIS as the target template. Brain surfaces were reconstructed using *recon-all* (FreeSurfer 6.0.1; Dale, Fischl, & Sereno, 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray matter of Mindboggle (Klein et al., 2017). Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template Version 2009c (RRID:SCR_008796; Fonov, Evans, McKinstry, Almli, & Collins, 2009) was performed through nonlinear registration with *antsRegistration* (ANTs 2.2.0; Avants, Epstein, Grossman, & Gee, 2008), using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid, white matter, and gray matter was performed on the brain-extracted T1w using *fast* (FSL 5.0.9; Zhang, Brady, & Smith, 2001).

Functional data preprocessing. For each of the 18 BOLD runs per participant (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. A deformation field to correct for susceptibility distortions was estimated based on two EPI references with opposing phase-encoding directions, using *3dQwarp* (AFNI 20160207; Cox & Hyde, 1997). Based on the estimated susceptibility distortion, an unwarped BOLD reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using *bbregister* (FreeSurfer), which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using *mcflirt* (FSL 5.0.9; Jenkinson, Bannister, Brady, & Smith, 2002). BOLD runs were slice-time corrected using *3dTshift* from AFNI 20160207 (RRID:SCR_005927; Cox & Hyde, 1997). The BOLD time series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head motion and susceptibility distortions. Gridded (volumetric) resamplings were performed using *antsApplyTransforms* (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964).

After preprocessing with fMRIPrep, the first nine volumes and last five volumes of each functional scan were discarded. Then, all functional scans were additionally high-pass filtered (1/128 Hz cutoff) and *z*-scored using Nilearn before further analysis.

ROIs

We were specifically interested in measuring neural differentiation in the hippocampus (Fernandez, Jiang, Wang, Choi, & Wagner, 2023; Wammes, Norman, & Turk-Browne, 2022; Molitor, Sherrill, Morton, Miller, & Preston, 2021; Wanjia, Favila, Kim, Molitor, & Kuhl, 2021; Dimsdale-Zucker, Ritchey, Ekstrom, Yonelinas, & Ranganath, 2018; Zeithamova, Gelman, Frank, & Preston, 2018; Chanales, Oza, Favila, & Kuhl, 2017; Hulbert & Norman, 2015; Schlichting, Mumford, & Preston, 2015). Within the hippocampus, we expected to observe differentiation in the left CA2/3/DG subfield, which was the locus of representational change in the Kim et al. (2017) study. We also preregistered right and bilateral CA2/3/DG and left, right, and bilateral CA1 as other ROIs. Neural activity is sparser in CA2/3/DG compared with CA1, making it more difficult for competing memories to come to mind strongly (Good-Smith et al., 2017; Schapiro, Turk-Browne, Botvinick, & Norman, 2017; West, Slomianka, & Gundersen, 1991; Barnes, McNaughton, Mizumori, Leonard, & Lin, 1990). According to the NMPH, lower levels of memory activation in CA2/3/DG should bias this region toward showing differentiation, whereas higher levels of activation in CA1 should bias this region toward showing integration (Ritvo et al., 2019, 2024). In line with this, most fMRI studies of differentiation that have looked at hippocampal subfields have found that differentiation effects tend to be localized in CA2/3/DG rather than in CA1 (e.g., Wammes et al., 2022; Wanjia et al., 2021; Molitor et al., 2021; Dimsdale-Zucker et al., 2018; but see Zheng, Gao, McAvan, Isham, & Ekstrom, 2021). Thus, based on both theoretical grounds (relating to

the NMPH) and empirical precedent, there are strong reasons to expect that differentiation effects in our study will be more readily observed in CA2/3/DG than in CA1.

Hippocampal subfields were defined using the Automated Segmentation of Hippocampal Subfields (ASHS) toolbox (Yushkevich et al., 2015) and a database of manual MTL segmentations from a separate set of participants (Aly & Turk-Browne, 2016a, 2016b). Each participant's fMRIPrep-preprocessed T1w template and their raw T2w TSE image were submitted as input to ASHS. The resulting segmentations were used to make masks for the CA1, CA2/3, DG, and combined CA2/3/DG subfields in both hemispheres. These masks were then transformed and resampled to match the functional data.

Measuring Neural Differentiation

We followed the same fMRI data analysis procedure as Kim et al. (2017). The goal of this analysis was to measure how much the neural representation of B items moved away from the original representation of their A pairmate and compare these pattern similarity values for violation and nonviolation pairs. We used the prelearning snapshot of A as the baseline for representational change to avoid confounds due to item frequency that could be introduced by using the postlearning snapshot of A since A items in the violation condition were presented two more times than A items in the nonviolation condition.

For each pair, we computed the Pearson correlation between the prelearning snapshot of A and postlearning snapshot of B. Pre- and postlearning snapshots were defined as the spatial pattern of activity elicited by each item in a particular ROI at the peak of the hemodynamic response (4.5 sec after image onset). We transformed Pearson's r to Fisher's z, computed the average pattern similarity for pairs within the violation and nonviolation task conditions, and then calculated the difference of violation minus nonviolation conditions. We refer to this difference score as the neural differentiation score; negative values indicate decreased pattern similarity (i.e., less neural overlap or more differentiation) in the violation compared with nonviolation condition.

To test if neural differentiation effects are item-specific (i.e., "Does B become more distinct from A specifically, not just generally more distinct from other items?"), we performed a randomization analysis. For each participant, we shuffled the pair assignments of A and B 1000 times within each task condition. For each shuffle, we recalculated the average pattern similarity for violation and nonviolation (shuffled) pairs and then computed the violation minus nonviolation neural differentiation score. If differentiation is item-specific, the original neural differentiation score should be more negative than the shuffled distribution. This was quantified by computing, for each participant, the z score of the true neural differentiation score relative to the mean and *SD* of the null distribution of 1000 shuffled differentiation scores.

Relating Prediction to Differentiation

The goal of this analysis is to examine how differentiation of A and B items relates to the amount of B activation during the two violation events, when scene A was followed by faces X and Y instead of the expected B pairmate. This analysis was only performed for the violation condition. To measure B prediction on violation trials, we calculated the Pearson correlation between the prelearning snapshot of B and the pattern of activity evoked by the X and Y items. Specifically, for each pair, we correlated Bpre/X and Bpre/Y, then Fisher z-transformed the resulting correlation coefficients, and averaged these two values to provide a single estimate of B prediction for each pair. Across pairs, we calculated the correlation of the B prediction score with the Apre/Bpost neural pattern similarity score, resulting in one Pearson's r value for each subject, which was transformed to Fisher's z for statistical analysis at the group level.

Measuring Integration of X and Y Faces with B Scenes

We ran a control analysis to rule out the alternative explanation that postlearning B snapshots appear less similar to prelearning A snapshots in the violation condition because the postlearning B snapshots include additional noise from faces X and Y (Greve, Abdulrahman, & Henson, 2018). To measure the amount of B-X-Y integration for each B item in the violation condition, we calculated the Pearson correlation between the postlearning snapshot of B and the pattern of activity evoked by the corresponding X and Y faces during their postlearning snapshot phase. Specifically, we correlated Bpost/Xpost and Bpost/Ypost, then transformed these values to Fisher's z, and averaged them to arrive at a single value of B-X-Y integration. Across all violation pairs within each subject, we correlated this B-X-Y integration score with the corresponding Apre/Bpost neural pattern similarity score, resulting in one Pearson's r value for each subject, which was transformed to Fisher's z for statistical analysis at the group level.

EEG Data Processing

EEG Data Acquisition

EEG data were acquired using Ag/AgCl active electrodes (Biosemi Active Two) from 64 scalp EEG locations. We followed standard polysomnography procedures and additionally recorded data from two EOG locations (LOC and ROC), two submental EMG locations, two electrocardiogram locations, and two mastoids. Data were sampled at 512 Hz.

EEG Data Preprocessing and Sleep Scoring

Although we recorded EEG data during Wake sessions, those data were only used to ensure wakefulness and were

not further analyzed. Nap EEG data were preprocessed using BrainVision Analyzer 2.0 (BrainProducts). For sleep scoring purposes, channels used for scoring (LOC, ROC, C3, C4, O1, O2) were re-referenced to the average signal of the left and right mastoid; in cases where one mastoid electrode became detached, a single mastoid was used ($n = 3$). The data were bandpass filtered between 0.3 and 35 Hz with a 60-Hz notch filter and down-sampled to 256 Hz. One EMG channel was subtracted from the other to create one bipolar EMG channel. The EMG channel was bandpass filtered between 10 and 70 Hz. Data were visually scored for sleep stages in 30-sec epochs following standard criteria (Berry et al., 2012) using the Hume 1.0.4 toolbox for MATLAB (https://github.com/jsaletin/hume). Note that sleep records for two participants were unable to be scored for sleep staging due to an excessively noisy signal ($n = 1$, NREM group) and the EEG amplifier battery dying in the middle of the nap session ($n = 1$, REM group). These data sets were not included in analyses that depend on knowing the precise amount of time spent in sleep stages, but they were able to be included in group-level analyses since the experimenter was still able to determine if their nap contained REM sleep or not.

For further analysis of sleep EEG data, artifacts (large movements; arousals; and rare, large deflections in single channels) during sleep were visually identified and rejected in 5-sec chunks. Problematic channels were interpolated. Sleep spindles were detected during Stage 2 and Stage 3 using a wavelet-based algorithm (Warby et al., 2014; Wamsley et al., 2012). Spindle densities were calculated by dividing the number of discrete spindle events by time spent in the corresponding sleep stage. We report spindle data from one exemplar centroparietal electrode (CPz).

Statistical Analyses

Data analysis was performed in RStudio using R 4.2.1. Accuracy on the behavioral cover task was not normally distributed. Therefore, we used the WRS2 package in R to perform robust ANOVA on 20% trimmed means (Mair & Wilcox, 2020). Specifically, we used the *t1way* function to test for overall group differences in accuracy and the *bwtrim* function to compute a two-way mixed model ANOVA for analyses including both within- and between-subject factors.

We used independent samples t tests to test for differences between the two nap groups on sleep architecture measures. When the data within either group were not normally distributed, we report a Wilcoxon rank sum test instead of the independent samples t test. If the assumption of homogeneity of variances was violated, we report Welch's t test.

To directly test our main hypotheses, we ran planned contrasts to examine differences between groups. The first contrast compared the REM group to the two groups that did not have REM sleep (contrast weights REM: −1, NREM: 0.5, Wake: 0.5). Next, we further asked if the NREM group was different from the wake group (contrast weights REM: 0, NREM: −1, Wake: 1). Since we ran the planned contrasts in six individual ROIs, the Bonferroni adjusted alpha level for each set of tests was .05/6 = .008.

Contrasts that suggested the REM group was different from the Wake and NREM groups were followed up with one-sample t tests against zero to test whether neural measures were significantly positive or negative in each group, with a Bonferonni-adjusted alpha level of .05/3 = .017 ($k = 3$ for three groups). Since we preregistered the direction of the expected effects in the REM group, we report one-tailed p values in the REM group (our preregistration did not specifically state we would use one-tailed tests but we did preregister the direction of effects).

For the randomization analysis, we computed the z score of the true neural differentiation score relative to the mean and *SD* of the null distribution of 1000 shuffled differentiation scores within each subject and tested the reliability of these z scores across participants with a one-sample t test against zero.

We used a two-way repeated-measures ANOVA to analyze the reward phase data with both Task Condition (violation/nonviolation) and Repetition as within-subject repeated measures. We used mixed-model ANOVAs for analyses including both within- and between-subject factors; for example, to test for differences in the Decision score we included Task Condition (violation/nonviolation) as the within-subject factor and Group (Wake/NREM/REM) as the between-subject factor.

Pearson correlations examined the relationship between neural measures and sleep, and between neural measures and behavior.

RESULTS

Behavioral Cover Task

In Session 1, participants completed six runs of incidental encoding while performing a subcategory judgment task ("Is the scene indoor or outdoor?" or "Is the face male or female?"). In Session 2, participants completed two runs of the same subcategory judgment task to obtain post-learning snapshots of scenes and faces.

Session 1 task performance was highly accurate (overall mean = 0.95, *SD* = 0.04) and did not differ between the three groups (Wake, NREM, REM; $F(2, 25.02) = 0.51$, $p = .61$). There was no interaction between Trial Type (prelearning, A–B pairings, violation events, restudy events) and Group ($F(6, 27.62 = 0.55$, $p = .77$; Figure S1A). Session 2 postlearning snapshot performance was also highly accurate (overall mean = 0.97, *SD* = 0.02) and did not differ between the three groups ($F(2, 27.37 = 0.84$, $p = .44$; Figure S1B).

Nap Sleep Architecture

Nap sleep architecture variables are summarized in Table 1 and Figure S2. By design, the REM group had significantly

Table 1. Sleep Architecture

	NREM Group	*REM Group*
Sleep Variable	*Mean (SD)*	*Mean (SD)*
TIB (min)***	65.20 (23.51)	113.86 (10.72)
TST (min)***	49.27 (9.41)	89.91 (13.63)
Sleep efficiency	79.95 (15.90)	79.52 (13.40)
Minutes		
NREM***	49.16 (9.47)	73.20 (10.92)
REM***	0.11 (0.34)	16.70 (9.49)
Stage 1*	6.66 (6.21)	9.18 (5.21)
Stage 2***	22.32 (9.61)	42.07 (11.91)
Stage 3	20.18 (14.33)	21.95 (9.86)
WASO**	8.75 (16.06)	14.05 (13.42)
Percent		
NREM***	99.75 (0.75)	82.10 (9.29)
REM***	0.25 (0.75)	17.90 (9.29)
Stage 1	13.74 (12.84)	10.61 (6.88)
Stage 2	46.09 (18.64)	46.59 (10.17)
Stage 3*	39.92 (25.37)	24.90 (10.94)
WASO	17.22 (32.57)	17.49 (19.24)
Sleep spindles		
Stage 2 spindle number***	94.64 (39.73)	180.73 (50.72)
Stage 3 spindle number	108.00 (73.24)	112.95 (51.81)
Stage 2 spindle density	4.31 (0.65)	4.34 (0.54)
Stage 3 spindle density	5.14 (0.86)	5.19 (0.61)

TIB = time in bed; Sleep efficiency = TST/TIB; NREM = time in Stages 1, 2, and 3 combined. The values reported here are from $n = 22$ in each group. Asterisks indicate a significant difference between groups.

* $p < .05$; ** $p < .01$; *** $p < .001$.

greater TST and minutes of REM sleep than the NREM group (Wilcoxon rank sum test, both ps $< .001$). The REM group also had significantly more minutes of Stage 1 sleep (Wilcoxon rank sum test, $p = .03$), Stage 2 sleep (Wilcoxon rank sum test, $p < .001$), and wake after sleep onset (WASO; Wilcoxon rank sum test, $p = .004$). However, the groups did not differ on these values as a percentage of TST (Stage 1 percent: Wilcoxon rank sum test, $p = .93$; Stage 2 percent: Welch's t test, $t(32.5) = -0.11$, $p = .91$; WASO percent: Wilcoxon rank sum test, $p = .06$). There was no difference between groups in minutes of Stage 3 ($t(42) = -0.48$, $p = .64$), but the NREM group had a greater percentage of time spent in Stage 3 than the REM group (Welch's t test, $t(28.5) = 2.55$, $p = .02$). The groups did not differ in sleep efficiency (TST/time spent in bed; Wilcoxon rank sum test, $p = .99$).

In keeping with the overall greater amount of time spent in Stage 2 sleep, the REM group also had significantly more discrete Stage 2 spindle events (Wilcoxon rank sum test, $p < .001$), but there was no difference in Stage 2 spindle density (spindles/min) between groups ($t(42) = -0.16$, $p = .87$). The groups also did not differ in the number or density of sleep spindles during Stage 3 (Wilcoxon rank sum test, spindle number: $p = .62$; spindle density: $p = .22$).

Neural Differentiation as a Function of Sleep

To examine how much the neural representation of B items moved away from their A pairmates, we used the same procedure as Kim et al. (2017). We correlated voxel patterns for the prelearning snapshot of A and postlearning snapshot of B (preA/postB) for all AB pairs within each task condition (violation and nonviolation) for each participant (Figure 1B). Next, we computed the average pattern similarity value for each task condition (Figure S3; note that numerically smaller pattern similarity values indicate less neural overlap or more neural differentiation). Since our general hypothesis was that pattern similarity should be lower for violation compared with nonviolation pairs (Kim et al., 2017), we computed the difference of the violation and nonviolation conditions in each subject. We will refer to this difference as the neural differentiation score; negative values indicate more violation-related neural differentiation.

Our main preregistered hypothesis was that violation-related neural differentiation should be greatest in the REM group, specifically in the left CA2/3/DG subfield of the hippocampus. As the most direct test of this hypothesis, we ran a planned contrast (contrast weights: REM: −1, NREM: 0.5, Wake: 0.5) to determine if having REM sleep, compared with not having REM sleep (i.e., NREM or Wake), resulted in a significantly more negative neural differentiation score. This contrast was not significant in left CA2/3/DG ($t(66) = 0.027$, $p = .98$, $d = 0.007$; Figure 2). We ran the same contrast in our five remaining ROIs (right and bilateral CA2/3/DG and left, right, and bilateral CA1). The predicted pattern of results was present in right CA2/3/DG ($t(66) = 2.19$, $p = .03$, $d = 0.54$), although it was not significant after correcting for multiple comparisons (adjusted alpha = .008). The contrast was not significant in bilateral CA2/3/DG ($t(66) = 1.12$, $p = .26$, $d = 0.28$) or any CA1 ROI (left CA1: $t(66) = -0.35$, $p = .73$, $d = -0.08$; right CA1: $t(66) = -0.83$, $p = .41$, $d = -0.20$; bilateral CA1: $t(66) = -0.81$, $p = .42$, $d = -0.20$).

We then ran post hoc tests to investigate the differentiation finding in right CA2/3/DG. Within the REM group, the neural differentiation score in right CA2/3/DG was significantly negative ($t(22) = -2.44$, $p = .01$, $d =$

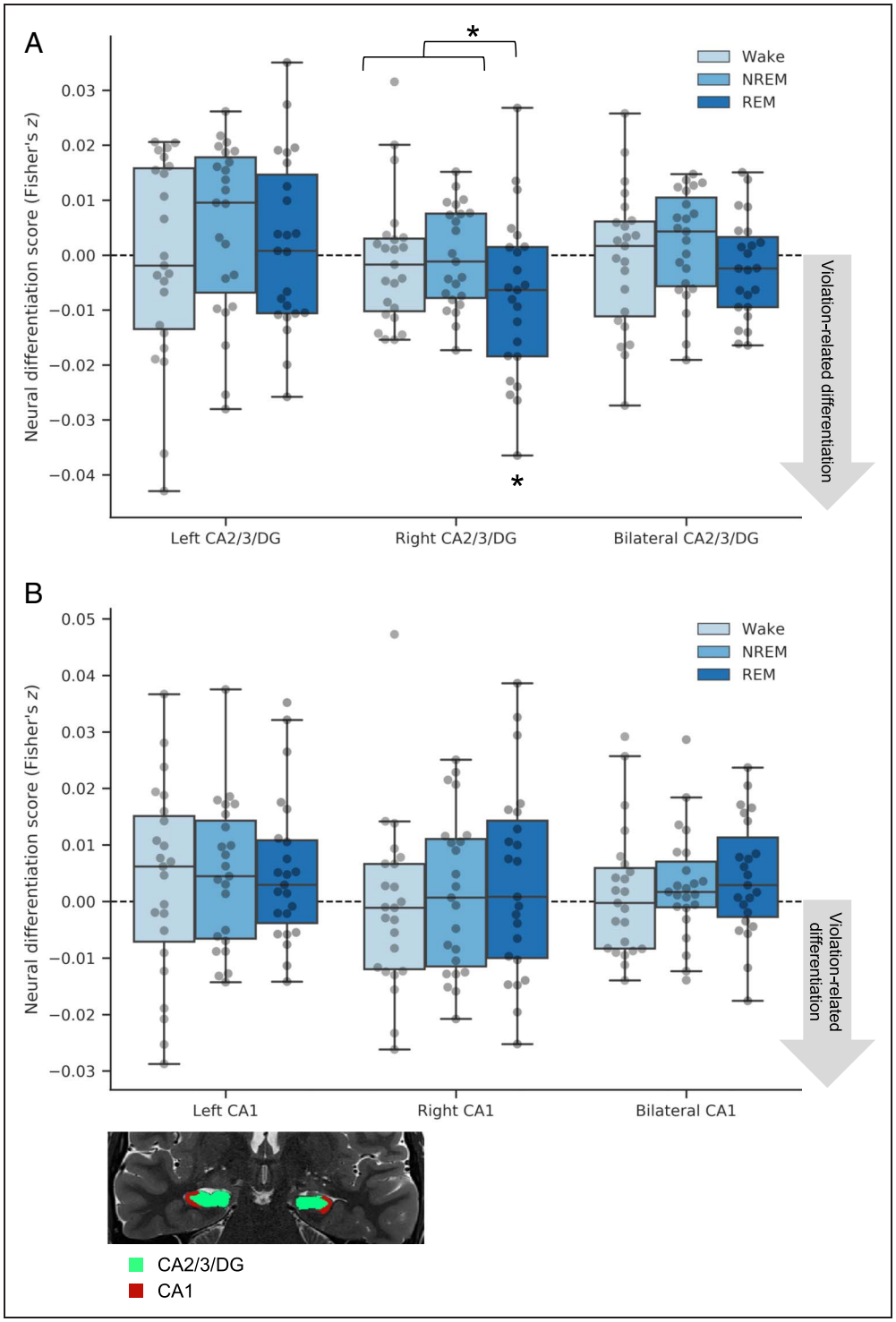

Figure 2. Neural differentiation. Neural differentiation scores were calculated as the difference in preA/postB pattern similarity for the violation minus nonviolation task conditions, in each sleep group in (A) CA2/3/DG and (B) CA1. A planned contrast in our six ROIs revealed more violation-related neural differentiation in the REM group than the Wake and NREM groups in right CA2/3/DG (p = .03, uncorrected for multiple comparisons). Within the REM group, post hoc tests showed that the neural differentiation score in right CA2/3/DG was significantly negative (p = .01, one-tailed) and reliably item-specific based on a randomization analysis (p = .02). Brain image shows segmented ROIs for one subject overlaid on their high-resolution T2w anatomical image. n = 23 in each group; *p < .05.

−0.51, one-tailed), indicating that the A and B items making up violation condition pairs became less similar to each other than nonviolation condition pairmates. As illustrated in Figure S3, which plots the violation and nonviolation conditions separately, the greater differentiation effect in the REM group appears to be driven by nonviolation pattern similarity values being higher in the REM group than the other groups, rather than violation pattern similarity values being lower in the REM group; indeed, an exploratory contrast revealed that nonviolation pattern similarity values were significantly higher with REM than without REM ($t(66)$ = 2.12, p = .04, d = 0.52). This result suggests that, in the REM group, there may have been some integration of A and B items in the nonviolation condition that did not occur in the violation condition.

Given the significant effect of REM sleep in right CA2/3/DG, we next asked if this effect was item-specific. In other words, does B become more distinct from its specific A pairmate, not just generally more distinct from other items? We employed a randomization analysis where we shuffled the pair assignments of A and B 1000 times within each task condition and recalculated the neural differentiation score. If differentiation is item-specific, the actual neural differentiation score should be more negative (negative values indicate violation-related differentiation) than the shuffled distribution. Within each participant, we computed the z score of the observed neural differentiation score relative to the mean and *SD* of 1000 shuffled scores and tested the reliability of these z scores across participants with a one-sample t test against zero. This confirmed that violation-related neural differentiation in right CA2/3/DG in the REM group was item-specific ($t(22) = -2.62$, $p = .02$, $d = -0.54$).

We did not have a specific prediction about whether the NREM and Wake groups would differ from each other. NREM sleep alone could yield some marginal benefit for neural differentiation compared with time spent awake. However, contrasting those two groups (contrast weights: REM: 0, NREM: −1, Wake: 1) revealed no difference in any subfield (left CA2/3/DG: $t(66) = -1.22, p = .23, d = -0.30$; right CA2/3/DG: $t(66) = -0.22, p = .83, d = -0.05$; bilateral CA2/3/DG: $t(66) = -0.88, p = .38, d = -0.22$; left CA1: $t(66) = -0.30, p = .76, d = -0.07$; right CA1: $t(66) = -0.52, p = .60, d = -0.13$; bilateral CA1: $t(66) = -0.47, p = .64, d = -0.12$).

Finally, we ran an exploratory analysis separating the differentiation scores in CA2/3 and DG subfields (Figure S4; raw pattern similarity values for each task condition can be found in Figure S5). The contrast testing the effect of REM versus NREM/Wake showed the predicted pattern in right DG only (right DG: $t(66) = 2.27, p = .03, d = 0.56$, not significant after correcting for multiple comparisons; left DG: $t(66) = -0.37, p = .72, d = -0.09$; bilateral DG: $t(66) = 1.03, p = .31, d = 0.25$; left CA2/3: $t(66) = 0.45, p = .66, d = 0.11$; right CA2/3: $t(66) = -0.14, p = .89, d = -0.03$; bilateral CA2/3: $t(66) = 0.37, p = .71, d = 0.09$), suggesting that violation-related neural differentiation in the REM group was primarily driven by the right DG. A post hoc test showed that the neural differentiation score in the REM group in right DG was significantly negative ($t(22) = -2.31, p = .015, d = -0.48$, one-tailed), and a randomization analysis further confirmed that this violation-related neural differentiation was item-specific ($t(22) = -2.48$, $p = .02$, $d = -0.52$). Clearer differentiation in DG than CA2/3 is consistent with prior findings from the small number of studies to isolate DG (Wammes et al., 2022). Mirroring the result from the combined CA2/3/DG ROI, an exploratory contrast revealed that nonviolation pattern similarity values in right DG were significantly higher with REM than without REM ($t(66) = 2.77, p = .007, d = 0.68$), suggesting there may have been some integration of A and B items in the nonviolation condition.

Our preferred explanation for why postlearning B snapshots become less similar to prelearning A snapshots is that the representation of B moves away from A (relinquishing shared features and acquiring new features). An alternative explanation for the differentiation effects we observed is that the postlearning B snapshots include additional "noise" from the X and Y faces (Greve et al., 2018). According to this account, during violation events, when a participant sees A followed by X or Y while simultaneously reactivating B, the X and Y faces are bound to (or integrated with) both A and B, resulting in decreased similarity between prelearning A (which does not incorporate this noise from X and Y) and postlearning B (which does). If true, then we might expect a negative relationship between preA/postB pattern similarity and a measure of B-X-Y integration. That is, the more that X-Y are integrated into B, the lower preA/postB pattern similarity should be. To measure B-X-Y integration, we calculated the Pearson correlation between the postlearning snapshot of B and the postlearning snapshots of X and Y. We then correlated this B-X-Y integration score with preA/postB pattern similarity across pairs within subject. This correlation was not reliably different from zero across subjects in the REM group in right CA2/3/DG ($t(22) = 0.93, p = .36, d = 0.19$), and thus, X-Y "noise" does not appear to be driving the observed violation-related neural differentiation effect.

Relationship between Prediction and Differentiation

As part of the hypothesized differentiation mechanism, we predicted that the degree to which A and B items differentiated would be related to the amount of B activation during violation trials (as observed by Kim et al., 2017). That is, after three instances of B following A in the stimulus sequence, how much did B come to mind when a participant was presented with A followed by a face? To measure this "B prediction" on violation trials (two violation trials per pair, with a unique face presented during each violation, which we refer to as faces X and Y), we calculated the Pearson correlation between the prelearning snapshot of B and the pattern of activity evoked by the X and Y violation events, then averaged these values, yielding one B prediction score per pair in the violation condition (Figure 1B). Across pairs, we computed the correlation of this prediction score with the preA/postB pattern similarity values within each subject. If greater B prediction is related to more neural differentiation, this should yield a negative correlation value (i.e., more B prediction, less preA/postB pattern similarity). We then analyzed these within-subject correlation values across subjects at the group level.

Again, we hypothesized that the prediction–differentiation relationship would be stronger in the REM group compared with the NREM and Wake groups, specifically in left CA2/3/DG. However, the planned

contrast testing for a difference between the REM group and the two other groups (contrast weights: REM: −1, NREM: 0.5, Wake: 0.5) revealed no significant difference in left CA2/3/DG ($t(66) = 1.26$, $p = .21$, $d = 0.31$), or any other ROI (right CA2/3/DG: $t(66) = 0.92$, $p = .36$, $d = 0.23$; bilateral CA2/3/DG: $t(66) = 1.34$, $p = .18$, $d = 0.33$; left CA1: $t(66) = 0.63$, $p = .53$, $d = 0.16$; right CA1: $t(66) = 0.41$, $p = .68$, $d = 0.10$; bilateral CA1: $t(66) = 0.59$, $p = .56$, $d = 0.14$; Figure S6).

An exploratory analysis testing the same contrast in CA2/3 and DG separately showed the predicted pattern in bilateral DG ($t(66) = 2.04$, $p = .046$, $d = 0.50$; Figure S7), with higher levels of B prediction associated with decreased A–B pattern similarity in the REM group (one-sample t test, $t(22) = -2.06$, $p = .03$, $d = -0.43$, one-tailed); neither of these tests remained significant following correction for multiple comparisons. The contrast between REM and Wake/NREM was not significant in the remaining ROIs (left CA2/3: $t(66) = 1.80$, $p = .08$, $d = 0.44$; right CA2/3: $t(66) = -0.17$, $p = .86$, $d = -0.04$; bilateral CA2/3: $t(66) = 0.44$, $p = .66$, $d = 0.11$; left DG: $t(66) = 1.18$, $p = .24$, $d = 0.29$; right DG: $t(66) = 1.32$, $p = .19$, $d = 0.32$). Nonetheless, the REM group numerically showed the predicted negative relationship between B prediction and A–B pattern similarity in all of these ROIs; in addition to bilateral DG, the negative relationship in the REM group also passed traditional significance levels in left CA2/3 ($t(22) = -2.50$, $p = .01$, $d = -0.52$, one-tailed). However, the group-level contrast in left CA2/3 did not show that the REM group was different than NREM/Wake, so we interpret this with caution.

Relationship between Sleep and Differentiation

We hypothesized that the degree to which items differentiated would not only depend on the amount of prediction during violation trials but also on the composition of the intervening sleep period. Specifically, we expected that greater amounts of REM sleep would be associated with more violation-related differentiation (numerically, this should yield a negative correlation between REM duration and the differentiation score). We found that, within the REM group, minutes of REM sleep was not significantly correlated with the neural differentiation score in the preregistered ROIs (left CA2/3/DG: $r = -.12$, $p = .59$; right CA2/3/DG: $r = .10$, $p = .66$; bilateral CA2/3/DG: $r = .01$, $p = .96$; left CA1: $r = .36$, $p = .10$; right CA1: $r = -.02$, $p = .94$; bilateral CA1: $r = .20$, $p = .38$), nor was REM duration correlated with neural differentiation in the exploratory DG ROI that showed a difference between REM and Wake/NREM (left DG: $r = -.30$, $p = .17$; right DG: $r = .14$, $p = .52$; bilateral DG: $r = -.09$, $p = .70$).

We also ran exploratory analyses exploring the relationship between other sleep variables (TST, minutes of Stage 1, Stage 2, Stage 3, spindle density) and the neural differentiation score in CA2/3/DG and CA1 in all nap participants (NREM and REM groups combined). Stage 3 spindle density was positively correlated with the differentiation score in left and bilateral CA1 (left: $r = .36$, $p = .02$; bilateral: $r = .34$, $p = .03$), indicating that greater spindle densities were associated with less violation-related differentiation in these regions, but these correlations were not significant after correcting for multiple comparisons. There were no other significant correlations (all remaining ps > .08; see Table S1). Given the lack of significant correlations between sleep features and differentiation, we did not proceed with the additional multiple linear regression analyses proposed in the preregistration.

Relationship between Neural Differentiation and Behavioral Reward Learning

At the end of the main experiment, participants completed a secondary task with the aim of detecting a behavioral consequence of neural differentiation. We based our task on Wimmer and Shohamy (2012), who showed that reward values can spread across implicitly associated memories. Participants explicitly learned to associate either a reward or neutral outcome with the A scene from each pair; B scenes were never directly associated with the reward or neutral outcome during the learning phase. The rationale of the task is that, if the neural representations of A and B are more overlapping (i.e., less differentiated), then the learned reward associations should generalize more to the B scene pairmate. Therefore, we predicted less generalization to B scenes for pairs that experience more differentiation, specifically violation pairs in the REM group. However, we did not find any evidence of differences in generalization due to task condition or group assignment (see Supplemental Materials for detailed methods and results).

DISCUSSION

Prior work from our lab found hippocampal differentiation when neural activity was measured approximately 24 hr after prediction violations (Kim et al., 2017). The current study provides a preregistered test of the hypothesis that REM sleep was a critical ingredient driving this representational change. In support of our hypothesis, we found greater differentiation in the REM group than the Wake and NREM groups in the same CA2/3/DG ROI as the previous study. Our effect was lateralized to the right, rather than the left, hemisphere and was only significant at an uncorrected threshold. An exploratory analysis found that this effect was driven by right DG. We also hypothesized that the degree of neural differentiation would be correlated with the amount of prediction during violation events and more so in the REM group than the Wake and NREM groups. This pattern was reliable at an uncorrected threshold in an exploratory bilateral DG ROI. Although our findings do not fully align with the preregistration, they are nevertheless consistent with the hypothesis that REM sleep is important for hippocampal neural

differentiation. Indeed, all of the differentiation effects we obtained were larger in the REM group than the other two groups or apparent within the REM group alone and were localized to the predicted region of the hippocampus (CA2/3/DG). More work is needed, but these results provide provisional evidence for our working hypothesis that neural dynamics during REM sleep allow spreading activation within the hippocampus to identify and coactivate memories marked for representational change (Norman et al., 2005). Below, we discuss notable aspects of the present work that can guide future research.

Our study builds on prior work that has examined representational change across periods of consolidation (i.e., time, including time spent asleep). Tompary and Davachi (2017) measured neural pattern similarity immediately postlearning and after 1 week. After a week of consolidation (but not immediately after encoding), unique objects associated with the same scene became integrated in medial prefrontal cortex (mPFC) and posterior hippocampus relative to objects associated with different scenes (see Ezzyat, Inhoff, & Davachi, 2018, for an example of consolidation-related differentiation in mPFC). Within the hippocampus, another study found that representations for object–word pairs that were studied at the same time (i.e., within a list) and subsequently remembered became more differentiated in the anterior compared with the posterior hippocampus (Cowan et al., 2021); this reorganization of memory representations along the long axis of the hippocampus only emerged after an overnight delay (also see Dandolo & Schwabe, 2018). Although sleep may have contributed to the effects reported in these studies, sleep was not included as an experimental variable. Cowan et al. (2020) reported correlational evidence linking sleep to the restructuring of memory representations, finding that fast spindle density during overnight sleep was associated with greater pattern similarity in ventromedial prefrontal cortex (vmPFC) for object–word pairs learned before sleep and that this relationship was mediated by anterior hippocampal–vmPFC functional connectivity. Here, we manipulated the content of sleep (NREM only vs. combined NREM and REM) and included a quiet wake control group; this design allowed us to isolate the role of REM sleep. Interestingly, violation-related neural differentiation in the REM group appears to have been driven by increased (relative to NREM and Wake) pattern similarity in the nonviolation condition. These results are compatible with an interpretation where integration occurs—and is facilitated by REM sleep—in the nonviolation condition, but violation events prevent this integration from taking place. We note that integration in the nonviolation condition (where A was always followed by B) is consistent with principles of the NMPH; every time participants saw A, there was also strong coactivation of B, which should lead to strengthening of the connections between A and B according to the NMPH (Ritvo et al., 2019). Our results thus contribute to an emerging picture in which learning demands determine if memories should be integrated or differentiated (see Antony & Schechtman, 2023 for the related consolidation trajectory hypothesis), and REM sleep drives those representational changes during a period of consolidation (Sterpenich et al., 2014).

The idea that REM sleep (as opposed to NREM sleep alone) was necessary for hippocampal differentiation appears to be at odds with other results showing that NREM sleep, as well as its associated electrophysiological signatures (i.e., sleep spindles and slow oscillations), is the key sleep stage for hippocampal memory consolidation (Cairney, Guttesen, El Marj, & Staresina, 2018; Mednick et al., 2013; Rasch, Büchel, Gais, & Born, 2007; Marshall, Helgadóttir, Mölle, & Born, 2006). One reason why it might be challenging to appreciate the effects of REM sleep on memory consolidation is that it may not cause a simple increase or decrease in memory strength. Rather, as shown here, offline learning during REM may be important for shaping how memory representations relate to one another, either by pushing memories closer together or pulling them apart. The behavioral measures typically used to quantify declarative memory consolidation (e.g., memory for paired associates) may not be sensitive enough to detect these representational changes from REM sleep. Even in the current study, we did not find group-level differences in performance on our postsleep behavioral measure (the reward learning task). Other studies that found a behavioral effect of REM sleep used measures sensitive to the structure of memory representations, such as interference between competing memories (McDevitt et al., 2015; Baran, Wilson, & Spencer, 2010) or multi-item integration (Abdou et al., 2024; Batterink, Oudiette, Reber, & Paller, 2014; Cai, Mednick, Harrison, Kanady, & Mednick, 2009). Our study highlights how experiments can be designed to target the aspects of memory hypothesized to be REM-dependent.

Even though NREM sleep alone was not sufficient to bring about differentiation, it is still possible that NREM sleep contributed to the effects seen in the REM group. We did not test a REM-only condition, since it is not biologically normal to enter REM sleep without some amount of preceding NREM sleep—as such, we cannot claim that REM alone is sufficient. Rather, the combination of the two in sequence (i.e., a full sleep cycle) may be critical for bringing about these representational changes, as suggested by the "sequential hypothesis" (Diekelmann & Born, 2010; Giuditta et al., 1995). An alternative explanation of our results is that the differentiation we observed in the REM group was due to this group having more Stage 2 sleep compared with the NREM group. However, we did not find any evidence of dose-dependent effects; no sleep variable, including time spent in Stage 2 sleep, was significantly correlated with differentiation, suggesting that—even if the NREM group had more Stage 2—it would not have contributed to more differentiation. Having some amount of NREM and REM sleep, rather than high absolute amounts of either, might be the critical factor. There was

also no dose-dependent effect of REM sleep. Simply correlating the amount of REM sleep with the overall amount of differentiation across participants is a relatively blunt measure and may not be sensitive to the likely nuanced ways the brain prioritizes information for replay during REM sleep, particularly in a shorter episode of sleep, such as a daytime nap. For example, memory processing in the early part of REM sleep may especially benefit weakly learned information (Schapiro, McDevitt, et al., 2017). Future research, including research using computational models that can manipulate the amount and structure of sleep (e.g., Singh et al., 2022), could address these open questions.

Whereas in our study differentiation required offline time spent asleep, other studies have found neural differentiation during or immediately following learning (Wanjia et al., 2021; Schlichting et al., 2015). For example, Wanjia et al. (2021) demonstrated that hippocampal differentiation abruptly happened at the "inflection point" in the learning process when participants showed clear behavioral evidence of discriminating similar cues. Our working hypothesis is that, if competition is not resolved during wake (e.g., there is not enough training to reach the inflection point), then offline learning during sleep can supplement and pull the competing representations further apart. An interesting extension of Wanjia et al. (2021) would be to test if pairs that do not reach the inflection point during learning, and presumably have not yet differentiated, show hippocampal differentiation following REM sleep without additional training. Future research is also needed to establish whether sleep is necessary to (1) stabilize already-differentiated hippocampal representations and (2) differentiate representations in long-term neocortical storage sites so that the behavioral memory benefits of differentiation (e.g., reduced interference; Favila et al., 2016) are upheld over the long-term (Favila & Aly, 2025; Ezzyat et al., 2018).

Although we obtained support for our hypothesis that REM sleep contributes to differentiation in CA2/3/DG (and DG alone), our effect was found in the right hemisphere instead of the left hemisphere, which was our main, preregistered ROI. Arguably, we erred in our preregistration by overfitting to a single prior study (Kim et al., 2017). Reviewing the literature more broadly, there is strong support for differentiation in CA2/3/DG (or DG alone), and more than in CA1 (e.g., Wammes et al., 2022; Molitor et al., 2021; Wanjia et al., 2021; Schapiro, Kustner, & Turk-Browne, 2012). However, there is less support for differentiation being localized to the left hemisphere (for another example, see Bein & Davachi, 2024). In fact, many studies did not separate or analyze their results by hemisphere (Wammes et al., 2022; Molitor et al., 2021; Wanjia et al., 2021; Dimsdale-Zucker et al., 2018). Schapiro et al. (2012) did show a stronger numerical pattern of differentiation in right versus left CA2/3/DG. In retrospect, it would have been wise to treat results from these additional studies as a "prior" when filing the preregistration—with this hindsight, we might have been more agnostic about hemisphere and predicted an effect in CA2/3/DG with bilateral, left, and right ROIs.

Apart from these shortcomings in the preregistration, there were also some important differences between the present study and Kim et al. (2017) that may have contributed to discrepancies in the results. Kim et al. (2017) included a full night of sleep between prediction violations and the final measurement of representational change, whereas our study used a nap that was approximately 90 min in duration in the REM group. Using a daytime nap as our sleep intervention is both a strength and limitation, allowing us to (1) minimize time-of-day effects, (2) implement a strong waking control without stressful sleep deprivation, and (3) experimentally manipulate the absence/presence of REM sleep (McDevitt et al., 2015; Mednick et al., 2003). However, a daytime nap is a shorter period of sleep, and it is possible that more representational change would occur over multiple sleep cycles in a full night. Another limitation is that we did not reach our target sample size of $n = 102$ (34 participants per group) because of pandemic-related data collection interruptions. Our final sample size of $n = 69$ (23 participants per group) may be underpowered to detect the full range of effects (for reference, Kim et al., 2017, reported their effects in one group of 32 participants). In particular, the REM group showed the expected direction of the prediction–differentiation relationship in right CA2/3/DG and right DG (the two ROIs that showed overall differentiation in the REM group), but the difference between groups was not statistically significant.

In conclusion, our findings provide suggestive evidence for how memory processing during REM sleep can complement the plasticity processes initiated during waking experience; when memories are identified as targets for representational change (here, as a result of prediction errors that occur during wake), some of the learning required to implement these changes may occur later, during REM sleep. These results provide initial, converging support for a link between REM sleep and representational change of individual memories in the hippocampus.

Acknowledgments

We thank Monika Schönauer and James Antony for helpful discussions and comments on an earlier version of this paper.

Corresponding authors: Elizabeth A. McDevitt, Princeton Neuroscience Institute, Princeton University, Princeton, NJ, e-mail: emcdevitt@princeton.edu and Kenneth A. Norman, Department of Psychology, Princeton University, Princeton, NJ, e-mail: knorman@princeton.edu.

Data Availability Statement

Neuroimaging data are available on OpenNeuro.org: https://openneuro.org/datasets/ds006576. Supplemental Material can be accessed on this article's homepage: https://doi.org/10.1162/JOCN.a.82.

Code Availability

Code related to this project is on GitHub: https://github.com/PrincetonCompMemLab/rem_viodiff_code.git.

Author Contributions

Elizabeth A. McDevitt: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Visualization; Writing—Original draft. Ghootae Kim: Conceptualization; Methodology; Resources; Writing—Review & editing. Nicholas B. Turk-Browne: Conceptualization; Funding acquisition; Methodology; Writing—Review & editing. Kenneth A. Norman: Conceptualization; Funding acquisition; Methodology; Supervision; Writing—Review & editing.

Funding Information

This work was supported by the National Institutes of Health (https://dx.doi.org/10.13039/100000025) (NIMH R01-MH069456 to K. A. N. and N. T.-B. and NIMH K99-MH126154 to E. A. M.). Funding for the acquisition of the data was provided in part by the Regina and John Scully '66 Center for the Neuroscience of Mind and Behavior.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this paper report its proportions of citations by gender category to be: M/M = .429; W/M = .254; M/W = .143; W/W = .175.

REFERENCES

Abdellahi, M. E. A., Koopman, A. C. M., Treder, M. S., & Lewis, P. A. (2023). Targeted memory reactivation in human REM sleep elicits detectable reactivation. *eLife*, *12*, e84324. https://doi.org/10.7554/eLife.84324, PubMed: 37350572

Abdou, K., Nomoto, M., Aly, M. H., Ibrahim, A. Z., Choko, K., Okubo-Suzuki, R., et al. (2024). Prefrontal coding of learned and inferred knowledge during REM and NREM sleep. *Nature Communications*, *15*, 4566. https://doi.org/10.1038/s41467-024-48816-x, PubMed: 38914541

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., et al. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*, 14. https://doi.org/10.3389/fninf.2014.00014, PubMed: 24600388

Adan, A., & Almirall, H. (1991). Horne & Östberg morningness-eveningness questionnaire: A reduced scale. *Personality and Individual Differences*, *12*, 241–253. https://doi.org/10.1016/0191-8869(91)90110-w

Aly, M., & Turk-Browne, N. B. (2016a). Attention stabilizes representations in the human hippocampus. *Cerebral Cortex*, *26*, 783–796. https://doi.org/10.1093/cercor/bhv041, PubMed: 25766839

Aly, M., & Turk-Browne, N. B. (2016b). Attention promotes episodic encoding by stabilizing hippocampal representations. *Proceedings of the National Academy of Sciences, U.S.A.*, *113*, E420–E429. https://doi.org/10.1073/pnas.1518931113, PubMed: 26755611

Antony, J. W., & Schechtman, E. (2023). Reap while you sleep: Consolidation of memories differs by how they were sown. *Hippocampus*, *33*, 922–935. https://doi.org/10.1002/hipo.23526, PubMed: 36973868

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*, 26–41. https://doi.org/10.1016/j.media.2007.06.004, PubMed: 17659998

Baran, B., Wilson, J., & Spencer, R. M. C. (2010). REM-dependent repair of competitive memory suppression. *Experimental Brain Research*, *203*, 471–477. https://doi.org/10.1007/s00221-010-2242-2, PubMed: 20401652

Barnes, C. A., McNaughton, B. L., Mizumori, S. J., Leonard, B. W., & Lin, L.-H. (1990). Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing. *Progress in Brain Research*, *83*, 287–300. https://doi.org/10.1016/s0079-6123(08)61257-1, PubMed: 2392566

Batterink, L. J., Oudiette, D., Reber, P. J., & Paller, K. A. (2014). Sleep facilitates learning a new linguistic rule. *Neuropsychologia*, *65*, 169–179. https://doi.org/10.1016/j.neuropsychologia.2014.10.024, PubMed: 25447376

Bein, O., & Davachi, L. (2024). Event integration and temporal differentiation: How hierarchical knowledge emerges in hippocampal subfields through learning. *Journal of Neuroscience*, *44*, e0627232023. https://doi.org/10.1523/JNEUROSCI.0627-23.2023, PubMed: 38129134

Berry, R. B., Brooks, R., Gamaldo, C. E., Harding, S. M., Marcus, C., Vaughn, B. V., et al. (2012). *The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications* (Version 2.4). American Academy of Sleep Medicine.

Cai, D. J., Mednick, S. A., Harrison, E. M., Kanady, J. C., & Mednick, S. C. (2009). REM, not incubation, improves creativity by priming associative networks. *Proceedings of the National Academy of Sciences, U.S.A.*, *106*, 10130–10134. https://doi.org/10.1073/pnas.0900271106, PubMed: 19506253

Cairney, S. A., Guttesen, A. Á. V., El Marj, N., & Staresina, B. P. (2018). Memory consolidation is linked to spindle-mediated information processing during sleep. *Current Biology*, *28*, 948–954. https://doi.org/10.1016/j.cub.2018.01.087, PubMed: 29526594

Cantero, J. L., Atienza, M., Stickgold, R., Kahana, M. J., Madsen, J. R., & Kocsis, B. (2003). Sleep-dependent θ oscillations in the human hippocampus and neocortex. *Journal of Neuroscience*, *23*, 10897–10903. https://doi.org/10.1523/JNEUROSCI.23-34-10897.2003, PubMed: 14645485

Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among spatial memories triggers repulsion of hippocampal representations. *Current Biology*, *27*, 2307–2317. https://doi.org/10.1016/j.cub.2017.06.057, PubMed: 28736170

Cowan, E., Liu, A., Henin, S., Kothare, S., Devinsky, O., & Davachi, L. (2020). Sleep spindles promote the restructuring of memory representations in ventromedial prefrontal cortex through enhanced hippocampal-cortical functional connectivity. *Journal of Neuroscience*, *40*, 1909–1919. https://doi.org/10.1523/JNEUROSCI.1946-19.2020, PubMed: 31959699

Cowan, E. T., Liu, A. A., Henin, S., Kothare, S., Devinsky, O., & Davachi, L. (2021). Time-dependent transformations of memory representations differ along the long axis of the hippocampus. *Learning & Memory*, *28*, 329–340. https://doi.org/10.1101/lm.053438.121, PubMed: 34400534

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, *10*, 171–178. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L, PubMed: 9430344

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*, *9*, 179–194. https://doi.org/10.1006/nimg.1998.0395, PubMed: 9931268

Dandolo, L. C., & Schwabe, L. (2018). Time-dependent memory transformation along the hippocampal anterior-posterior axis. *Nature Communications*, *9*, 1205. https://doi.org/10.1038/s41467-018-03661-7, PubMed: 29572516

Detre, G. J., Natarajan, A., Gershman, S. J., & Norman, K. A. (2013). Moderate levels of activation lead to forgetting in the think/no-think paradigm. *Neuropsychologia*, *51*, 2371–2388. https://doi.org/10.1016/j.neuropsychologia.2013.02.017, PubMed: 23499722

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, *11*, 114–126. https://doi.org/10.1038/nrn2762, PubMed: 20046194

Dimsdale-Zucker, H. R., Ritchey, M., Ekstrom, A. D., Yonelinas, A. P., & Ranganath, C. (2018). CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal subfields. *Nature Communications*, *9*, 294. https://doi.org/10.1038/s41467-017-02752-1, PubMed: 29348512

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., et al. (2018). fMRIPrep 1.2.3. *Software*. https://doi.org/10.5281/zenodo.852659

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., et al. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*, 111–116. https://doi.org/10.1038/s41592-018-0235-4, PubMed: 30532080

Ezzyat, Y., Inhoff, M. C., & Davachi, L. (2018). Differentiation of human medial prefrontal cortex activity underlies long-term resistance to forgetting in memory. *Journal of Neuroscience*, *38*, 10244–10254. https://doi.org/10.1523/JNEUROSCI.2290-17.2018, PubMed: 30012697

Favila, S. E., & Aly, M. (2025). Hippocampal mechanisms resolve competition in memory and perception. *bioRxiv*. https://doi.org/10.1101/2023.10.09.561548, PubMed: 37873400

Favila, S. E., Chanales, A. J. H., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature Communications*, *7*, 11066. https://doi.org/10.1038/ncomms11066, PubMed: 27925613

Fernandez, C., Jiang, J., Wang, S.-F., Choi, H. L., & Wagner, A. D. (2023). Representational integration and differentiation in the human hippocampus following goal-directed navigation. *eLife*, *12*, e80281. https://doi.org/10.7554/elife.80281, PubMed: 36786678

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage*, *47(Suppl. 1)*, S102. https://doi.org/10.1016/S1053-8119(09)70884-5

Giuditta, A., Ambrosini, M. V., Montagnese, P., Mandile, P., Cotugno, M., Zucconi, G. G., et al. (1995). The sequential hypothesis of the function of sleep. *Behavioural Brain Research*, *69*, 157–166. https://doi.org/10.1016/0166-4328(95)00012-i, PubMed: 7546307

GoodSmith, D., Chen, X., Wang, C., Kim, S. H., Song, H., Burgalossi, A., et al. (2017). Spatial representations of granule cells and mossy cells of the dentate gyrus. *Neuron*, *93*, 677–690. https://doi.org/10.1016/j.neuron.2016.12.026, PubMed: 28132828

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., et al. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, *5*, 13. https://doi.org/10.3389/fninf.2011.00013, PubMed: 21897815

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P., et al. (2018). Nipype. *Software*. https://doi.org/10.5281/zenodo.596855

Greve, A., Abdulrahman, H., & Henson, R. N. (2018). Neural differentiation of incorrectly predicted memories. *Frontiers in Human Neuroscience*, *12*, 278. https://doi.org/10.3389/fnhum.2018.00278, PubMed: 30050419

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, *48*, 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060, PubMed: 19573611

Guerreiro, I. C., & Clopath, C. (2024). Memory's gatekeeper: The role of PFC in the encoding of congruent events. *Proceedings of the National Academy of Sciences, U.S.A.*, *121*, e2403648121. https://doi.org/10.1073/pnas.2403648121, PubMed: 39018188

Hulbert, J. C., & Norman, K. A. (2015). Neural differentiation tracks improved recall of competing memories following interleaved study and retrieval practice. *Cerebral Cortex*, *25*, 3994–4008. https://doi.org/10.1093/cercor/bhu284, PubMed: 25477369

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*, 825–841. https://doi.org/10.1006/nimg.2002.1132, PubMed: 12377157

Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience*, *10*, 100–107. https://doi.org/10.1038/nn1825, PubMed: 17173043

Johns, M. W. (1992). Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep*, *15*, 376–381. https://doi.org/10.1093/sleep/15.4.376, PubMed: 1519015

Kim, G., Lewis-Peacock, J. A., Norman, K. A., & Turk-Browne, N. B. (2014). Pruning of memories by context-based prediction error. *Proceedings of the National Academy of Sciences, U.S.A.*, *111*, 8997–9002. https://doi.org/10.1073/pnas.1319438111, PubMed: 24889631

Kim, G., Norman, K. A., & Turk-Browne, N. B. (2017). Neural differentiation of incorrectly predicted memories. *Journal of Neuroscience*, *37*, 2022–2031. https://doi.org/10.1523/JNEUROSCI.3272-16.2017, PubMed: 28115478

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., et al. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*, e1005350. https://doi.org/10.1371/journal.pcbi.1005350, PubMed: 28231282

Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience*, *22*, 1598–1610. https://doi.org/10.1038/s41593-019-0467-3, PubMed: 31451802

Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, *1*, 76–85. https://doi.org/10.1137/0701007

Louie, K., & Wilson, M. A. (2001). Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron*, *29*, 145–156. https://doi.org/10.1016/S0896-6273(01)00186-6, PubMed: 11182087

Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, *52*, 464–488. https://doi.org/10.3758/s13428-019-01246-w, PubMed: 31152384

Marshall, L., Helgadóttir, H., Mölle, M., & Born, J. (2006). Boosting slow oscillations during sleep potentiates memory. *Nature*, *444*, 610–613. https://doi.org/10.1038/nature05278, PubMed: 17086200

McDevitt, E. A., Duggan, K. A., & Mednick, S. C. (2015). REM sleep rescues learning from interference. *Neurobiology of Learning and Memory*, *122*, 51–62. https://doi.org/10.1016/j.nlm.2014.11.015, PubMed: 25498222

McDevitt, E. A., Rowe, K. M., Brady, M., Duggan, K. A., & Mednick, S. C. (2014). The benefit of offline sleep and wake for novel object recognition. *Experimental Brain Research*, *232*, 1487–1496. https://doi.org/10.1007/s00221-014-3830-3, PubMed: 24504196

Mednick, S., Nakayama, K., & Stickgold, R. (2003). Sleep-dependent learning: A nap is as good as a night. *Nature Neuroscience*, *6*, 697–698. https://doi.org/10.1038/nn1078, PubMed: 12819785

Mednick, S. C., McDevitt, E. A., Walsh, J. K., Wamsley, E., Paulus, M., Kanady, J. C., et al. (2013). The critical role of sleep spindles in hippocampal-dependent memory: A pharmacology study. *Journal of Neuroscience*, *33*, 4494–4504. https://doi.org/10.1523/JNEUROSCI.3127-12.2013, PubMed: 23467365

Molitor, R. J., Sherrill, K. R., Morton, N. W., Miller, A. A., & Preston, A. R. (2021). Memory reactivation during learning simultaneously promotes dentate gyrus/$CA_{2,3}$ pattern differentiation and CA_1 memory integration. *Journal of Neuroscience*, *41*, 726–738. https://doi.org/10.1523/JNEUROSCI.0394-20.2020, PubMed: 33239402

Nemeth, D., Gerbier, E., Born, J., Rickard, T., Diekelmann, S., Fogel, S., et al. (2024). Optimizing the methodology of human sleep and memory research. *Nature Reviews Psychology*, *3*, 123–137. https://doi.org/10.1038/s44159-023-00262-0

Newman, E. L., & Norman, K. A. (2010). Moderate excitation leads to weakening of perceptual representations. *Cerebral Cortex*, *20*, 2760–2770. https://doi.org/10.1093/cercor/bhq021, PubMed: 20181622

Norman, K. A., Newman, E. L., & Perotte, A. J. (2005). Methods for reducing interference in the complementary learning systems model: Oscillating inhibition and autonomous memory rehearsal. *Neural Networks*, *18*, 1212–1228. https://doi.org/10.1016/j.neunet.2005.08.010, PubMed: 16260116

Rasch, B., Büchel, C., Gais, S., & Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science*, *315*, 1426–1429. https://doi.org/10.1126/science.1138581, PubMed: 17347444

Reuter, M., Rosas, H. D., & Fischl, B. (2010). Highly accurate inverse consistent registration: A robust approach. *Neuroimage*, *53*, 1181–1196. https://doi.org/10.1016/j.neuroimage.2010.07.020, PubMed: 20637289

Ritvo, V. J. H., Nguyen, A., Turk-Browne, N. B., & Norman, K. A. (2024). A neural network model of differentiation and integration of competing memories. *eLife*, *12*, RP88608. https://doi.org/10.7554/eLife.88608, PubMed: 39319791

Ritvo, V. J. H., Turk-Browne, N. B., & Norman, K. A. (2019). Nonmonotonic plasticity: How memory retrieval drives learning. *Trends in Cognitive Sciences*, *23*, 726–742. https://doi.org/10.1016/j.tics.2019.06.007, PubMed: 31358438

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, *22*, 1622–1627. https://doi.org/10.1016/j.cub.2012.06.056, PubMed: 22885059

Schapiro, A. C., McDevitt, E. A., Chen, L., Norman, K. A., Mednick, S. C., & Rogers, T. T. (2017). Sleep benefits memory for semantic category structure while preserving exemplar-specific information. *Scientific Reports*, *7*, 14869. https://doi.org/10.1038/s41598-017-12884-5, PubMed: 29093451

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *372*, 20160049. https://doi.org/10.1098/rstb.2016.0049, PubMed: 27872368

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, *6*, 8151. https://doi.org/10.1038/ncomms9151, PubMed: 26303198

Schönauer, M., Alizadeh, S., Jamalabadi, H., Abraham, A., Pawlizki, A., & Gais, S. (2017). Decoding material-specific memory reprocessing during sleep in humans. *Nature Communications*, *8*, 15404. https://doi.org/10.1038/ncomms15404, PubMed: 28513589

Singh, D., Norman, K. A., & Schapiro, A. C. (2022). A model of autonomous interactions between hippocampus and neocortex driving sleep-dependent memory consolidation. *Proceedings of the National Academy of Sciences, U.S.A.*, *119*, e2123432119. https://doi.org/10.1073/pnas.2123432119, PubMed: 36279437

Sterpenich, V., Schmidt, C., Albouy, G., Matarazzo, L., Vanhaudenhuyse, A., Boveroux, P., et al. (2014). Memory reactivation during rapid eye movement sleep promotes its generalization and integration in cortical stores. *Sleep*, *37*, 1061–1075. https://doi.org/10.5665/sleep.3762, PubMed: 24882901

Tompary, A., & Davachi, L. (2017). Consolidation promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex. *Neuron*, *96*, 228–241. https://doi.org/10.1016/j.neuron.2017.09.005, PubMed: 28957671

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., et al. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*, 1310–1320. https://doi.org/10.1109/TMI.2010.2046908, PubMed: 20378467

Wammes, J., Norman, K. A., & Turk-Browne, N. (2022). Increasing stimulus similarity drives nonmonotonic representational change in hippocampus. *eLife*, *11*, e68344. https://doi.org/10.7554/elife.68344, PubMed: 34989336

Wamsley, E. J., Tucker, M. A., Shinn, A. K., Ono, K. E., McKinley, S. K., Ely, A. V., et al. (2012). Reduced sleep spindles and spindle coherence in schizophrenia: Mechanisms of impaired memory consolidation? *Biological Psychiatry*, *71*, 154–161. https://doi.org/10.1016/j.biopsych.2011.08.008, PubMed: 21967958

Wanjia, G., Favila, S. E., Kim, G., Molitor, R. J., & Kuhl, B. A. (2021). Abrupt hippocampal remapping signals resolution of memory interference. *Nature Communications*, *12*, 4816. https://doi.org/10.1038/s41467-021-25126-0, PubMed: 34376652

Warby, S. C., Wendt, S. L., Welinder, P., Munk, E. G., Carrillo, O., Sorensen, H. B., et al. (2014). Sleep-spindle detection: Crowdsourcing and evaluating performance of experts, non-experts and automated methods. *Nature Methods*, *11*, 385–392. https://doi.org/10.1038/nmeth.2855, PubMed: 24562424

West, M. J., Slomianka, L., & Gundersen, H. J. G. (1991). Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator. *Anatomical Record*, *231*, 482–497. https://doi.org/10.1002/ar.1092310411, PubMed: 1793176

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, *265*, 676–679. https://doi.org/10.1126/science.8036517, PubMed: 8036517

Wimmer, G. E., & Shohamy, D. (2012). Preference by association: How memory mechanisms in the hippocampus bias decisions. *Science*, *338*, 270–273. https://doi.org/10.1126/science.1223252, PubMed: 23066083

Yushkevich, P. A., Pluta, J. B., Wang, H., Xie, L., Ding, S.-L., Gertje, E. C., et al. (2015). Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Human Brain Mapping*, *36*, 258–287. https://doi.org/10.1002/hbm.22627, PubMed: 25181316

Zeithamova, D., Gelman, B. D., Frank, L., & Preston, A. R. (2018). Abstract representation of prospective reward in the hippocampus. *Journal of Neuroscience*, *38*, 10093–10101. https://doi.org/10.1523/JNEUROSCI.0719-18.2018, PubMed: 30282732

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*, 45–57. https://doi.org/10.1109/42.906424, PubMed: 11293691

Zheng, L., Gao, Z., McAvan, A. S., Isham, E. A., & Ekstrom, A. D. (2021). Partially overlapping spatial environments trigger reinstatement in hippocampus and schema representations in prefrontal cortex. *Nature Communications*, *12*, 6231. https://doi.org/10.1038/s41467-021-26560-w, PubMed: 34711830