

Festschrift in Honor of Dr. Robert Stickgold

Original Article

Sleep and retrieval practice both strengthen and distort story recollection

Dan Denis^{1,*}, Carissa DiPietro², R. Nathan Spreng³, Daniel L. Schacter⁴, Robert Stickgold^{5,6} and Jessica D. Payne²

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

²Department of Psychology, University of Notre Dame, Notre Dame, IN, USA,

³Montreal Neurological Institute, McGill University, Montreal, QC, Canada,

⁴Department of Psychology, Harvard University, Cambridge, MA, USA,

⁵Department of Psychiatry, Beth Israel Deaconess Medical Center, Boston, MA, USA and

⁶Department of Psychiatry, Harvard Medical School, Boston, MA, USA

*Corresponding author. Dan Denis, Department of Psychology, PS/B/213 Psychology Building, University of York, Heslington, York YO10 5DD, UK. Email: dan.denis@york.ac.uk

Abstract

Over time, memories lose episodic detail and become distorted, a process with serious ramifications for eyewitness identification. What are the processes contributing to such transformations over time? We investigated the roles of post-learning sleep and retrieval practice in memory accuracy and distortion, using a naturalistic story recollection task. Undergraduate students listened to a recording of the “War of the Ghosts,” a Native American folktale, and were assigned to either a sleep or wake delay group, and either a retrieval practice or listen-only study condition. We found higher accuracy after sleep compared to wake in the listen-only condition, but not in the retrieval practice condition. This effect was driven by participants in the wake, retrieval practice condition showing superior memory compared to the wake, listen-only condition. A similar pattern was found for memory distortion, with both sleep and retrieval practice being associated with more inferences of nonpresented, but story-related information, compared to the wake, listen-only condition. These findings suggest both sleep and retrieval practice contribute to narrative memory stabilization and distortion.

Key words: sleep; retrieval practice; memory; testing; consolidation

Statement of Significance

A critical question for our understanding of human memory is how memories are transformed over time. Our long-term memories tend to lack episodic detail, with the “gist” of the episode being preserved. We also integrate related memories, allowing us to generalize new information based on our past experiences. How does this evolution of memory occur? Using the classic “War of the Ghosts” story, we found that overnight sleep (compared to a day awake) both strengthens memory, while also increasing the number of inferences made about plausible, but nonpresented information. Strikingly, the same effects were achieved through post-learning retrieval practice. In conclusion, our results suggest two different pathways (post-learning sleep and retrieval practice) that promote long-term memory transformation.

Human memories are inevitably transformed and distorted over time, causing us to remember past events quite differently than the true original experience. Episodic detail is lost, with the “gist” of the narrative being preserved [1, 2]. Distortions of memory also appear, with memory for different but overlapping events combined, or forgotten details being replaced or fabricated in order

to fit a comprehensive narrative in tune with our cultural schemas [3–6]. Such transformative processes confer a flexibility and generalizability to memory that may at times be more advantageous than a veridical reinstatement of the original experience. However, this flexibility comes at the cost of memory accuracy, which has important real-world implications when considering

the reliability of eyewitness testimony [5, 7]. As such, understanding when and how these transformations of memory occur remains a central question in the cognitive psychology of human memory.

In this *Festschrift*, we honor and celebrate Dr. Robert Stickgold's major contributions to the fields of sleep and circadian science. Of greatest relevance to the present article is Dr. Stickgold's work in the domain of sleep-related memory consolidation. Going beyond the classic definition of memory consolidation that refers simply to the process of memories becoming more resistant to interference, Dr. Stickgold argued for an expansion of the term consolidation to encompass a much broader set of processes, particularly pertaining to consolidation during sleep [8–11]. Coining the phrase *memory evolution* [10], empirical work from Dr. Stickgold's laboratory has shown how sleep actively improves [12–14] and selectively enhances the most relevant aspects of memory [15]. Work from Dr. Stickgold's lab has also shown that sleep-dependent memory evolution fosters the integration of new memory traces into preexisting schemas [16–18], as well as facilitating the generation of new insights and inferences of previously learned information [19, 20]. This body of work has gone on to be replicated and extended by multiple independent labs around the world [21–24].

In the present article, we continue to explore Dr. Stickgold's idea of sleep-dependent memory evolution, through the examination of the learning and consolidation of an unfamiliar story. For this, we turned to another eminent scholar of memory, Dr. Frederik Bartlett and his famous "War of the Ghosts" experiments [3]. The War of the Ghosts is a Native American folktale written in a disjointed, non-Western style unfamiliar to most college students [3]. In his original experiments, Bartlett had participants read the story twice before testing their memory for the story over varying time delays. He found that over time, not only did forgetting occur, but memory also became distorted and infused with misinformation. Such distortions in recall became magnified over time and with repeated recall attempts, as participants held onto the schema or gist while replacing story details to keep a complete narrative. The importations could be quite dramatic, especially if the delay between recall attempts was long [3]. Since Bartlett's initial study, his results have been replicated under stricter experimental conditions [25].

It is striking that many of the memory transformations observed by Dr. Bartlett map onto the types of memory evolution shown by Dr. Stickgold to occur during sleep, particularly the generalization of new information (the War of the Ghosts story) into something that made sense based on existing schemas (growing up in a Western culture more than 100 years after War of the Ghosts was written). Thus, guided by Dr. Stickgold's ideas regarding memory evolution during sleep [10, 26, 27], we set out to investigate whether the time-dependent changes shown in the recall of the War of the Ghosts are heightened across a night of sleep compared to a day spent awake.

An important debate in the sleep and memory literature is what the exact boundary conditions are that allow for a sleep benefit to occur [28]. In other words, which aspects of memory consolidation are truly *sleep-dependent*. Numerous studies have demonstrated that simply attempting to recall a memory shortly after the original event can bolster memory strength, a phenomenon known as the testing effect [29–31]. Several recent studies have suggested memory consolidation processes once thought of as being dependent on sleep can also be achieved through post-learning retrieval practice. For example, using a

paired associates task, researchers found that when participants engaged in retrieval practice, this eliminates the typical benefit of sleep compared to wake [32, 33]. Similar results have also been observed for recall of spatial locations [34].

While these studies suggest that time-dependent memory strengthening is not uniquely dependent on sleep, prior work has not addressed whether other transformative aspects of memory evolution, such as memory integration and inference generation, can also be achieved through retrieval practice. Although evidence of retrieval-induced distortion of memory has been observed [35, 36], the degree and exact nature of these distortions relative to overnight sleep have yet to be examined. The War of the Ghosts story has several advantages as a tool for studying memory evolution, and for examining which aspects of evolution may (or may not) be dependent on sleep. Story recall represents the kind of recollection that people engage in naturalistically, in which memory for the "gist" of an event often replaces veridical recall. It enables us to study not only accurate retrieval of facts, but also the distortions and importations of new information that result from schema-based representations. As such, we had two main research aims:

1. How does sleep and retrieval practice impact memory accuracy? This was tested through examining the number of story elements correctly recalled, as well as the number of intrusions of nonstory elements (i.e. false memories).
2. For inaccurate memories, does sleep and/or retrieval practice primarily contribute to a simple loss of episodic detail, or a distortion of the original story? For memory distortions, we categorized distortions into four separate types, allowing us to carefully examine where exactly any *sleep-dependent* memory evolution may occur.

Methods

Participants

A total of 114 Harvard University undergraduate students ($M_{\text{age}} = 20.1$ years old, range 18–30 years, 62 females) participated in the study and were compensated for their time with either course credit or cash. Because the research design and effects of interest were novel, it was not possible to estimate effect size from previous studies to perform an a priori power analysis. However, we aimed to recruit 30 participants per group, which was considered achievable given time and financial constraints, and was consistent with many sleep and memory studies [15, 37–40]. The participants were native English speakers with normal or corrected-to-normal vision and free from any psychiatric disorders, sleep, drug, or alcohol problems. Participants were not taking any medications affecting the central nervous system. Informed consent was obtained for all participants, and the study was approved by the Harvard University and Beth Israel Deaconess Medical Center Internal Review Boards. The research was performed in accordance with the Declaration of Helsinki.

Materials

A recorded version of the Native American folktale "The War of the Ghosts" served as the to-be-remembered material [3] (Supplementary Table S1). All participants completed a three-night sleep log and the Stanford Sleepiness Scale (SSS; a one-item questionnaire used to assess participant's alertness in the moment [41]) at the start of each session (Supplementary Table

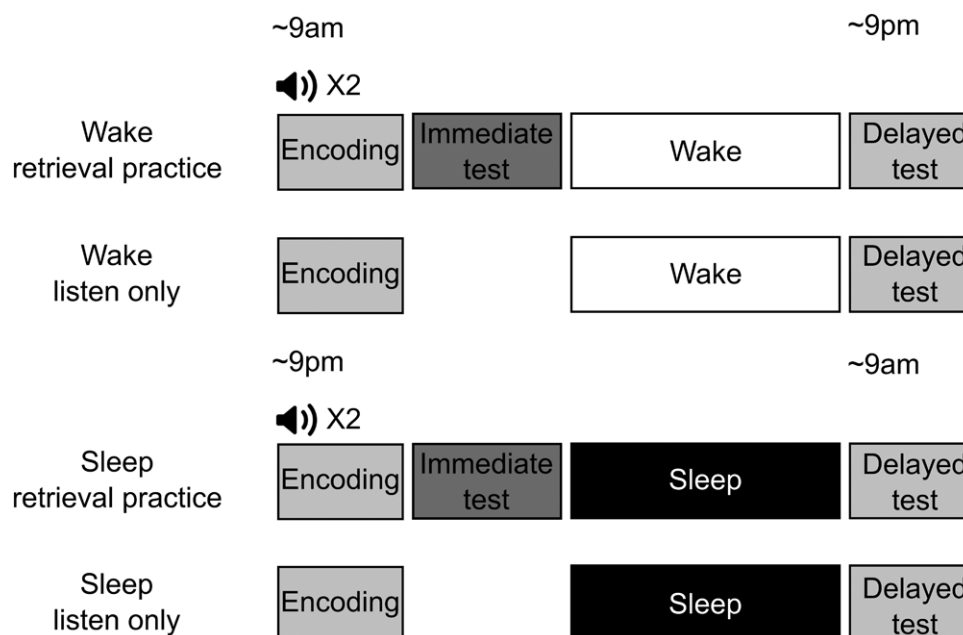


Figure 1. Experimental design. During encoding, participants listened twice to an audio recording of The War of the Ghosts. After a 15-minute distractor task, half of the participants were given a sheet of paper and were asked to recall the story in as much detail as possible. The other half were dismissed after listening to the story. After a 12-hour delay, all participants returned to the lab and were required to recall the story in as much detail as possible. In half of the participants, the delay included a day of wakefulness. In the other half, the delay included a night of sleep.

S2). None of the included participants reported prior experience with the story.

Participants were assigned to one of four experimental groups (Figure 1): (1) a wake + retrieval practice group ($n = 30$); (2) a wake, listen-only group ($n = 29$); (3) a sleep + retrieval practice group ($n = 27$), and (4) a sleep, listen-only group ($n = 28$). The groups did not differ in terms of age ($F(3,110) = 0.27, p = .84, \eta^2 = .007$) or sex ($\chi^2(3) = 0.30, p = .96$), see also [Supplementary Table S2](#). Participants listened to the War of the Ghosts story twice in succession, either in the morning (9 am), or the evening (9 pm) depending on the sleep/wake group assignment (i.e. wake groups listened in the morning, sleep groups listened in the evening). All participants then completed a 15-minute distractor task consisting of simple arithmetic problems. Next, half of the participants were dismissed (i.e. those in the two listen-only groups; wake, listen-only; sleep, listen-only) while the other half of participants were given a blank sheet of notebook paper and were instructed to engage in free recall by recollecting the story as accurately as possible (i.e. those in the two retrieval practice groups; wake + retrieval practice; sleep + retrieval practice). All participants returned 12 hours after their first session for a delayed recall test, following either a day of wakefulness (wake groups) or a night of sleep (sleep groups). For half of the participants, it was their first time recalling the story (listen-only groups), while for the other half, it was their second recall of the story (retrieval practice groups).

During each recall test, participants were encouraged to use the exact same words as heard in the story (veridical recall), and to write down as many facts and events as possible. Participants were given 10 minutes for each recall test and were encouraged to use the full-time.

Scoring

The War of the Ghosts story was subdivided into 42 propositions following previously established methods [25, 42] ([Supplementary](#)

[Table S1](#)). To facilitate identification of incorrect placements (see below), the story was demarcated into three episodes (episode 1 = lines 1–18, episode 2 = lines = 19–28, episodes 3 = lines 29–42). Handwritten transcripts of recalled material were typed into a computer and divided into units of propositions. These transcripts were then scored by two researchers who were blind to experimental condition. Interrater reliability between the two scorers showed a Cohen's kappa of 0.84. As in previous studies using this task, all data reported below are those of scorer 1 [25]. Results were unchanged when data from scorer 2 was used. Sentences from the participant transcripts were matched to the propositions in the story that were most similar in content. We then identified propositions from the recalled material that fit into one of several categories (see [Table 1](#) for an overview and example of each category).

First, propositions were categorized as being either exactly correct (i.e. verbatim) or essentially correct (i.e. verbatim with minor omission that did not result in a loss of detail). These two categories were collapsed together to represent accurate memory [25]. Propositions were categorized as inaccurate if they included at least one major distortion, or if changes to the phrasing of the proposition rendered it incorrect. Inaccurate propositions were then categorized as either containing omission (a loss of detail without any distortion, i.e. detail is omitted, but nothing is added to the proposition) or containing a distortion (if the recalled proposition contains anything that modified or changes the meaning of the proposition as it is recalled). Finally, the type of distortion was identified. *Modification* was defined as a change in the phrasing of a proposition that does not add new information (see *Inference*) and does not change into something more common (see *Normalization*) but nevertheless uses a distinctly different phrasing or involves an accurate summarizing statement. *Inference* involved a change or addition to the proposition such that what was not stated before but was implied or can be inferred from the story is stated explicitly. *Normalization*

Table 1. Rules for scoring each proposition in the story

Proposition	Rule	Example
Accurate	Propositions that were recalled verbatim, or verbatim with minor omission that did not result in a loss of detail	Verbatim: "They escaped to the shore, and hid behind some logs" With minor omission: "They escaped and hid behind some logs"
Omission	Omission of key details of the proposition	"Now canoes came up, and they heard the sound of paddles" changed to "canoes were approaching, and they started to hear noises"
Modification	Contained all elements of the proposition but used distinctly different wording, but the meaning was essentially correct	"I will not go along. I might be killed. My relatives don't know where I have gone" changed to "But my family won't know where I have gone and I might die"
Inference	Wrote an event in a proposition that was not explicitly stated, but could reasonably be assumed to have occurred between two events	Writing that one of the boys got into the canoe, when it is never explicitly stated that he traveled with the warriors
Normalization	Conventionalization of a story element	Recalling that the two young men went on a fishing trip, rather than going down to the river to hunt seals
Incorrect placement	The proposition contained an element from a different portion of the story	Referring to the warriors as ghosts at the beginning of the story, when they are not referred to as ghosts until the middle
Importation	An entirely new element was added to the story	Recalling that the warriors fought with guns (the story only mentions bows and arrows)

was defined as changing an element in the proposition from the actual element to something more in line with that of the target social group (in this case, Harvard University undergraduates) to be the normal experience or conception of the world. An *Incorrect placement* was where a story element is imported from another episode within the story. Finally, an *Importation* occurred when an entirely new element (essentially an intrusion from an unknown source) was added to the story. The categorization of propositions was consistent with prior research [25].

Statistical analysis

All statistical analyses were performed in R [43] using the *rstatix* package [44].

Time of day influences.

We compared immediate memory performance between the two retrieval practice groups via a series of unpaired t-tests, with each of the proposition types entered as a DV.

Baseline sleep influences.

We compared the four groups on baseline subjective sleep metrics (as obtained through sleep diaries) and alertness (as obtained through Stanford sleepiness scale assessments) via a series of 2 (delay group: sleep, wake) $\times$ 2 (study condition: retrieval practice, listen-only) ANOVAs with the following dependent variables: bedtime the night before the encoding session, waking up time the morning of the encoding session, total sleep time the night before encoding, subjective sleep quality the night before encoding, SSS score prior to encoding, and SSS score prior to the delayed test. For the two sleep groups, we also compared differences in bedtime, waking up time, total sleep time, and sleep quality for the night of sleep between encoding and delayed test between the retrieval practice and listen-only group, via a series of unpaired t-tests.

Aim 1: How does sleep and retrieval practice impact memory accuracy?

We examined the roles of sleep and retrieval practice on the number of accurate and false (i.e. importations) propositions at the

delayed test by way of two separate 2 (delay group: sleep, wake) $\times$ 2 (study condition: retrieval practice, listen-only) ANOVAs. A corrected significance level of p value $\times$ 2 tests was used. Post hoc tests were corrected for multiple comparisons as appropriate.

Aim 2a: For inaccurate memories, does sleep and/or retrieval practice primarily contribute to a simple loss of episodic detail, or a distortion of the original story?

We tested how the number of omissions and distortions varied between our factors of interest. We again performed two 2 $\times$ 2 ANOVAs, with the number of omissions and the number of distortions at the delayed test entered as DVs. A corrected significance level of p value $\times$ 2 tests was used. Post hoc tests were corrected for multiple comparisons as appropriate.

Aim 2b: Which kind of memory distortions are most impacted by sleep and retrieval practice?

A series of four 2 $\times$ 2 ANOVAs were run to disentangle the effects of sleep and retrieval practice on each type of distortion made during the delayed test (modification, inference, normalization, incorrect placement). A corrected significance level of p value $\times$ 4 tests was used. Pairwise follow-up tests, corrected for multiple comparisons, were performed as appropriate.

We also sought to quantify the relative effects of sleep and retrieval practice on memory performance. Using the wake, listen-only group as a baseline (i.e. the group with neither sleep nor retrieval practice), we compared the size of improvements gained via either sleep (sleep, listen-only group), retrieval practice (wake, retrieval practice group), or both (sleep, retrieval practice group) by subtracting the average wake, listen-only delayed test score from each participant's score in the other three groups [37, 45]. We then tested whether the size of improvement was significantly different from zero via a series of one-sample t-tests, separately for the three groups. We also directly compared the three groups (sleep, listen-only; wake, retrieval practice; sleep, retrieval practice) on their relative improvement size via ANOVA. Multiple comparison correction was applied as appropriate. To reduce the number of comparisons, we only performed these analyses for proposition categories that were significantly modulated by either

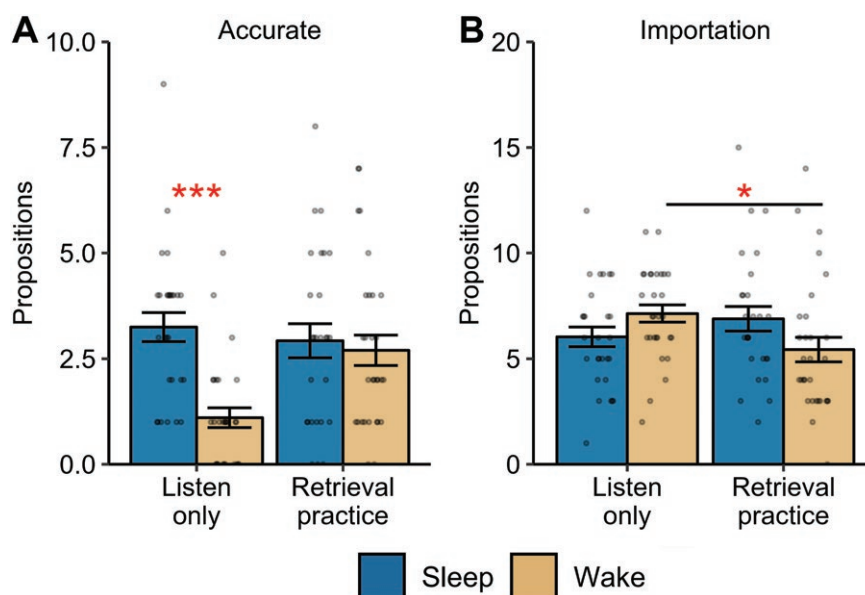


Figure 2. Number of propositions classified as accurate or importation (false memory) at the delayed test. Dots represent individual data points, and error bars display the between-subjects standard error of the mean. *** $p < .001$, * $p < .05$.

delay group or test session in the ANOVA analyses described above under Aims 1 and 2.

Results

Ruling out time of day and baseline sleep influences

First, we checked to see whether time of day affected recall by comparing memory metrics between the sleep and the wake group who engaged in retrieval practice. There were no differences in immediate free recall performance based on delay condition for any of the metrics (all $ps > .08$, all $ds < 0.46$; [Supplementary Table S3](#)). Similarly, there was no difference between the sleep and wake groups in terms of total number of propositions produced at the initial test ($t(54.80) = 0.96$, $p = .34$, $d = 0.26$). At the delayed test, a 2 (delay group: sleep, wake) $\times$ 2 (study condition; retrieval practice, listen-only) ANOVA found a significant main effect of delay on total number of propositions produced ($F(1,110) = 11.82$, $p < .001$, $np^2 = .10$), with more propositions in the sleep groups ($M = 29.3$, $SD = 4.06$) compared to the wake groups ($M = 26.4$, $SD = 5.02$). The main effect of study condition and the interaction were both nonsignificant ($ps > .07$).

Subjective sleep quality and alertness measures can be found in [Supplementary Table S2](#). On the night before encoding, participants in the wake groups reported waking up significantly earlier than participants in the sleep groups ($F(1,109) = 23.72$, $p < .001$, $np^2 = .18$). There was no effect of study condition ($F(1,109) = 2.22$, $p = .14$, $np^2 = .02$), nor was there an interaction ($F(1,109) = 0.42$, $p = .52$, $np^2 = .004$). Despite this, there were no delay group differences in subjective alertness at either session ($ps > .096$), and time of awakening did not correlate with the number of propositions made at either the immediate ($r = 0.10$, $p = .62$), or delayed ($r = -0.002$, $p = .98$) test. No other effects of delay group, study condition, on any interactions, were significant ($ps > .054$).

Effects of sleep and retrieval practice on delayed recall performance

Participants in the sleep groups reported sleeping on average for 424 minutes ($SD = 56$ minutes) during the night between

encoding and the delayed test. There was no difference between the listen-only and retrieval practice study conditions in any of the sleep diary measures (all $ps > .11$).

Number of propositions at the delayed test, as a function of category, is displayed in [Supplementary Table S4](#). As a first analysis, we examined the effect of delay group (sleep, wake) and study condition (retrieval practice, listen-only) on memory accuracy and number of importations. For memory accuracy, we observed a main effect of delay group ($F(1,110) = 12.23$, $p_{\text{corr}} = .001$, $np^2 = .10$), with more accurate propositions in the sleep groups ($M = 3.09$, $SD = 1.95$) compared to the wake groups ($M = 1.91$, $SD = 1.83$). We also observed a significant delay group $\times$ study condition interaction ($F(1,110) = 8.02$, $p_{\text{corr}} = .012$, $np^2 = .068$; [Figure 2A](#)). Follow-up tests revealed that accuracy was significantly higher after sleep compared to wake only for those participants who did not engage in retrieval practice ($t(48) = 5.16$, $p < .001$, $d = 1.37$). When participants engaged in retrieval practice, there was no difference in accuracy between the sleep and wake groups ($t(53.5) = 0.42$, $p = .68$, $d = 0.11$). This effect was driven by higher accuracy in the wake, retrieval practice group compared to the wake, listen-only group ($t(49.7) = 3.72$, $p < .001$, $d = 0.97$), with no differences between the two sleep conditions (sleep, retrieval practice vs sleep, listen-only) ($t(51.5) = 0.61$, $p = .54$, $d = 0.17$). We note that the significant interaction remained when a potential outlier was removed ($F(1,109) = 6.88$, $p_{\text{corr}} = .020$, $np^2 = .059$).

For the importation of new story elements, we again observed a significant delay group $\times$ study condition interaction ($F(1,110) = 6.21$, $p_{\text{corr}} = .028$, $np^2 = .05$; [Figure 2B](#)), driven by a significantly higher rate of importation of nonstory elements in the wake, listen-only group, compared to the wake, retrieval practice group ($t(51.3) = 2.41$, $p = .019$, $d = 0.63$). All other pairwise comparisons were nonsignificant (all $ps > .10$).

Next, we examined the effects of sleep and testing on types of inaccurate memory (i.e. the number of omissions and distortions). There were no significant main effects or interactions between delay group and study condition on the number of omissions or distortions (all $ps > .057$). To explore potential differences within the subtypes of distortion, we examined the four different kinds of distortion separately ([Figure 3](#)). For the

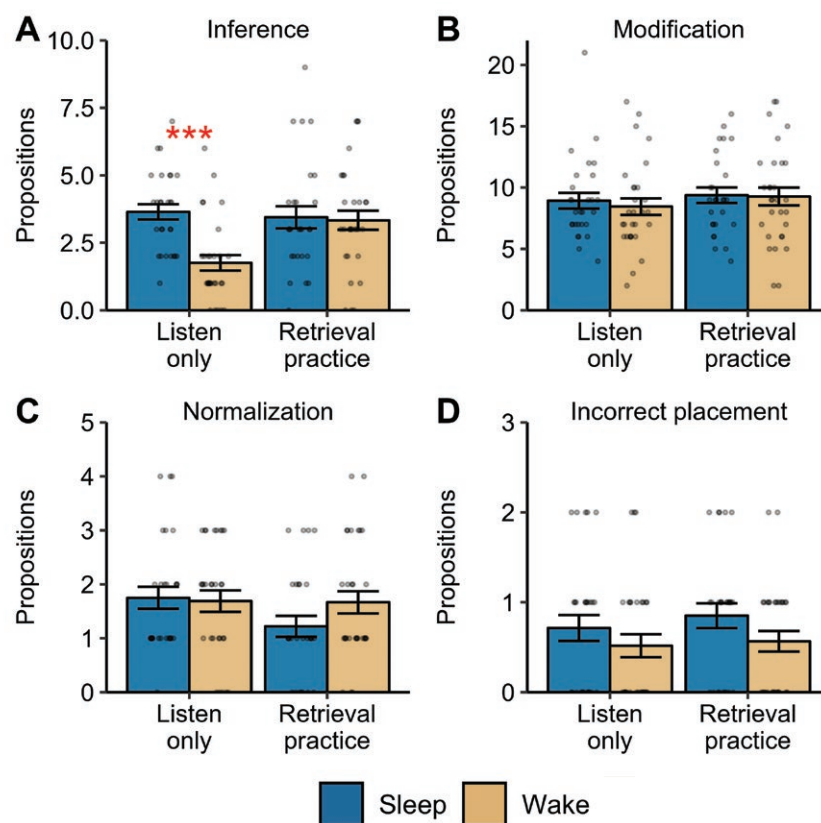


Figure 3. Effects of sleep and retrieval practice on distortion of the original story material. Dots represent individual data points, and error bars display the between-subjects standard error of the mean. *** $p < .001$.

number of inferences made at the delayed test, there was a significant main effect of delay group ($F(1,110) = 8.82$, $p_{\text{corr}} = .016$, $\eta^2 = .074$), with overall more inferences in participants who slept ($M = 3.55$, $SD = 1.81$) compared with participants who stayed awake ($M = 2.56$, $SD = 1.91$). This effect was superseded by a significant delay group $\times$ study condition interaction ($F(1,110) = 6.97$, $p_{\text{corr}} = .04$, $\eta^2 = .060$; Figure 3A). The number of inferences made by the sleep group was higher than the wake group when there was no retrieval practice ($t(55) = 4.67$, $p < .001$, $d = 1.24$), but there was no difference when retrieval practice occurred ($t(53) = 0.21$, $p = .84$, $d = 0.05$). The number of inferences made by the sleep group did not differ depending on whether participants engaged in retrieval practice or not ($t(53) = 0.40$, $p = .69$, $d = 0.11$). However, wake participants made significantly more inferences in the retrieval practice condition, compared to the listen-only condition ($t(57) = 3.44$, $p = .001$, $d = 0.90$). There were no main effects or interactions for modifications, normalizations, or incorrect placements made at the delayed memory test (all $p_{\text{corr}} > .067$; Figure 3B–D). No significant main effects or interactions for modifications were found after removal of an outlier ($ps > .17$).

These results suggest that there was an effect of sleep on increasing both accuracy and inference. However, this was true only in the absence of retrieval practice. We next sought to quantify the relative effects of sleep and retrieval practice on memory performance (Figure 4). Compared to the wake, listen-only group, both sleep and retrieval practice led to a significant improvement in accuracy, and significantly more inferences being made (all $ps < .001$, all $ds > 0.66$). Sleep and retrieval practice produced nearly identical changes in memory (accuracy increase: sleep M

[SD] = 1.58 [1.91], retrieval practice M [SD] = 1.32 [1.97]; inference increase: sleep M [SD] = 1.39 [1.69], retrieval practice M [SD] = 1.33 [1.94]). The influences of sleep and retrieval practice were not additive, with the sleep, retrieval practice group not showing any significant differences compared to the sleep, listen-only or wake, retrieval practice groups (main effect of group: $F(2,82) = 0.14$, $p = .87$, $\eta^2 = .003$). Aside from an overall reduction in number of propositions made, we did not observe any significant changes in propositions between the initial and delayed test in the two retrieval practice groups (all $ps > .07$).

Discussion

Here, we studied how two key processes, namely overnight sleep and post-learning retrieval practice, impacted memory for the War of the Ghosts story over a 12-hour delay. Our key findings were that, compared to a day of wake without retrieval practice, both sleep and retrieval practice independently resulted in more accurate propositions being produced 12 hours later, but at the same time, led to significantly more inferences of plausible, but not actually presented, story elements. This finding adds further evidence that the testing effect can be as powerful as sleep for long-term memory accuracy [33], and also extends the literature to show testing also contributes to memory evolution, again in a manner similar to sleep. Although such transformations can be beneficial in terms of memory flexibility and generalization, there are also times that such distortions of the original event are unwanted (e.g. eyewitness testimony). This is also an important consideration for educators, when considering the impacts of testing and sleep on the retention and accurate recall of educational materials.

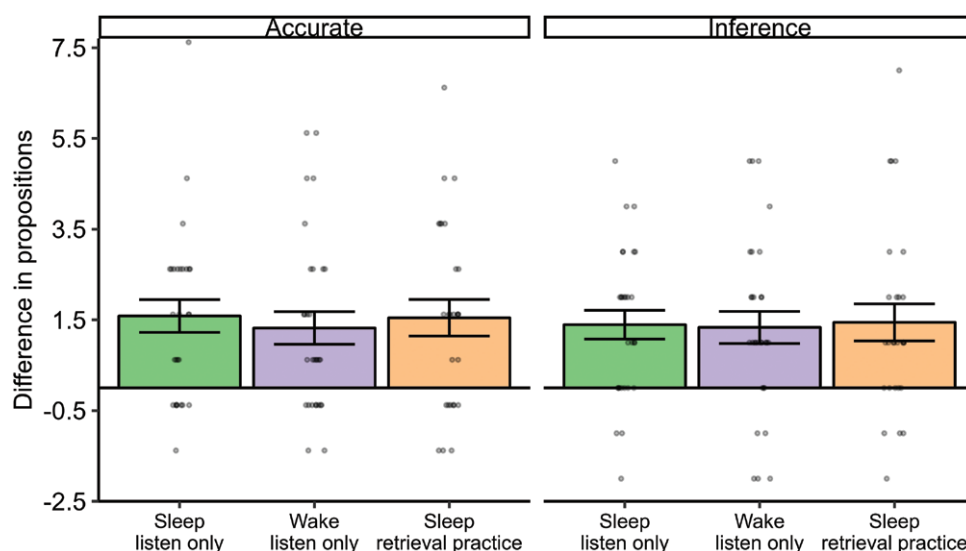


Figure 4. Relative effects of sleep and wake + retrieval practice on delayed memory performance. Difference scores were calculated by subtracting mean accuracy/inference from the wake, listen-only group from each participant's score in the other three groups. As such, a positive value indicates a relative increase with sleep and/or retrieval practice, and a negative value indicates a relative decrease. Error bars show the between-subjects standard error or the mean.

Our finding that sleep resulted in improved accuracy compared to wake fits with the large body of evidence that sleep strengthens memories for facts and events. For example, memory for word pairs is significantly better after a period of sleep compared to an equivalent period of wake [37, 45]. The present work shows that this benefit extends to a story-based task, similar to other work [46]. This finding is important, as it provides evidence that sleep's benefit to memory translates to a more ecologically valid test of memory.

We also found that participants who slept made more inferences than participants who stayed awake. As well as being strengthened, memories undergo transformation processes during sleep. One of these transformations involves improvements in relational memory, the ability to generalize new information across existing stores of knowledge. Prior experimental work has suggested that making new inferences about the relationships between abstract visual stimuli benefits from time, and especially sleep [47, 48]. In the present study, we defined an inference as a proposition that contained an event not explicitly stated in the story but that conceivably could have occurred between two explicitly described events. The fact that these occurred more frequently after sleep may reflect the sleeping brain's ability to generalize aspects of the story and utilize existing knowledge structures to form gist-consistent inferences regarding how events in the story likely linked together.

These effects were not unique to sleep. Post-learning retrieval practice, followed by a day of wake, produced similar changes in terms of behavior (i.e. numerically similar changes in proposition number), when compared to the wake, listen-only group (i.e. retrieval practice increased both accuracy and inference). There was no additive effect of sleep and retrieval practice, as the group that engaged in both retrieval practice and a night of sleep did not differ in delayed recall performance from the groups that received one or the other. One interpretation of this finding is that the changes in memory induced by retrieval practice altered the underlying memory representations to such a degree that sleep could not act on that representation further. This interpretation is supported in that, aside from an overall decline in total number

of propositions, there were no significant changes in propositions between the immediate and delayed test in the retrieval practice groups. A growing body of work has demonstrated other instances in which presleep retrieval practice precludes any benefit of sleep, compared to a restudy condition [32–34]. Although our findings appear to support this literature, we note that in the present study, we contrasted retrieval practice to a no retrieval, listen-only condition.

Together, our results replicate and extend previous work showing that memory effects that were once ascribed to being sleep-dependent can also be achieved through retrieval practice. Thus, future work should address the similarities and differences between how retrieval practice and sleep work to benefit memory. In a recent study, Liu and Ranganath found that while retrieval practice benefitted memory for associations that occurred in close temporal proximity, sleep uniquely facilitated the retention of related information that was learned in temporally distant episodes [49]. The authors proposed that this sleep-specific effect could be due to properties of the consolidation process that are unique to sleep, whereby sleep consolidation mechanisms extend the reach of retrieval practice by allowing the brain to discover links between temporally distant experiences [49].

At a neural level, it is proposed that sleep-associated memory benefits occur through a process of active systems consolidation, whereby memory traces are repeatedly reactivated, leading to a strengthening of neocortical memory representations and altering hippocampal–cortical connectivity [50–53]. Recent work has shown that repeated testing can induce similar systems consolidation effects on the neural substrates of memory [54], leading to the suggestion that retrieval practice initiates a rapid consolidation process, resulting in similar structural and functional changes as a night of sleep [55]. As such, it is plausible that sleep and retrieval practice may impact memory via similar changes to the underlying memory, albeit on different timescales. Although we did not see any additive effects of sleep and retrieval practice, recent work has shown that long-lasting changes to memory representations achieved via retrieval practice will only be

long-lasting if they are solidified by subsequent sleep [56]. An intriguing hypothesis for how sleep and retrieval practice may differ is that retrieval practice can rapidly initiate systems consolidation, but sleep is required for long-term stabilization.

Another potentially fruitful area for future research is to further examine how different types of memory might benefit differently from retrieval practice and sleep. Earlier studies have established that sleep and retrieval practice can both benefit memory for word pair associates and spatial location [32–34]. The present study extends this to memory accuracy and distortion in a more naturalistic story-based task. One explanation of these findings is that retrieval practice strengthens the underlying memory to such an extent that further consolidation during sleep is unneeded [33]. This is consistent with Dr. Stickgold's sleep-dependent memory-triage model, which suggests that sleep consolidates new information in a discriminatory manner, targeting the most relevant memories for consolidation [26]. Regarding memory strength in particular, Dr. Stickgold has suggested that memories with a very low or high initial strength may be triaged out of subsequent sleep-related processing [10, 37]. Therefore, if retrieval practice pushes memory strength above the "sweet spot" for sleep-dependent triage, the relative benefit of sleep on these memories will be reduced.

We found that the overall number of propositions made was higher in the sleep groups compared to the wake groups, irrespective of whether participants engaged in retrieval practice or not. Thus, this effect appeared to be unique to sleep. One possible interpretation is that sleep influenced the type of strategy participants engaged in during retrieval, such that sleep participants took a more liberal approach to writing down propositions they thought they recalled. It is also possible that participants in the sleep groups were simply more motivated at the final test. Though we do note that levels of subjective alertness did not differ between the sleep and the wake groups.

A number of limitations of the work should be considered. First, as already stated, we did not contrast retrieval practice to a restudy condition as is common in the testing effect literature. We do note however that some restudy occurred in all groups, as participants listened to the War of the Ghosts story twice during encoding. Even with this limitation, our results show that reengaging with studied material can boost memory in a manner equivalent to sleep, which is important in the context of other literature showing pre-sleep testing reduces the magnitude of the sleep effect [32–34]. Second, although our sample size was consistent with many sleep and memory studies [38–40], we did not perform an a priori power analysis due to the lack of a known effect size for this paradigm. Although the lack of an a priori power analysis is a limitation of the present work, the reported data provide an important baseline for which future high-powered studies can estimate expected effect sizes from. Third, we did not collect any data on the experiences of the participants in the wake group between encoding and the delayed test. Therefore, we don't know whether participants rehearsed any of the material between the sessions. Fourth, we did not collect any objective measurements regarding participant's sleep. As such, the quality and architecture of sleep following learning were unknown, and we do not know if any sleep stage of feature was specifically associated with any of the observed memory transformations. Finally, the scores in some of the proposition categories were very low, leaving open the possibility of floor effects in the case of those variables.

We found that both sleep and retrieval practice strengthened accurate recall of the story, while also inducing an increase in story-related inferences. These findings illustrate the importance

of both processes in memory evolution. When considering our work in the light of Dr. Stickgold's contributions to the field, two key themes emerge. First, our finding that inferences increase over time aligns with his suggestion that *memory evolution* is an appropriate phrase to refer to the many ways in which memories are transformed by post-encoding processes beyond simple strengthening [10]. Second, our finding that post-learning retrieval practice can promote the same memory changes as sleep can be viewed in the context of a memory-triage process [26], whereby the act of retrieval practice strengthens memory to a point that no additional reprocessing during sleep is needed [10, 33]. Future research should now further seek to determine memory processes that are unique to sleep, with a particular focus on the temporal distance between related episodes [49] and initial memory strength [10] as cues for sleep-dependent memory triage.

Supplementary Material

Supplementary material is available at *SLEEP Advances* online.

Acknowledgments

We would like to thank Jon Chamberlain for assistance in data collection, and Sara Kim for helpful comments on the manuscript.

Author Contributions

Dan Denis (Formal analysis [equal], Software [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]), Carissa DiPietro (Writing—original draft [equal], Writing—review & editing [equal]), R. Nathan Spreng (Conceptualization [equal], Methodology [equal], Writing—review & editing [equal]), Daniel L. Schacter (Conceptualization [equal], Funding acquisition [equal], Methodology [equal], Writing—review & editing [equal]), Robert Stickgold (Conceptualization [equal], Funding acquisition [equal], Methodology [equal], Writing—review & editing [equal]), and Jessica Payne (Conceptualization [equal], Funding acquisition [equal], Investigation [equal], Methodology [equal], Writing—review & editing [equal])

Funding

This study was supported by a National Institutes of Health grant (NIH MH048832) awarded to RS, a National Institute on Aging grant (R01 AG008441) awarded to DLS, and a National Science Foundation grant (BCS-2001025) awarded to JDP.

Conflict of Interest Statement

The authors declare no competing interests.

Disclosure Statement

Financial disclosure: none. Nonfinancial disclosure: none.

Data Availability Statement

The data sets generated and/or analyzed during the current study are available on the Open Science Framework, <https://doi.org/10.17605/OSF.IO/3B4HG>.

References

1. Lutz ND, Diekelmann S, Hinse-Stern P, Born J, Rauss K. Sleep supports the slow abstraction of gist from visual perceptual memories. *Sci Rep*. 2017;**7**(1):42950. doi:[10.1038/srep42950](https://doi.org/10.1038/srep42950)
2. Zeng T, Tompary A, Schapiro AC, Thompson-Schill SL. Tracking the relation between gist and item memory over the course of long-term memory consolidation. *eLife*. 2021;**10**:e65588. doi:[10.7554/eLife.65588](https://doi.org/10.7554/eLife.65588)
3. Bartlett FC. *Remembering: A Study in Experimental and Social Psychology*. Cambridge, UK: Cambridge University Press; 1932. doi:[10.1017/CBO9780511759185](https://doi.org/10.1017/CBO9780511759185)
4. Loftus EF. Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learn Mem*. 2005;**12**(4):361–366. doi:[10.1101/lm.94705](https://doi.org/10.1101/lm.94705)
5. Roediger HL, McDermott KB. Distortions of memory. In: Tulving E, Craik FIM, eds. *The Oxford Handbook of Memory*. Oxford, UK: Oxford University Press; 2000:149–162.
6. Schacter DL. *The Seven Sins of Memory: How the Mind Forgets and Remembers (Updated Edition)*. Boston, MA: Houghton Mifflin Harcourt; 2021.
7. Wetmore SA, Neuschatz JS, Gronlund SD, Wooten A, Goodsell CA, Carlson CA. Effect of retention interval on showup and lineup performance. *J Appl Res Mem Cogn*. 2015;**4**(1):8–14. doi:[10.1016/j.jarmac.2014.07.003](https://doi.org/10.1016/j.jarmac.2014.07.003)
8. Stickgold R. Sleep: off-line memory reprocessing. *Trends Cogn Sci*. 1998;**2**(12):484–492. doi:[10.1016/s1364-6613\(98\)01258-3](https://doi.org/10.1016/s1364-6613(98)01258-3)
9. Stickgold R. Sleep-dependent memory consolidation. *Nature*. 2005;**437**(7063):1272–1278. doi:[10.1038/nature04286](https://doi.org/10.1038/nature04286)
10. Stickgold R. How do I remember? Let me count the ways. *Sleep Med Rev*. 2009;**13**(5):305–308. doi:[10.1016/j.smrv.2009.05.004](https://doi.org/10.1016/j.smrv.2009.05.004)
11. Walker MP, Stickgold R. Sleep-dependent learning and memory consolidation. *Neuron*. 2004;**44**(1):121–133. doi:[10.1016/j.neuron.2004.08.031](https://doi.org/10.1016/j.neuron.2004.08.031)
12. Stickgold R, James L, Hobson JA. Visual discrimination learning requires sleep after training. *Nat Neurosci*. 2000;**3**(12):1237–1238. doi:[10.1038/81756](https://doi.org/10.1038/81756)
13. Stickgold R, Whidbee D, Schirmer B, Patel V, Hobson JA. Visual discrimination task improvement: a multi-step process occurring during sleep. *J Cogn Neurosci*. 2000;**12**(2):246–254. doi:[10.1162/089892900562075](https://doi.org/10.1162/089892900562075)
14. Walker MP, Brakefield T, Morgan A, Hobson JA, Stickgold R. Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron*. 2002;**35**(1):205–211. doi:[10.1016/s0896-6273\(02\)00746-8](https://doi.org/10.1016/s0896-6273(02)00746-8)
15. Payne JD, Stickgold R, Swanberg K, Kensinger EA. Sleep preferentially enhances memory for emotional components of scenes. *Psychol Sci*. 2008;**19**(8):781–788. doi:[10.1111/j.1467-9280.2008.02157.x](https://doi.org/10.1111/j.1467-9280.2008.02157.x)
16. Payne JD, Schacter DL, Propper RE, et al. The role of sleep in false memory formation. *Neurobiol Learn Mem*. 2009;**92**(3):327–334. doi:[10.1016/j.nlm.2009.03.007](https://doi.org/10.1016/j.nlm.2009.03.007)
17. Tamminen J, Payne JD, Stickgold R, Wamsley EJ, Gaskell MG. Sleep spindle activity is associated with the integration of new memories and existing knowledge. *J Neurosci*. 2010;**30**(43):14356–14360. doi:[10.1523/JNEUROSCI.3028-10.2010](https://doi.org/10.1523/JNEUROSCI.3028-10.2010)
18. Wamsley EJ, Tucker MA, Payne JD, Stickgold R. A brief nap is beneficial for human route-learning: the role of navigation experience and EEG spectral power. *Learn Mem*. 2010;**17**(7):332–336. doi:[10.1101/lm.1828310](https://doi.org/10.1101/lm.1828310)
19. Barsky MM, Tucker MA, Stickgold R. REM sleep enhancement of probabilistic classification learning is sensitive to subsequent interference. *Neurobiol Learn Mem*. 2015;**122**:63–68. doi:[10.1016/j.nlm.2015.02.015](https://doi.org/10.1016/j.nlm.2015.02.015)
20. Djonlagic I, Rosenfeld A, Shohamy D, Myers C, Gluck M, Stickgold R. Sleep enhances category learning. *Learn Mem*. 2009;**16**(12):751–755. doi:[10.1101/lm.1634509](https://doi.org/10.1101/lm.1634509)
21. Cai DJ, Mednick SA, Harrison EM, Kanady JC, Mednick SC. REM, not incubation, improves creativity by priming associative networks. *Proc Natl Acad Sci USA*. 2009;**106**(25):10130–10134. doi:[10.1073/pnas.0900271106](https://doi.org/10.1073/pnas.0900271106)
22. Durrant SJ, Taylor C, Cairney S, Lewis PA. Sleep-dependent consolidation of statistical learning. *Neuropsychologia*. 2011;**49**(5):1322–1331. doi:[10.1016/j.neuropsychologia.2011.02.015](https://doi.org/10.1016/j.neuropsychologia.2011.02.015)
23. Hanert A, Weber FD, Pedersen A, Born J, Bartsch T. Sleep in humans stabilizes pattern separation performance. *J Neurosci*. 2017;**37**(50):12238–12246. doi:[10.1523/JNEUROSCI.1189-17.2017](https://doi.org/10.1523/JNEUROSCI.1189-17.2017)
24. Sanders KEG, Osburn S, Paller KA, Beeman M. Targeted memory reactivation during sleep improves next-day problem-solving. *Psychol Sci*. 2019;**30**(11):1616–1624. doi:[10.1177/0956797619873344](https://doi.org/10.1177/0956797619873344)
25. Bergman ET, Roediger HL. Can Bartlett's repeated reproduction experiments be replicated. *Mem Cognit*. 1999;**27**(6):937–947. doi:[10.3758/bf03201224](https://doi.org/10.3758/bf03201224)
26. Stickgold R, Walker MP. Sleep-dependent memory triage: evolving generalization through selective processing. *Nat Neurosci*. 2013;**16**(2):139–145. doi:[10.1038/nrn.3303](https://doi.org/10.1038/nrn.3303)
27. Walker MP, Stickgold R. Overnight alchemy: sleep-dependent memory evolution. *Nat Rev Neurosci*. 2010;**11**(3):218–218. doi:[10.1038/nrn2762-c1](https://doi.org/10.1038/nrn2762-c1)
28. Cordi MJ, Rasch B. How robust are sleep-mediated memory benefits? *Curr Opin Neurobiol*. 2021;**67**:1–7. doi:[10.1016/j.conb.2020.06.002](https://doi.org/10.1016/j.conb.2020.06.002)
29. Adesope OO, Trevisan DA, Sundararajan N. Rethinking the use of tests: a meta-analysis of practice testing. *Rev Educ Res*. 2017;**87**(3):659–701. doi:[10.3102/0034654316689306](https://doi.org/10.3102/0034654316689306)
30. Roediger HL, Butler AC. The critical role of retrieval practice in long-term retention. *Trends Cogn Sci*. 2011;**15**(1):20–27. doi:[10.1016/j.tics.2010.09.003](https://doi.org/10.1016/j.tics.2010.09.003)
31. Rowland CA. The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychol Bull*. 2014;**140**(6):1432–1463. doi:[10.1037/a0037559](https://doi.org/10.1037/a0037559)
32. Abel M, Haller V, Köck H, et al. Sleep reduces the testing effect—but not after corrective feedback and prolonged retention interval. *J Exp Psychol Learn Mem Cogn*. 2019;**45**(2):272–287. doi:[10.1037/xlm0000576](https://doi.org/10.1037/xlm0000576)
33. Bäuml KHT, Holtermann C, Abel M. Sleep can reduce the testing effect: it enhances recall of restudied items but can leave recall of retrieved items unaffected. *J Exp Psychol Learn Mem Cogn*. 2014;**40**(6):1568–1581. doi:[10.1037/xlm0000025](https://doi.org/10.1037/xlm0000025)
34. Antony JW, Paller KA. Retrieval and sleep both counteract the forgetting of spatial information. *Learn Mem*. 2018;**25**(6):258–263. doi:[10.1101/lm.046268.117](https://doi.org/10.1101/lm.046268.117)
35. Bridge DJ, Paller KA. Neural correlates of reactivation and retrieval-induced distortion. *J Neurosci*. 2012;**32**(35):12144–12151. doi:[10.1523/JNEUROSCI.1378-12.2012](https://doi.org/10.1523/JNEUROSCI.1378-12.2012)
36. Hupbach A, Gomez R, Nadel L. Episodic memory reconsolidation: updating or source confusion? *Memory*. 2009;**17**(5):502–510. doi:[10.1080/09658210902882399](https://doi.org/10.1080/09658210902882399)
37. Denis D, Schapiro AC, Poskanzer C, et al. The roles of item exposure and visualization success in the consolidation of memories across wake and sleep. *Learn Mem*. 2020;**27**(11):451–456. doi:[10.1101/lm.051383.120](https://doi.org/10.1101/lm.051383.120)
38. Klinzing JG, Nienborg H, Rauss K. Sleep does not aid the generalisation of binocular disparity-based learning to the other visual hemifield. *J Sleep Res*. 2021;**30**:e13335. doi:[10.1111/jsr.13335](https://doi.org/10.1111/jsr.13335)

39. Noack H, Doeller CF, Born J. Sleep strengthens integration of spatial memory systems. *Learn Mem.* 2021;**28**(5):162–170. doi:[10.1101/lm.053249.120](https://doi.org/10.1101/lm.053249.120)
40. Scullin MK, McDaniel MA. Remembering to execute a goal: sleep on it! *Psychol Sci.* 2010;**21**(7):1028–1035. doi:[10.1177/0956797610373373](https://doi.org/10.1177/0956797610373373)
41. Hoddes E, Dement W, Zarcone V. The development and use of the Stanford sleepiness scale. *Psychophysiology.* 1972;**9**:150.
42. Mandler JM, Johnson NS. Remembrance of things parsed: story structure and recall. *Cogn Psychol.* 1977;**9**(1):111–151. doi:[10.1016/0010-0285\(77\)90006-8](https://doi.org/10.1016/0010-0285(77)90006-8)
43. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. <https://www.R-project.org>. Accessed February 25, 2019.
44. Kassambara A. Rstatix: Pipe-Friendly Framework for Basic Statistical Tests; 2021. <https://CRAN.R-project.org/package=rstatix>. Accessed February 25, 2019.
45. Talamini LM, Nieuwenhuis ILC, Takashima A, Jensen O. Sleep directly following learning benefits consolidation of spatial associative memory. *Learn Mem.* 2008;**15**(4):233–237. doi:[10.1101/lm.771608](https://doi.org/10.1101/lm.771608)
46. Aly M, Moscovitch M. The effects of sleep on episodic memory in older and younger adults. *Memory.* 2010;**18**(3):327–334. doi:[10.1080/09658211003601548](https://doi.org/10.1080/09658211003601548)
47. Ellenbogen JM, Hu PT, Payne JD, Titone D, Walker MP. Human relational memory requires time and sleep. *Proc Natl Acad Sci USA.* 2007;**104**(18):7723–7728. doi:[10.1073/pnas.0700094104](https://doi.org/10.1073/pnas.0700094104)
48. Huguet M, Payne JD, Kim SY, Alger SE. Overnight sleep benefits both neutral and negative direct associative and relational memory. *Cogn Affect Behav Neurosci.* 2019;**19**(6):1391–1403. doi:[10.3758/s13415-019-00746-8](https://doi.org/10.3758/s13415-019-00746-8)
49. Liu XL, Ranganath C. Resurrected memories: sleep-dependent memory consolidation saves memories from competition induced by retrieval practice. *Psychon Bull Rev.* 2021;**28**:2035. doi:[10.3758/s13423-021-01953-6](https://doi.org/10.3758/s13423-021-01953-6)
50. Gais S, Albouy G, Boly M, et al. Sleep transforms the cerebral trace of declarative memories. *Proc Natl Acad Sci USA.* 2007;**104**(47):18778–18783. doi:[10.1073/pnas.0705454104](https://doi.org/10.1073/pnas.0705454104)
51. Klinzing JG, Niethard N, Born J. Mechanisms of systems memory consolidation during sleep. *Nat Neurosci.* 2019;**22**(10):1598–1610. doi:[10.1038/s41593-019-0467-3](https://doi.org/10.1038/s41593-019-0467-3)
52. Payne JD, Kensinger EA. Sleep leads to changes in the emotional memory trace: evidence from fMRI. *J Cogn Neurosci.* 2011;**23**(6):1285–1297. doi:[10.1162/jocn.2010.21526](https://doi.org/10.1162/jocn.2010.21526)
53. Takashima A, Nieuwenhuis ILC, Jensen O, Talamini LM, Rijpkema M, Fernández G. Shift from Hippocampal to neocortical centered retrieval network with consolidation. *J Neurosci.* 2009;**29**(32):10087–10093. doi:[10.1523/JNEUROSCI.0799-09.2009](https://doi.org/10.1523/JNEUROSCI.0799-09.2009)
54. Brodt S, Gais S, Beck J, Erb M, Scheffler K, Schönauer M. Fast track to the neocortex: a memory engram in the posterior parietal cortex. *Science.* 2018;**362**(6418):1045–1048. doi:[10.1126/science.aau2528](https://doi.org/10.1126/science.aau2528)
55. Antony JW, Ferreira CS, Norman KA, Wimber M. Retrieval as a fast route to memory consolidation. *Trends Cogn Sci.* 2017;**21**(8):573–576. doi:[10.1016/j.tics.2017.05.001](https://doi.org/10.1016/j.tics.2017.05.001)
56. Himmer L, Schönauer M, Heib DPJ, Schabus M, Gais S. Rehearsal initiates systems memory consolidation, sleep makes it last. *Sci Adv.* 2019;**5**(4):eaav1695. doi:[10.1126/sciadv.aav1695](https://doi.org/10.1126/sciadv.aav1695)