BRIEF REPORT

Consolidation without intention: Sleep strengthens veridical and gist representations of information after incidental encoding

Elle M. D. Wernette[1] · Kimberly M. Fenn[1]

Abstract

Sleep strengthens declarative memory, but research investigating the effect of sleep on memory for information that is not explicitly studied for a test is sparse. In two experiments, we investigated the effect of sleep on gist-based and veridical representations of incidentally encoded information. Participants rated words from Deese–Roediger–McDermott (DRM) lists in either a deep or shallow encoding task and completed a surprise memory test after either sleep or wake. In Experiment 1, words were presented in lists, in order of descending associativity with the unpresented critical lure. Memory for list words and critical lures in both encoding tasks was stronger after sleep than wake, suggesting that sleep consolidated gist-based memory. In Experiment 2, the same words were presented in a random order across the experiment to minimize gist-based processing. Sleep strengthened veridical memory for list words following deep, but not shallow, encoding and did not affect critical lures. These results suggest sleep consolidates gist and veridical representations of information after incidental encoding, and that sleep-dependent consolidation processes may depend on processes at encoding, such as overlapping context and the strength of veridical memory traces.

Keywords Episodic memory · Sleep-dependent consolidation · Incidental encoding · Recognition memory · Gist

Individuals encode an overwhelming amount of information every day, but not all is remembered and only some is strengthened during sleep (Ellenbogen et al., 2006). A question remains as to what information is subject to sleep-dependent consolidation processes. Current research largely focuses on consolidation of information that is remembered for a subsequent test. When participants study paired associates for a test, recall performance is better after a period of sleep than waking (Ellenbogen et al., 2009; Fenn & Hambrick, 2012; Marshall et al., 2006; Tucker et al., 2006).

It is not clear if the expectation of a memory test influences sleep-dependent consolidation. One study found that sleep benefits information only if a test was expected (Wilhelm et al., 2011). After training on paired associates, participants who were told that there would be a test showed less forgetting after sleep than wake, but for participants who did not expect a test, performance was similar after sleep and wake. This finding suggests that sleep selectively consolidates memory based on factors such as intentionality (Born & Wilhelm, 2012; Rasch & Born, 2013), but was not replicated in later work (Ashton & Cairney, 2021). Due to these mixed results, it remains unclear if sleep affects memory for information that is not being remembered for a subsequent test.

Most studies on sleep-dependent consolidation focus on declarative memory for information that has been intentionally studied. Less is known about the extent to which sleep affects memory for information that is simply encountered. Although sparse, some studies have investigated memory for information that was incidentally encoded, meaning that individuals worked with information in some way but did not purposefully study it. The general finding is that sleep strengthens veridical memory after deep processing. Sleep strengthened memory for negative and neutral images when participants made arousal judgments (Baran et al., 2012), and strengthened recall after deep encoding but did not affect recognition after shallow encoding (Jurewicz et al., 2016). Importantly, there has not been a study that directly compared the effect of sleep after deep versus shallow encoding, using the same test. Thus, it remains unclear if sleep affects

✉ Elle M. D. Wernette
delgros1@msu.edu

1   Department of Psychology, Michigan State University, 316 Physics Road, Room 213, East Lansing, MI 48824, USA

information similarly based on depth of processing and what factors influence consolidation after incidental encoding.

Furthermore, although most consolidation research focuses on veridical representations, evidence from a diverse set of paradigms shows that sleep can abstract beyond studied information. In the number reduction task (NRT), seven steps are required to obtain the correct answer, but the task can be successfully solved in two steps; the answers to the second and seventh steps are identical. Participants are more likely to gain insight into this hidden rule after sleep than wake (Verleger et al., 2013; Wagner et al., 2004). Sleep also promotes resolution of remote associate test (RAT) problems. Participants are more likely to solve problems that they could not initially solve, if tested after sleep than wake (Cai et al., 2009; Sio et al., 2013). Thus, sleep can abstract conceptual representations that may not have been explicitly recognized during encoding.

Similarly, sleep can increase abstraction in the Deese–Roediger–McDermott (DRM) paradigm (Deese, 1959; Roediger III & McDermott, 1995), wherein participants study lists of semantically related words (*door, glass, pane*) that converge on a common theme, the critical lure (*window*). This critical lure is often falsely retrieved at test, potentially due to the use of gist-based representations to retrieve information (Brainerd & Reyna, 2001). When a recall test is used, participants show higher false recall of critical lures after sleep than wake (Chatburn et al., 2014; Diekelmann et al., 2010; Newbury & Monaghan, 2019; Payne et al., 2009). When a recognition test is used, sleep reduces false recognition of critical lures (Fenn et al., 2009; Lo et al., 2014), although a recent meta-analysis has cast doubt on the strength of these recognition effects (Newbury & Monaghan, 2019). There is no reason to assume these disparate results are due to differences in encoding or consolidation processes because the methods between these studies are similar until the test. Instead, differences in performance may suggest that different processes govern retrieval at test. During recall, participants may rely more on gist-based than veridical memory, especially at later positions in response output (Colbert & McBride, 2007; Roediger III & McDermott, 1995). In the present study, we were not interested in retrieval processes; instead, we focused on encoding, an underexplored area in the field.

In two experiments, we investigated consolidation of veridical and gist-based memory after incidental encoding and assessed the extent to which depth of processing affected sleep-dependent consolidation. Participants rated DRM words in a deep or shallow encoding task and completed a surprise recognition test after a 12-hour interval containing wake or sleep. In Experiment 1, participants rated words in order of descending associative strength with the critical lures; in Experiment 2, participants rated the same words in random order across the experiment. In Experiment 1, we predicted sleep would strengthen gist-based memory for list words and critical lures after deep, but not shallow, encoding. In Experiment 2, we predicted sleep would strengthen veridical memory for list words after deep, but not shallow, encoding and would not consolidate gist.

Experiment 1

Method

Participants

We recruited native English speaking undergraduate students from Michigan State University ($N = 193$) with no history of memory or sleep disorders. Although there are a few studies on sleep and memory for incidentally encoded information in the literature, we began data collection prior to their publication. Therefore, we used our prior work on sleep and memory for DRM stimuli (Fenn et al., 2009, Experiment 3) as a guide, and we aimed to recruit approximately the same number of participants (i.e., 30) in each of our Sleep and Wake groups as in the prior work. Thus, we aimed to recruit 120 participants between all experimental groups (Sleep, Deep; Sleep, Shallow; Wake, Deep; Wake Shallow), with 30 participants in each cell. We aimed to recruit an additional 120 participants for our control group (Morning and Evening). Our goal was to collect all of our data in one year; therefore, we also instituted a stopping rule such that data collection would end at the end of one academic year.

In total, we only collected data from 193 participants in the time we allotted for this experiment and stopped data collection at the end of one academic year. This resulted in 47 fewer participants than we originally planned. Some participants were excluded from analyses due to attrition ($N = 13$), incomplete data (i.e., experimental error or not performing the encoding task correctly, $N = 5$), or for napping during the waking retention interval ($N = 2$). The final sample contained 173 participants (101 female; $M_{age} = 20.22$ years, $SD_{age} = 3.0$). Participants provided informed consent and were compensated with course credit. This study was approved by the Michigan State University Institutional Review Board.

Design

Participants were assigned to either an experimental group (Wake, Sleep) or a time of day control group (Morning, Evening). Within each group, participants were randomly assigned to one of two encoding tasks (Deep, Shallow).

All participants completed an encoding phase and a surprise recognition test. For the experimental group, half of the participants completed encoding in the morning

(09:00–10:00) and the test in the evening (21:00–22:00), following a period of wakefulness (Wake). The other half completed encoding in the evening (21:00–22:00) and the test in the morning (09:00–10:00), after a night of sleep (Sleep). In the control group, participants completed encoding and test in a single session; half did so in the morning (09:00) and the other half, in the evening (21:00).

Materials

We chose 20 DRM lists based on backward associative strength (BAS), which is a measure of the association of each list word to their respective critical lure (Roediger III et al., 2001). These 20 lists were then divided into two sets, each containing 10 lists (Table S1). Sets were normed for BAS, false recognition rates, and connectivity (see the Supplemental Online Materials [SOM] for more information). List set was counterbalanced within each condition and performance across sets was similar, so we collapsed across the two versions in our analyses (for more information, see the SOM).

The test contained 80 items. All participants completed the same test, but the items varied based on which set of lists participants studied; items that were targets for some participants were lures for others. Of the 80 items, 30 were list words (from list positions 1, 8, and 10) and 10 were critical lures from the studied lists. The other 40 items were unrelated lures: 10 critical lures and 30 list words from the unstudied lists.

Procedure

Prior to the start of the experiment, we told participants that the experiment was investigating how time of day affects subjective word ratings and that they would be doing the same task in both sessions. This cover story was designed to ensure that encoding was incidental and that participants did not try to memorize the words.

During the encoding phase, participants first rated subjective sleepiness using the Stanford Sleepiness Scale (Hoddes et al., 1973; for details and data, see SOM) and then rated individual words in either a deep or shallow encoding task. In the deep encoding task, participants assessed how abstract each word was on a 7-point Likert scale ranging from concrete to abstract. We defined concrete as something that physically exists (e.g., *piano*) and abstract as a quality that does not physically exist in nature (e.g., *thought*). In the shallow encoding task, participants simply indicated the number of vowels in each word on a 7-point scale. Thus, in the deep encoding task, participants encoded semantic features of each word whereas in the shallow encoding task, participants simply encoded visual features. List words were presented individually on the computer screen, in order of

descending associativity with the critical lure. No time limit was imposed; participants had as much time as they needed to make a decision. Each experimental block contained the 15 words from a single DRM list (Roediger III et al., 2001). After each block, participants had the opportunity to take a five second break. After this, experimental participants left the laboratory and control participants listened to a 2-minute audio clip to reduce rehearsal.

During the test phase, either immediately after the audio clip (control group) or after a 12-hour retention interval (experimental group), participants were informed that they would take a surprise recognition memory test. Experimental participants first completed the Stanford Sleepiness Scale and then the test. On the test, participants provided old/new judgments on each word, presented in a random order. They were instructed to indicate that a word was "old" if they had rated it during the encoding phase and to indicate that a word was "new" if they had not rated it. After each "old" response, participants were asked to rate their confidence on a 7-point Likert scale ranging from "Not confident at all" to "Extremely confident." Participants were instructed to take as much time as needed to maximize accuracy. After completing the test, participants completed the Morningness-Eveningness Questionnaire, an 18-item survey that measures chronotype or preferred time of day (Horne & Östberg, 1976; see SOM for more details) and a demographic questionnaire.

Results

To assess memory discrimination, we used hits and false alarms to compute d-prime (d'); higher d' represents better discriminability (Macmillan & Creelman, 2004). For list words, hits were the proportion of times participants responded "old" to list words from studied lists and false alarms were the proportion of times participants responded "old" to list words from unstudied lists. Because we were interested in gist-based representations, we computed d' for critical lures by using the proportion of "old" responses to critical lures as hits (even though these are actually false alarms) and using the proportion "old" responses to critical lures from unstudied lists as false alarms. This is the appropriate comparison because critical lures tend to be more frequent and familiar in the English language and have higher false-alarm rates than list words (Roediger III & McDermott, 1995). In cases where the hit rates were 1 or false-alarm rates were 0, we replaced these values with $1-(1/2n)$ and $1/2n$, respectively (Macmillan & Creelman, 2004; Macmillan & Kaplan, 1985).

We first performed a mixed-design ANOVA on d' with item type (List Words, Critical Lures) as a within-subjects factor and condition (Wake, Sleep) and encoding (Deep, Shallow) as between-subjects factors. Descriptive statistics are displayed in Table 1, and results are displayed in Fig. 1.

Table 1 Experiment 1 means (and standard deviations) for the proportion of "Old" responses to the various types of test stimuli and d' scores for list words and critical lures across experimental and control groups

Group	Encoding	n	List Words	List Words Unstudied Lists	Critical Lures	Critical Lures Unstudied Lists	d' List Words	d'' Critical Lures
Wake	Deep	22	.79 (.17)	.23 (.11)	.62 (.19)	.37 (.19)	1.73 (0.78)	0.72 (0.66)
	Shallow	24	.60 (.15)	.38 (.16)	.50 (.21)	.45 (.21)	0.63 (0.41)	0.16 (0.56)
Sleep	Deep	22	.90 (.09)	.20 (.10)	.67 (.18)	.27 (.19)	2.33 (0.59)	1.17 (0.73)
	Shallow	25	.68 (.16)	.34 (.15	.58 (.21)	.41 (.24)	0.97 (0.55)	0.49 (0.59)
Morning control	Deep	18	.87 (.12)	.16 (.12)	.57 (.20)	.26 (.21)	2.37 (0.79)	0.97 (0.68)
	Shallow	21	.71 (.15)	.30 (.22)	.48 (.20)	.40 (.23)	1.28 (0.83)	0.23 (0.63)
Evening control	Deep	21	.88 (.18)	.15 (.11)	.59 (.22)	.23 (.18)	2.59 (0.80)	1.06 (0.81)
	Shallow	20	.74 (.14)	.36 (.15)	.54 (.24)	.48 (.28)	1.12 (0.71)	0.17 (0.76)
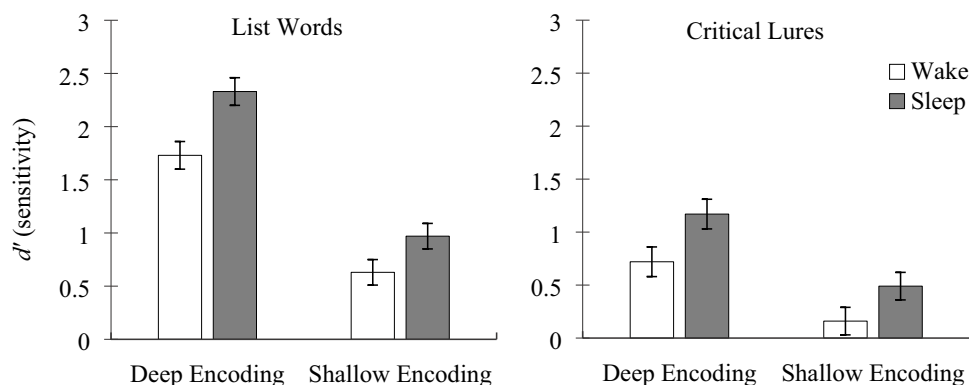
Fig. 1 Sensitivity (d') for the Wake and Sleep conditions after Deep and Shallow encoding for list words and critical lures in Experiment 1. *Note.* Error bars represent the standard error of the mean

As expected, there was a main effect of encoding, $F(1, 89) = 90.50$, $p < .001$, $\eta_p^2 = 0.50$; deep encoding led to higher d' than shallow, and a main effect of item type, $F(1, 89) = 89.36$, $p < .001$, $\eta_p^2 = 0.50$; d' was higher for list words than critical lures. Importantly, there was a main effect of condition, $F(1, 89) = 19.41$, $p < .001$, $\eta_p^2 = 0.18$. Participants showed higher d' after sleep than waking. There was not an interaction between condition and encoding, $F(1, 89) = 0.92$, $p = .34$. There was an interaction between encoding and item type, $F(1, 89) = 13.43$, $p < .001$, $\eta_p^2 = 0.13$; but this interaction was not relevant to our research question. Finally, there were no interactions between condition and item type or condition, item type, and encoding, $Fs < 1$. Results regarding differences in response bias and confidence did not inform the main analyses and can be found in the Supplemental Online Materials (SOM).

To ensure our results were not affected by diurnal or circadian effects, we conducted a multivariate repeated-measures ANOVA on d' for control participants with item type (List Words, Critical Lures) as a within-subjects factor, and encoding (Deep, Shallow) and time (Morning, Evening) as between-subjects factors (Table 1). Importantly, there was

not a main effect of time, $F(1, 76) = 0.03$, $p = .87$, and there were not any interactions between time and encoding or item type, $Fs < 1$ (Fig. S4). Finally, there was no evidence that sleepiness or chronotype affected our primary results (see SOM).

Thus, we found evidence of sleep-dependent consolidation after both deep and shallow encoding for list words and critical lures. This suggests that sleep may have consolidated the overall theme of the list, or gist, which can account for both increased correct memory and increased false memory of critical lures (Brainerd & Reyna, 1990; Cann et al., 2011). Thus, it remains unclear if sleep consolidated veridical memory. We explore this possibility in Experiment 2.

Experiment 2

To better elucidate the effect of sleep on veridical memory following incidental encoding, we used the same words and encoding tasks as in Experiment 1; however, to reduce gist-based processing, participants rated words in random order across the encoding phase (Mather et al., 1997). We

predicted that after deep encoding, the Sleep group would show higher sensitivity to list words than Wake. We were unsure if sleep would affect memory in the shallow encoding task. We also expected critical lure sensitivity would not differ between Wake and Sleep after either encoding task.

Method

Participants

We conducted an a priori power analysis in G*Power to estimate sample size for the experimental groups (i.e., participants in the Sleep and Wake conditions). The power analysis was conducted for an ANOVA with fixed effects and interactions to find a moderate effect ($f = 0.35$) with similar power to that in Experiment 1 ($1 - \beta = .945$). This analysis revealed a necessary sample size of 141 total participants across the two delay conditions and two encoding tasks. To balance groups, we aimed to have a sample of 144 participants in the experimental group and another 144 participants in the control group, for a total of 288 participants. To account for attrition and data loss due to napping, we recruited 393 undergraduates from Michigan State University who did not participate in Experiment 1. All participants were native English speakers with no history of memory or sleep disorders. Participants did not have strong time of day preferences (scores on the Morningness-Eveningness Questionnaire between 42 and 58; Horne & Östberg, 1976). Additionally, all participants had generally healthy sleep quality (score between 0 and 12 on the sleep disturbances scale of the Pittsburgh Sleep Quality Index [PSQI]; Buysse et al., 1989). Several participants were excluded from all analyses due to attrition ($N = 24$), napping during the waking retention interval ($N = 22$), or data loss caused by a program error ($N = 13$) or experimenter error ($N = 35$). The final sample included 299 participants (200 females; $M_{age} = 19.32$, $SD_{age} = 1.21$).

Procedure

This experiment was nearly identical to Experiment 1 with two exceptions. Critically, in Experiment 1, we presented words grouped by list and in order of descending associativity with the critical lure. In this experiment, we presented the same words, but the words were not grouped in lists. Instead, they were presented in random order across the encoding phase. The basic structure of the experiment was the same as in Experiment 1; participants rated 15 words in each of ten blocks, with a five-second break between blocks. At the beginning of each session, participants completed both the Stanford Sleepiness Scale and the Positive and Negative Affect Schedule, which assesses mood (Thompson, 2007; see SOM for more information and data). The recognition memory test was the same test used in Experiment 1, except participants rated their confidence for both "old" and "new" judgments instead of only "old" judgments.

Results

We used a mixed-design ANOVA on d' with item type (List Words, Critical Lures) as a within-subjects factor and condition (Wake, Sleep) and encoding (Deep, Shallow) as between-subjects factors (Fig. 2). There were main effects of encoding, $F(1, 140) = 57.06$, $p < .001$, $\eta_p^2 = 0.29$, and item type, $F(1, 140) = 68.01$, $p < .001$, $\eta_p^2 = 0.33$; d' was higher after deep than shallow encoding and higher for list words than critical lures (Table 2). There was not a significant main effect of condition, $F(1, 140) = 2.24$, $p = .14$, but there was a three-way interaction between condition, item type, and encoding, $F(1, 140) = 4.33$, $p = .04$, $\eta_p^2 = 0.03$. To follow up this interaction, we ran a pair of mixed-design
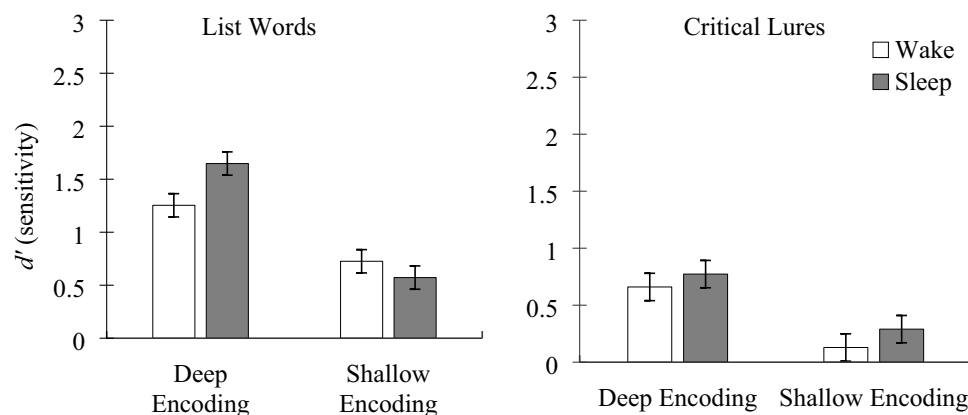

Fig. 2 Sensitivity (d') to list words and critical lures across condition and encoding task when DRM list words were randomly presented at encoding in Experiment 2. *Note.* Errors bars represent the standard error of the mean

Table 2 Experiment 2 means (and standard deviations) for the proportion of "Old" responses to the various types of test stimuli across all experimental and control groups and encoding tasks

Group	Encoding	n	List Words	List Words Unstudied Lists	Critical Lures	Critical Lures Unstudied Lists	d' List Words	d' Critical Lures
Wake	Deep	35	.81 (.15)	.36 (.16)	.67 (.19)	.45 (.24)	1.25 (0.77)	0.66 (0.77)
	Shallow	36	.69 (.14)	.45 (.17)	.58 (.18)	.52 (.18)	0.72 (0.43)	0.13 (0.56)
Sleep	Deep	38	.81 (.15)	.28 (.13)	.66 (.21)	.40 (.23)	1.65 (0.74)	0.77 (0.73)
	Shallow	35	.69 (.14)	.48 (.17)	.68 (.16)	.58 (.20)	0.57 (0.57)	0.29 (0.70)
Morning controls	Deep	40	.90 (.09)	.21 (.11)	.55 (.17)	.30 (.19)	2.34 (0.67)	0.75 (0.62)
	Shallow	37	.66 (.15)	.39 (.16)	.57 (.21)	.43 (.18)	0.79 (0.79)	0.41 (0.64)
Evening controls	Deep	40	.86 (.16)	.24 (.15)	.60 (.21)	.34 (.16)	2.06 (0.96)	0.74 (0.66)
	Shallow	38	.71 (.15)	.41 (.16)	.62 (.18)	.52 (.21)	0.89 (0.57)	0.29 (0.71)

ANOVAs on list words and critical lures with condition and encoding as factors. For list words, there was a main effect of encoding, $F(1, 140) = 55.16$, $p < .001$, $\eta_p^2 = 0.28$; d' was higher after deep than shallow encoding. There was not a main effect of condition, $F(1, 140) = 1.24$, $p = .27$. The interaction between encoding and condition was significant, $F(1, 140) = 6.42$, $p = .01$, $\eta_p^2 = 0.04$; d' was higher after sleep after deep encoding, $t(71) = 2.21$, $p = .03$, $d = 0.52$, but not shallow, $t(69) = 1.28$, $p = .20$. For critical lures, we conducted another mixed-design ANOVA, which showed a main effect of encoding, $F(1, 140) = 18.94$, $p < .001$, $\eta_p^2 = 0.12$. There was not an effect of condition, $F(1, 140) = 1.43$, $p = .23$, or an interaction between condition and encoding, $F(1, 140) = 0.04$, $p = .84$. Although we designed this experiment to minimize gist, critical lure sensitivity was greater than zero in the Sleep group after both deep, $t(37) = 6.49$, $p < .001$, $d = 1.05$, and shallow encoding, $t(35) = 2.47$, $p = .02$, $d = 0.42$, and in the Wake group, after deep encoding, $t(36) = 5.06$, $p < .001$, $d = 0.86$, but not after shallow, $t(35) = 1.38$, $p = .18$.

There was an interaction between encoding and item type, $F(1, 140) = 4.34$, $p = .04$, $\eta_p^2 = 0.03$, but the interactions between condition and item type, $F(1, 140) = 0.02$, $p = .90$, and condition and encoding, $F(1, 140) = 2.11$, $p = .15$, were not significant. Results regarding sleepiness, mood, confidence, and response bias did not inform the primary analyses (see SOM).

Again, we conducted a repeated-measures ANOVA on d' on the control data with item type (List Words, Critical Lures) as a within-subjects factor and time (Morning, Evening) and encoding (Deep, Shallow) as between-subject factors. Importantly, there was not a main effect of time, $F(1, 151) = 0.75$, $p = .39$ (Fig. S8), or an interaction between encoding and time, $F(1, 151) = 0.60$, $p = 0.44$, or time, encoding, and item type, $F(1, 151) = 3.03$, $p = .08$ (SOM).

General discussion

This work provides the first evidence that sleep consolidates gist, as well as veridical memory, following incidental encoding. In Experiment 1, list words were presented in order of associative strength, and gist-based memory was stronger after sleep than wake following both deep and shallow encoding; sensitivity was higher to list words and critical lures. Although the effect in list words could reflect consolidation of veridical memory, the effect in critical lures suggests gist consolidation. In Experiment 2, list words were presented in a random order across blocks, and sleep strengthened veridical memory for list words after deep but not shallow encoding, and there was no evidence of gist consolidation after sleep for either encoding task.

Thus, across two experiments, we found that sleep increased gist-based memory when list words were presented in order of associative strength, and veridical memory when list words were presented randomly. The veridical memory effect is consistent with prior work showing veridical consolidation after deep but not shallow encoding (Jurewicz et al., 2016), potentially due to memory strength. It is possible that there is a strength threshold that needs to be reached for memories to be consolidated; sleep may consolidate memories above this threshold but not weaker memories. Deep encoding produces stronger memory than shallow (Craik & Lockhart, 1972; Goldman & Pellegrino, 1977) and may produce memories that reach the threshold for consolidation.

The results in gist memory are a novel finding and are a bit more surprising. First, the increased false memory of critical lures after sleep in Experiment 1, suggestive of gist consolidation, more strongly resembles DRM studies that use recall than recognition. When DRM lists are intentionally remembered, sleep consolidates gist-based memory of

critical lures, in free recall (Chatburn et al., 2014; Diekelmann et al., 2010; Payne et al., 2009), but when recognition is tested, studies show either lower critical lure false recognition after sleep (Fenn et al., 2009; Lo et al., 2014), or no effect of sleep (Newbury & Monaghan, 2019). One possible explanation for these disparate results is that participants may rely more on gist-based representations to generate target words during recall, whereas during recognition, they may be more likely to use veridical representations to monitor memory to accept and reject words. We propose that incidental encoding may also affect strategies used at test and encourage individuals to rely more on gist-based representations. During incidental encoding, individuals have no reason to suspect that the information will be important in the future. As such, they may not encode source information or item-specific information. Prior research has shown that participants are more likely to use gist-based representations at test if they do not have access to source information, or if veridical memory is weak (Dodhia & Metcalfe, 1999; Johnson et al., 1993; Lindsay & Johnson, 2000). Thus, conditions at encoding, such as intentionality, may affect retrieval processes and strategies.

The gist results are also surprising because we did not find increased gist memory after sleep in Experiment 2, even though the same encoding tasks were used and critical lure sensitivity (measured by d') was reliably greater than zero. Thus, despite random presentation, some gist memory was formed that could have been consolidated by sleep, likely because participants rated some words from the same DRM lists in the same block. A question, therefore, remains as to why gist memory was consolidated in Experiment 1 and not Experiment 2. It is possible that the strength of the gist memory in Experiment 2 was simply not sufficient for consolidation, similar to the lack of an effect of sleep on veridical memory after shallow encoding. It is also possible that the results of Experiment 1 do not actually reflect gist consolidation. Critical lure sensitivity in Experiment 1 may be explained by familiarity at retrieval since memory for list words was better after sleep than wake in both encoding groups. Given that participants were tested on three list words and one critical lure, participants may have responded based on familiarity to other test items or based on their responses to similar test items (Coane & McBride, 2006; Dewhurst et al., 2011). An alternative, albeit not exclusive, explanation is that spreading activation (Anderson, 1983; Collins & Loftus, 1975) and internal generation of the critical lure at encoding (Marsh & Bower, 2004) may have contributed to critical lure consolidation in Experiment 1. The order of words in each block may have caused the critical lure to be activated at encoding and subsequently consolidated during sleep. In Experiment 2, it is less likely that the critical lure would be internally-generated. Further research is necessary to distinguish between these alternatives.

The above arguments are all suggestive of an active consolidation process, such as memory replay during sleep. However, it is also possible that our results reflect a passive account. Specifically, individuals experience very little interference during sleep compared to waking; memories encoded prior to a period of sleep are less likely to be disrupted by interference than memories encoded prior to waking (Yonelinas et al., 2019). This account would predict better memory after sleep, than waking, albeit by a different mechanism. If passive protection from interference was the only mechanism underlying these results, we would have expected stronger performance after sleep for all types of items- both list words and critical lures, after both encoding tasks- which we did not find. The present experiment was not designed to distinguish between active and passive accounts, but we believe that our results more strongly suggest an active process.

An important consideration when interpreting our results with respect to the literature is our measure of signal detection. We analyzed our data using d'; however, some (Lo et al., 2014; Pardilla-Delgado & Payne, 2017; Shaw & Monaghan, 2017), but not all (Diekelmann et al., 2008; Huan et al., 2022; Jano et al., 2021), studies on sleep and false memory use A', an alternate measure of signal detection. We chose to use d' because it is the most widely used and well-understood. Importantly, we replicated our primary results in both experiments using A' (see SOM for further detail).

In conclusion, we found that sleep consolidated information that is processed, but not actively remembered. Our results suggest that what individuals remember after sleep likely depends on processes at encoding, such as the presence of shared context and strength of veridical memory prior to sleep. Thus, the present study provides new insight into the nature of sleep-dependent consolidation processes and suggests that memory about everyday life may be strengthened during sleep. Although memory advantages of sleep-dependent consolidation processes may not be ubiquitous, they act on more information than was previously demonstrated.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13423-023-02247-9.

References

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal Learning and Verbal Behavior, 22*(3), 261–295.

Ashton, J. E., & Cairney, S. A. (2021). Future-relevant memories are not selectively strengthened during sleep. *PLOS ONE, 16*(11), Article e0258110.

Baran, B., Pace-Schott, E. F., Ericson, C., & Spencer, R. M. C. (2012). Processing of emotional reactivity and emotional memory over sleep. *Journal of Neuroscience, 32*(3), 1035–1042.

Born, J., & Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychological Research, 76*(2), 192–203.

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review, 10*, 3–47.

Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience. In H. W. Reese & R. Kail (Eds.), *Advances in child development and behavior* (pp. 41–100). Academic Press.

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

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, 106*(25), 10130–10134.

Cann, D. R., McRae, K., & Katz, A. N. (2011). False recall in the Deese–Roediger–McDermott paradigm: The roles of gist and associative strength. *Quarterly Journal of Experimental Psychology, 64*(8), 1515–1542.

Chatburn, A., Lushington, K., & Kohler, M. J. (2014). Complex associative memory processing and sleep: A systematic review and meta-analysis if behavioral evidence and underlying EEG mechanisms. *Neuroscience and Biobehavioral Reviews, 47*, 645–666.

Coane, J. H., & McBride, D. M. (2006). The role of test structure in creating false memories. *Memory & Cognition, 34*, 1026–1036.

Colbert, J. M., & McBride, D. M. (2007). Comparing decay rates for accurate and false memories in the DRM paradigm. *Memory & Cognition, 35*(7), 1600–1609.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*(1), 17–22.

Dewhurst, S. A., Knott, L. M., & Howe, M. L. (2011). Test-induced priming impairs source monitoring accuracy in the DRM procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(4), 1001.

Diekelmann, S., Landolt, H. P., Lahl, O., Born, J., & Wagner, U. (2008). Sleep loss produces false memories. *PLOS ONE, 3*(10), Article e3512.

Diekelmann, S., Born, J., & Wagner, U. (2010). Sleep enhances false memories depending on general memory performance. *Behavioral Brain Research, 208*(2), 425–429.

Dodhia, R. M., & Metcalfe, J. (1999). False memories and source monitoring. *Cognitive Neuropsychology, 16*(3/5), 489–508.

Ellenbogen, J. M., Payne, J. D., & Stickgold, R. (2006). The role of sleep in declarative memory consolidation: Passive, permissive, active or none? *Current Opinions in Neurobiology, 16*(6), 716–722.

Ellenbogen, J. M., Hulbert, J. C., Jiang, Y., & Stickgold, R. (2009). The sleeping brain's influence on verbal memory: boosting resistance to interference. *PLOS ONE, 4*(1), Article e4117.

Fenn, K. M., & Hambrick, D. Z. (2012). Individual differences in working memory capacity predict sleep-dependent memory consolidation. *Journal of Experimental Psychology: General, 141*(3), 404–410.

Fenn, K. M., Gallo, D. A., Margoliash, D., Roediger, H. L., III, & Nusbaum, H. C. (2009). Reduced false memory after sleep. *Learning & Memory, 16*, 509–513.

Goldman, S. R., & Pellegrino, J. W. (1977). Processing domain, encoding elaboration, and memory trace strength. *Journal of Verbal Learning and Verbal Behavior, 16*(1), 29–43.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: A new approach. *Psychophysiology, 10*, 431–436.

Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology, 4*, 97–110.

Huan, S. Y., Xu, H. Z., Wang, R., & Yu, J. (2022). The different roles of sleep on false memory formation between young and older adults. *Psychological Research, 86*, 443–451.

Jano, S., Romeo, J., Hendrickx, M. D., Schlesewsky, M., & Chatburn, A. (2021). Sleep influences neural representations of true and false memories: An event-related potential study. *Neurobiology of Learning and Memory, 186*, Article 107553.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3.

Jurewicz, K., Cordi, M. J., Staudigl, T., & Rasch, B. (2016). No evidence for memory decontextualization across one night of sleep. *Frontiers in Human Neuroscience, 10*, 7.

Lindsay, D. S., & Johnson, M. K. (2000). False memories and the source monitoring framework: Reply to Reyna and Lloyd (1997). *Learning and Individual Differences, 12*(2), 145–161.

Lo, J. C., Sim, S. K. Y., & Chee, M. W. L. (2014). Sleep reduces false memory in healthy older adults. *Sleep, 37*(4), 665–671.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). Psychology Press.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin, 98*(1), 185.

Marsh, E., & Bower, G. (2004). The role of rehearsal and generation in false memory creation. *Memory, 12*(6), 748–761.

Marshall, L., Helgadóttir, H., Mölle, M., & Born, J. (2006). Boosting slow oscillations during sleep potentiates memory. *Nature, 444*, 610–613.

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition, 25*(6), 826–837.

Newbury, C. R., & Monaghan, P. (2019). When does sleep affect veridical and false memory consolidation? A meta-analysis. *Psychonomic Bulletin & Review, 26*(2), 387–400.

Pardilla-Delgado, E., & Payne, J. D. (2017). The impact of sleep on true and false memory across long delays. *Neurobiology of Learning and Memory, 137*, 123–133.

Payne, J. D., Schacter, D. L., Propper, R., Huang, L.-W., Wamsley, E., Tucker, M. A., Walker, M. P., & Stickgold, R. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory, 92*(3), 327–334.

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews, 93*(2), 681–766.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814.

Roediger, H. L., III, McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*(3), 385–407.

Shaw, J. J., & Monaghan, P. (2017). Lateralised sleep spindles relate to false memory generation. *Neuropsychologia, 107*, 60–67.

Sio, U. N., Monaghan, P., & Ormerod, T. (2013). Sleep on it, but only if it is difficult: Effects of sleep on problem solving. *Memory & Cognition, 41*(2), 159–166.

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology, 38*(2), 227–242.

Tucker, M. A., Hirota, Y., Wamsley, E. J., Lau, H., Chaklader, A., & Fishbein, W. (2006). A daytime nap containing solely non-REM sleep enhances declarative but not procedural memory. *Neurobiology of Learning and Memory, 86*(2), 241–247.

Verleger, R., Rose, M., Wagner, U., Yordanove, J., & Kolev, V. (2013). Insights into sleep's role for insight: Studies with the number reduction task. *Advances in Cognitive Psychology, 9*(4), 160–172.

Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature, 427*, 352–355.

Wilhelm, I., Diekelmann, S., Molzow, I., Ayoub, A., Mölle, M., & Born, J. (2011). Sleep selectively enhances memory expected to be of future relevance. *Journal of Neuroscience, 31*(5), 1563–1569.

Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: systems consolidation reconsidered. *Nature Reviews Neuroscience, 20*(6), 364–375.